Carbohydrate Binding Module 74 (CBM74) Is a Specialized Resistant Starch Binding Domain

by

Amanda L. Photenhauer

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in the University of Michigan
2023

Doctoral Committee:

Associate Professor Nicole Koropatkin, Chair
Professor Melanie Ohi
Professor Maria Sandkvist
Professor Joel Swanson
Professor Vince Young

Amanda L. Photenhauer
alphote@umich.edu
ORCID iD: 0000-0002-1553-4772

ACKNOWLEDGEMENTS

This dissertation really comes full circle, from my undergraduate degree to Ph.D. in this same department. I am so fortunate to have been supported by some really key players in my doctoral journey.

First, thank you to Theresa Keeley who saw something special in me among the contenders for a laboratory assistant position in the Samuelson lab as an undergraduate student. Once I started working in a research lab I knew I had found where I belonged. To my mentors in the Samuelson lab who showed me the ropes and how to be a good scientist: Dr. Linda Samuelson, Theresa Keeley, Dr. Kelli VanDussen, and Dr. Elise Demitrack - Thank you for teaching me and believing in me.

To Dr. Juanita Merchant who hired me as a technician but trained me like a graduate student - Thank you for trusting me to lead a project and advocating for me to be a part of the scientific community. Thank you also to Dr. Sinju Sundaresan and Dr. Nataliya Razumilava.

To my parents and siblings who always see the best in me; my top-tier patrons: Sarah, Sarah, and Steph (British Sarah); and my best friend and incredibly supportive partner, André - Thank you for the unconditional love and support.

To my friends who supported me during this long process, thank you for the memories of D&D and pizza, the twilight hours acting out Shakespeare in the Arb, and the lock-down marathons and long bike rides. You've taught me to see the joy in the journey and the perseverance to reach the finish line.

To our compatriots in the Martens lab - May you always be the other side of the coin. I want to issue a special thank you to Ana Sofia Luis who taught me to hit the ground running in a research project.

To the undergraduate researchers who worked on these projects with me: Anna, Adeline, Cameron, and Jack - You've left your imprint in all the proteins purified, data points collected, and crystal trays set up. Thank you.

To the Koropatkin Lab members past and present - Thank you for your companionship and sound advice. I have been so fortunate to be a part of this team. I have truly grown as a scientist from the guidance of these people especially: Dr. Krista Armbruster, Dr. Haley Brown, and Dr. Rebecca Pollet. Thank you for everything you have taught me. I want to issue a special thank you

to Dr. Filipe Cerqueira who started this journey with me in the Koropatkin Lab. Thank you for being my friend, my collaborator, my sounding board, and sailing this river with me the whole way.

To all of my collaborators - Thank you for expanding the breadth of my research and sharing in the joys and accomplishments. To my Sas6 collaborators: Krista, Filip and Stefan, Tiantian and Craig, Rosie and Brandon, Filipe and Jesse - We did it! I need to issue a special thank you to Dr. Zdzislaw Wawrzak, the beamline scientist from LS-CAT. I've never seen your face but I've seen your datasets. Thank you so much for always helping me collect and process datasets, answering all of my crystallography questions, and consoling me when my crystals didn't diffract. I hope these structures bring you joy. To my collaborators for Appendix B: Dr. Tom Schmidt, Dr. Clegg Waldron, Haiyan Tang, and Dr. Ethan Hillman - Thank you for planting the seeds of this project. To the Ohi lab members: Dr. Katie Meze and Dr. Louise Chang - Thank you for bringing me into the cryo-EM world and sharing laughs at our structure. The resources provided by the Microscopy & Image Analysis Laboratory and Cryo-EM facility have been "*instrumental*".

To my committee members Dr. Maria Sandkvist, Dr. Joel Swanson, Dr. Vince Young, and Dr. Melanie Ohi. Thank you for all of your guidance and feedback along the way. Thank you, Maria, for teaching me about pili and advocating for me at conferences. Thank you, Joel, for all of the microscopy help and theatre discussions. Thank you, Vince, for always challenging me to think deeper about the science. Thank you, Mel, for training me, collaborating with me, and letting me join your team as a postdoc. The advice you all have given me has helped me get here.

Last but not least, I would like to thank my mentor, Dr. Nicole Koropatkin. Thank you for always sharing the joy of the research with me from that first day I came in for my interview. You've nurtured me as a person, consoled me when my crystals didn't diffract, rejoiced with me when they did, and always pushed me to do better, think deeper, because you knew I could. I don't know what my future scientific journey holds, but I know I'm a better scientist and person because of you. I couldn't have asked for a better mentor. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

LIST OF ACRONYMS

RS2 Resistant Starch Type 2

CBM Carbohydrate Binding Module

GH Glycoside Hydrolase

ABSTRACT

Starch is a polymer of glucose that is used for energy storage in plants and is a major constituent of the Western diet. Starch that is not digestible by human amylases is termed resistant starch (RS) and acts as a prebiotic that preferentially promotes the growth of specialized beneficial microbes. The consumption of dietary RS has been linked to a lower incidence of colorectal cancer and intestinal inflammation, but it can only be processed by select gut bacteria in the human colon. *Bifidobacterium adolescentis* and *Ruminococcus bromii* are constituents of the healthy gut microbiota that can utilize RS as a sole carbon source and increase in abundance during host RS consumption. However, the genes involved in RS degradation as well as how the encoded proteins synergize to attack this insoluble fiber have not yet been elucidated.

Uncooked starch granules (type 2 RS) are resistant to the action of human amylases due to their supramolecular structure in which adjacent starch chains form double helices which are packed together in crystalline layers, alternating with amorphous layers rich with inter-chain branch points. To access granular starch, bacteria need a way to bind at the granule surface and hydrolyze the glycosidic bonds of the constitutive amylose and amylopectin. These functions can occur on either the same multi-domain protein as in *B. adolescentis* or on multi-protein complexes like the *R. bromii* amylosome. Both human and bacterial amylases contain glycoside hydrolase family 13 (GH13) domains, which catalyze the breakdown of starch. However, bacterial amylases differ from human enzymes in that they encode discrete domains known as carbohydrate binding modules (CBMs). One notable CBM family, CBM74, was first characterized as part of an α-amylase produced by a soil bacterium isolated from a potato starch-processing plant. Recombinantly expressed CBM74 binds to insoluble starch and was shown to be important for maximum granule digestion. In nearly every case, CBM74 is encoded adjacent to a CBM25 or CBM26 starch binding domain. I bioinformatically interrogated the genomes of three RS-degrading gut symbionts and found that each of them had at least one CBM74-containing protein. The CBM74-containing proteins in *R. bromii* and *B. adolescentis* are Sas6 and BaAmy7, respectively.

I first structurally and biochemically characterized CBM74 and the role it plays in conjunction with CBM26 within the context of Sas6 of *R. bromii*. Sas6 does not encode a GH13 domain but acts as a binding protein that can coordinate with amylases through amylosome assembly. I solved the first structure of a CBM74 family member and determined the molecular basis of granular

starch binding. I have determined the protein structure of CBM74, how it binds double helical starch motifs in its elongated binding groove, and how it cooperates with adjacent CBM26 for enzymatic activity on RS2. I then extrapolated these results to determine the role of CBM74 in starch breakdown by enzymatically characterizing BaAmy7 of *B. adolescentis*. I found that BaAmy7 is uniquely able to break down granular potato starch and that this activity is reliant on the CBM74-CBM26 motif encoded as part of its multi-domain architecture. These results support a model in which CBM74 domains target proteins to the crystalline regions of starch granules.

CHAPTER 1

Introduction

There is an old adage, "You are what you eat." In some ways this is absolutely true. Humans are host to a complex consortium of microorganisms, shaped in large part by our diet (152). While we have an intricate immune system to keep our organs largely free of microbes, the gastrointestinal tract is essentially an "outside" tube that runs through the center of the body (50). The gut microbial community is separated from the "inside" components through a layer of intestinal epithelial cells that secrete antimicrobial peptides and make a thick glycoprotein layer to keep beneficial microbes close without allowing them access to the internal barrier (57). While the intestinal microbial community includes fungi, viruses and protists, bacteria are the most numerically abundant organisms and the most well-studied with regards influence on host physiology (103).

The human gastrointestinal tract is thought to be seeded with microbes during childbirth with early microbes being passed from mother to baby (2). The most prevalent first colonizers of the infant gut include Enterococci such as *Enterococcus faecalis* and *Enterococcus faecium*, Enterobacteria such as *Escherichia coli* and Gram-positive Bifidobacteria like *Bifidobacterium infantis* among others (2). These organisms encode special enzymes for breaking down the human milk oligosaccharides that feed these colonizers in the infant host (215; 248). These enzymes are part of a functional group of enzymes known as Carbohydrate Active Enzymes (CAZymes) that include proteins from all domains of life that are active on carbohydrates: sugar polymers consisting of carbon, hydrogen, and oxygen (101).

Throughout human life the changing and developing gut microbiota will continue to be determined by the nutrients that pass through, which dictate the permanent and transient residents (8; 120; 142). The fermentation, or anaerobic metabolism, of dietary carbohydrates available to the gut microbiota generates short chain fatty acids that act as an energy source for our gut epithelium, a signal to our immune system, and even a defense to keep other microbes in check (234; 129). Humans can only encode enzymes to break down three carbohydrates: sucrose (a disaccharide of glucose and fructose), lactose (a disaccharide of glucose and galactose), and polymers of glucose (maltose, maltooligosaccharides, glycogen, and starch) (211). The carbohydrates we can't

break down constitute dietary fiber, a recommended component of a healthy diet to maintain gut health, increase fecal bulk, and control blood sugar (126). In the absence of adequate dietary fiber commensal and pathogenic organisms compete for resources (30). Some microbes utilize a suite of CAZymes that target mucins and break down the protective glycoprotein barrier between the microbiota and intestinal epithelium (49).

While we have long known that fiber is good for us, we still don't have a clear idea of which organisms are breaking down which carbohydrate and how these influence the metabolic ecosystem in the gut (114). It is then imperative that we endeavor to understand: 1) who the "good" bacteria are (probiotics) (75), 2) which carbohydrates feed these organisms (prebiotics) (81; 82), and 3) how we can combine prebiotics and probiotics (synbiotics) to selectively promote the growth of those beneficial microbes (40; 47).

A promising prebiotic currently under investigation is resistant starch (239). Though we encode enzymes to break down starch, the susceptibility of this glucose polymer to host enzyme breakdown is intrinsically tied to how it is synthesized and packaged (19). In its raw, granular form made by plants (admittedly not typically a large part of the human diet), starch is resistant to human amylases but susceptible to a few specialized bacterial amylases, mainly from *Bifidobacterium adolescentis* and *Ruminococcus bromii* (241; 17).

This thesis will introduce resistant starch and its tertiary structure that is recalcitrant to mammalian amylases, the bacterial resistant starch degraders and enzymes they use to break down resistant starch, and will conclude with future perspectives of how resistant starch and its bacterial degraders might be used as a synbiotic for improved gut health.

1.1 RESISTANT STARCH

1.1.1 Starch Structure

Starch is a polymer of glucose that plants use to store energy, analogous to the way mammals store excess glucose as glycogen in the liver and muscles. Starch is produced in the amyloplast of plants as dense granules that feature concentric layers of alternating crystallinity organized around a central hilium (**Fig. 1.1A**), giving starch birefringence under polarized light and generating the typical "maltese cross" image (20). Starch is composed of two polysaccharides: linear amylose composed of almost exclusively α1,4-linked glucose and amylopectin that features α1,6-linked branch points every 15-80 glucose units with differences in average chain length based on botanical source (91) (**Fig. 1.1A**). Starch is called an α-glucan because all of the glucose in starch is α-linked. In contrast β-linked glucose forms cellulose and other β-glucans. The type of linkages between glucose monomers has important implications for the overall structure of starch. In the case of

amylose, the extensive α1,4-linked chains form tightly wound single helices that make amylose largely insoluble in water (159). In the case of amylopectin, the α1,6 branch points create accessible regions for enzymes but also enable adjacent chains of α1,4-linked glucose to be in close enough proximity that they can interact and wrap around one another, forming crystalline double helices (70). Current models suggest that these double helices start to form between chains at a degree of polymerization of 10 residues (170).

In the supramolecular structure, starch granules are formed of alternating crystalline and amorphous layers that are formed by the double helical rich regions (crystalline) and regions of branching (amorphous) (20) (**Fig. 1.1A**). A major difference between starch from different plant sources is the arrangement of those double helices (20). In A type starch, as seen in corn starch granules, the double helices arrange into a trapezoid formation allowing for water molecules to intercalate between the helices (104; 78). In B type starch, as in potato starch, the double helices exhibit a hexagonal arrangement with water only being allowed in the center of the hexagon (105; 78). High amylose corn starch exhibits more of a V type of crystallinity due to the prevalence of amylose-lipid complexes (20; 79). In most starch granules amylose makes up 30% of the granule and current models suggest that amylose is either sequestered in the amorphous layers or threaded through the granule parallel to the amylopectin helices (20; 216). Another difference between starch of different sources that have implications for granular digestibility is the granule size. Corn starch granules for example vary in diameter from 5-50μm with the mode at 20μm (187). In contrast, potato starch granules range in diameter from 10-100μm with the mode at 50μm (187).

The breakdown of starch is mediated by several types of hydrolytic enzymes, amylases which act at α1,4-linkages and pullulanases which act at α1,6-linkages (227). The ability of these enzymes to break down granular starch is mediated by several factors including crystallinity type, granule size, amylose content, and surface structure (52). Due to these factors, the inherent digestibility of corn starch is higher than that of potato starch (53). This is important to keep in mind as we explore the breakdown of starch through the human digestive tract: structural and chemical differences in starch from different plants have distinct effects on the gut microbial community and consequently, digestive health.

Our human salivary and pancreatic amylases digest the majority of the cooked starch that we consume. However, uncooked granular starch is resistant to the activity of these enzymes and is therefore a type of resistant starch (11; 61). But not all resistant starch is created equal. There are four types of resistant starch: RS1 is physically inaccessible due to its packaging into seeds or grains, RS2 is resistant due to its granular semi-crystalline structure, RS3 is retrograded starch that was cooked and cooled, and RS4 is starch that has been chemically modified to be resistant to the action of amylases (21; 63).

1.1.2 RS2 in clinical trials

Colonic bacterial fermentation of RS2 promotes intestinal health and homeostasis (77; 36; 236). When bacteria ferment RS2, they produce short chain fatty acids such as acetate, propionate, and butyrate as end-products (18; 234). Butyrate is a bacterial fermentation end-product of particular interest as it prevents colorectal cancer (32; 100), reduces inflammation (128), and acts as a primary energy source for colonocytes (234). A major complexity of all human microbiome studies is that while population-level trends show that RS2 administration increases fecal butyrate levels, there is significant inter-individual variability (226; 15). This is likely due to differences in how the resident microbial community responds to RS2 breakdown (15). There is therefore a need to understand the mechanisms by which RS2 breakdown is initialized and the cascade of RS2 utilization to design prebiotics that reliably generate a butyrogenic response.

1.1.3 Types of RS2

Though the definition of RS2 as raw granular starch seems straightforward, there is great variability due to the structure of starch granules and how they differ between plant sources. RS2 includes raw granular starch from potato, corn, wheat, rice, and green banana to name a few. RS2 can also refer to granular starch that has been pretreated with digestive amylases to deplete the soluble fraction and enhance the resistant fraction. A commonly used form of RS2 in clinical trials and dietary interventions is raw starch from genetically modified plants such that the granules deviate in the proportion of amylose:amylopectin from the normal 30:70 to nearly 100:0 (high amylose starch) or 0:100 ("waxy" starch) (172). Starch granules with a high amylose content are more resistant to host intestinal amylases (19). Many of the human clinical trials use high amylose maize/corn starch (HACS, commercially known as Hi-Maize 260). For the purposes of this dissertation, in place of RS2, we will specify what source of starch is being investigated and endeavor to be clear about how that starch is pretreated.

1.1.4 Effects of Different Types of RS2 on Gut Community

In a dietary intervention study of 20 healthy individuals consuming raw potato starch (RPS), the increase in butyrate was variable among individuals (226). Those who exhibited an increase in butyrate had an increase in relative abundance of either *Bifidobacterium adolescentis* or *Ruminococcus bromii* (226; 15). The inter-individual variability of responses in these studies supports a need to better understand the inner workings of starch digestion by these species and how they modulate the rest of the community for desired health outcomes. The increase in either *B. adolescentis* or *R. bromii* upon RPS administration has been recapitulated in other human cohorts (68) and in a model

colon with human fecal sample inocula (178). Different starches differently influence the diversity of the microbiota community with a larger increase in relative abundance in *B. adolescentis* on RPS and a larger increase in *R. bromii* on HACS among other differentially responding species (178; 15).

Together, *B. adolescentis* and *R. bromii* encode specialized protein machinery for breaking down RS and are known as "primary degraders" of RS. These species initialize a cascade that promotes the growth of butyrate producers via crossfeeding. The mechanisms by which the primary degraders crossfeed butyrate producers may include modifications to the starch granule surface, generation of short chain fatty acids that can be utilized by other species, and release of sugar byproducts (maltooligosaccharides) that can be scavenged by nearby amylolytic species (221; 17; 183; 56). *R. bromii* in particular is a poor competitor for the sugar byproducts resulting in an outsized influence on the nearby community, qualifying it as a keystone species in the breakdown of RS within the gut ecosystem (241).

1.1.4.1 Ruminococcus bromii

Ruminococcus bromii, named for Bromius, the god of alcohol, is a Gram-positive anaerobic coccoid species that has strict nutritional requirements (158). *R. bromii* produces acetate, formate, and ethanol as byproducts of fermentation (158). It is an amylolytic bacterium that breaks down starch and its constituents but grows little if at all on glucose (241). The lack of glucose import machinery makes it uniquely capable of leaving leftovers for neighboring gut microbes to feed on (17; 160). This is likely what makes *R. bromii* a keystone species in the breakdown of resistant starch in the gut (241). The increase in relative abundance of *R. bromii* upon RS2 administration in the human diet frequently co-occurs with an increase in *Eubacterium rectale*, a starch utilizer and butyrate producer (15; 132). Though *E. rectale* is unable to grow on potato, corn, or HACS as a sole carbon source, it grows well in co-culture with *R. bromii*, highlighting that either the byproducts of *R. bromii* or the initiation of starch granule breakdown can facilitate the growth of *E. rectale* (241).

1.1.4.2 Bifidobacterium adolescentis

Bifidobacterium adolescentis consistently appears as a primary responder when RS2 is administered in the diet, both with potato and high amylose corn starch though it seems to respond to a higher degree to potato starch (68; 178; 15; 132; 226). The ability of *B. adolescentis* to adhere to and grow on raw starch in pure culture is strain specific and includes 22L, P2P3, U269, and L2-32, the latter of which is investigated in Chapter 3 (43; 17; 110). *B. adolescentis* produces acetate and lactate as end-products of fermentation which can aid in starch crossfeeding with other nearby

bacteria including butyrate producers (17; 183). The ability of *B. adolescentis* strains to degrade RS2 is likely determined by the 5-7 extracellular cell-bound amylases encoded by select strains (139; 111). Several of these amylases are upregulated when *B. adolescentis* 22L is grown on starch over glucose (58) and are differentially encoded in strains that adhere to starch granules (111).

1.2 COORDINATED STARCH BREAKDOWN SYSTEMS

1.2.1 Glycoside Hydrolases

CAZymes include enzymes that break down, build, or modify carbohydrates. The focus of this thesis is the class of enzymes that break down polysaccharides by hydrolysis, called glycoside hydrolases (GHs). GHs are classified into families based on their sequence similarity and their target substrates (204). Since starch is a polymer of mostly α1,4-linked glucose, it is known as an a α-glucan. The family primarily responsible for starch breakdown is GH family 13 (GH13), though several other families break down α-glucans including GH31, GH97, GH4, GH14, GH15, GH57, GH63, GH119, and GH126 (107; 231). These GH families are further classified into subfamilies that inform specific substrate motifs recognized and products generated (204). These subfamilies are listed after the family, i.e. GH13_28 would indicate a Glycoside Hydrolase family 13 subfamily 28 which are abundantly encoded in Bifidobacterial α-amylases (237; 242). These classifications will be important as GH13 is the largest family of glycoside hydrolases and we will be delving into several different enzymes in this thesis that belong to different subfamilies and differ slightly in how starch substrates are bound and coordinated within the active site (204).

1.2.2 Human Starch Degradation

Starch is a predominant component of nearly every human diet and is the only polysaccharide that can be broken down by human digestive enzymes. Starch digestion begins in the mouth with the action of salivary amylases (**Fig. 1.1B**). These α-amylases break down the α1,4-linkages of starch revealed during mastication. Salivary amylases (GH13_24) are encoded by the AMY1 gene (192). The amount of salivary α-amylase produced is proportional to AMY1 copy number, which can vary in the human genome, typically between 2-17 copies (192; 245). Several studies have investigated the role copy number variation may have on metrics such as BMI, insulin resistance, glycemic response, and others with inconsistent results (201). The process proceeds to the stomach where the low pH environment denatures salivary α-amylases (188). An additional caveat is that starch is readily hydrolyzed by acid hydrolysis at low pH. This effect targets mostly amorphous regions of starch, leaving intact double helical or crystalline starch (230). The extent of degradation at this

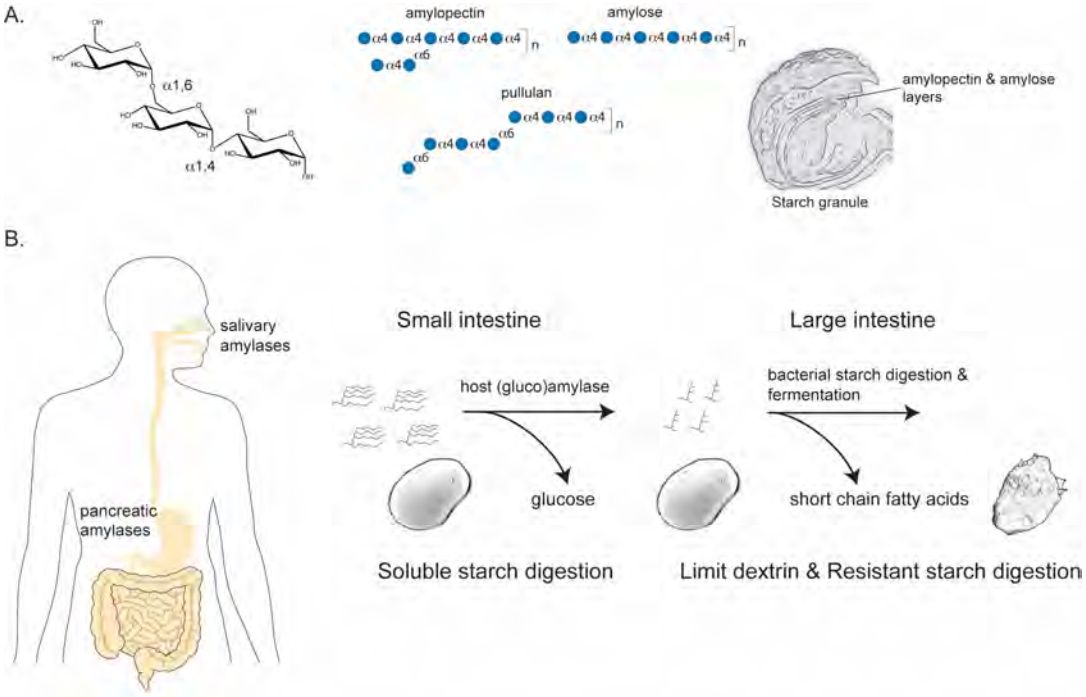


Figure 1.1: Model of starch breakdown in the gut

(**A**) Starch is composed of α1,4 linear and α1,6 branched glucose linkages. Starch granules, often resistant to host digestion, are composed of concentric rings of amorphous amylopectin and tightly packed helices of amylose. (**B**) Starch degradation is initiated in the oral cavity by the action of salivary amylases. Digestion continues via pancreatic amylases and small intestinal brush-border glucoamylases. The liberated glucose is absorbed through enterocytes in the small intestine. The remaining resistant starch (RS), as α-limit dextrin or starch granules, is degraded in the colon by specialized bacterial species which release short-chain fatty acids for the host. Modified from (36).

stage can have a profound effect on the portion of starch that eventually reaches the colon (53).

Following gastric emptying, starch transits to the duodenum of the small intestine where a flush of digestive enzymes is released from the pancreas including pancreatic α-amylase (GH13_24) (245). Pancreatic α-amylase continues the process of starch breakdown into short maltooligosaccharides and disaccharides that are further processed to glucose which is imported by intestinal epithelial cells (191). Because pancreatic α-amylase only hydrolyzes α1,4-linkages, it leaves behind limit dextrins, the highly branched cores of amylopectin, that can traverse the large intestine as food for bacteria (36).

At the brush border of the intestinal villi, mainly in the jejunum, enterocytes produce sucrase-isomaltase (SI), a dual-catalytic GH31-GH31 enzyme that gets cleaved by trypsin once it is secreted into the lumen (191; 211). Single nucleotide polymorphisms in the SI gene are associated with irritable bowel syndrome (IBS) and decreased efficacy of the main dietary intervention used to alleviate IBS, the low-FODMAP (fermentable oligo-, di-, monosaccharide, and polyols) diet (246). This implicates a connection between starch breakdown and digestive symptoms and is an active area of research. The other enzyme active at the brush border of intestinal enterocytes is maltase-glucoamylase (162). Like SI, it is made as a dual-catalytic precursor (162). Glucoamylase and maltase hydrolyze α1,4-linkages to produce glucose while isomaltase hydrolyzes α1,6-linkages in limit dextrins and isomaltase, a low-caloric sweetener (191) (**Fig. 1.1B**). These enzymes break down accessible starch to glucose that can be imported through glucose transporters in enterocyte membranes (191). The fraction of starch that bypasses this process is the resistant starch fraction and transits to the colon where it can be used by starch-degrading bacteria for fermentation (**Fig. 1.1B**) (62).

1.2.3 Bacterial Starch Degradation Systems

Bacterial α-amylases use granular starch far beyond the extent of human α-amylases but what molecular features of bacterial amylases set them apart from those of the upper digestive tract? Simply encoding GH13 genes is not selective to only those gut bacteria able to utilize RS2. GH13-containing proteins are ubiquitous in gut bacteria with numbers of GH13 genes ranging from just a few to up to 40 (Ceratobasidium sp. AG-Ba JN) (www.cazy.org) (54). It may be that certain GH13 subfamilies are structurally primed to take on granular starch or that the additional surface starch-binding site on select GH13s assists with RS2 activity. These surface starch binding sites are usually comprised of a cluster of residues on the catalytic domain, but distinct from the active site, that bind soluble or granular starch substrates (38). Additionally, many bacterial amylases encode discrete starch binding domains remote from the GH13 (106).

1.2.3.1 Carbohydrate Binding Modules

Unlike mammalian amylases, bacterial amylases are frequently multi-modular proteins with ancillary modules specific for substrate binding, called carbohydrate binding modules (CBMs) (25). CBMs are divided into families by sequence that dictate their substrate specificity. At the time of this writing, there are 98 known CBM families, 16 of which bind starch (www.cazy.org) (54; 106). These auxiliary domains are typically around 100 amino acids by definition must be appended to enzymes. There are four proposed roles for starch-binding CBMs (87): 1) Targeting the enzyme to the interfacial surface starch granules (promoting enzyme-substrate productive binding); 2) Assisting the catalytic domain with substrate hydrolysis and facilitating the next catalytic event; 3) Structural disruption of the substrate; 4) Anchoring proteins to the cell surface (25; 87).

An enzyme can have more than one CBM and there are some families which more commonly associate or exhibit tandem repeats (106). While some bacteria encode CBMs on the same polypeptide as amylases, some bacteria make non-catalytic starch binding proteins that can then assemble with amylases in trans (240). The common thread among these systems is coordination of catalytic domains with starch-specific CBMs that allows flexible recognition of different aspects of the starch granular structure.

In this chapter we will briefly introduce some of the paradigms of how starch binding and starch hydrolysis can be coordinated in different bacterial systems. Then we will take a deep dive into the molecular features of some of the starch-binding CBMs that will be covered in this thesis. We will conclude with some specific examples of how bacteria use these starch binding domains to access starch.

1.2.3.2 Coordinated PULs – Cell membrane bound and secreted

In Gram negative Bacteroides species, the degradation of starch is a highly coordinated affair (146). The degradation of starch in *Bacteroides thetaiotaomicron*, as a prototypical system for carbohydrate degradation, is mediated by a co-regulated eight gene cluster known as a polysaccharide utilization locus (PUL) (22). The starch utilization system (Sus) was the first of these PULs to be described and characterized (69; 210). The locus is made up of *susRABCDEFG* (210) (**Fig. 1.2A**). SusR is a sensor/regulator positioned in the inner membrane that senses the presence of maltose and upregulates the rest of the sus genes (48). The starch breakdown process begins at the cell surface with four lipoproteins, SusD, SusE, SusF, and SusG, that are tethered to the cell membrane like balloons by a flexible linker (37; 181). SusE and SusF are dedicated starch binding proteins that have multiple CBM-like starch-binding domains (31). They are largely immobile in the membrane (219). Hydrolysis of starch substrates is catalyzed by the outer membrane α-amylase, SusG. SusG has a GH13_36 that hydrolyzes α1,4-linkages, but can accommodate α1,6-linkages in its active site

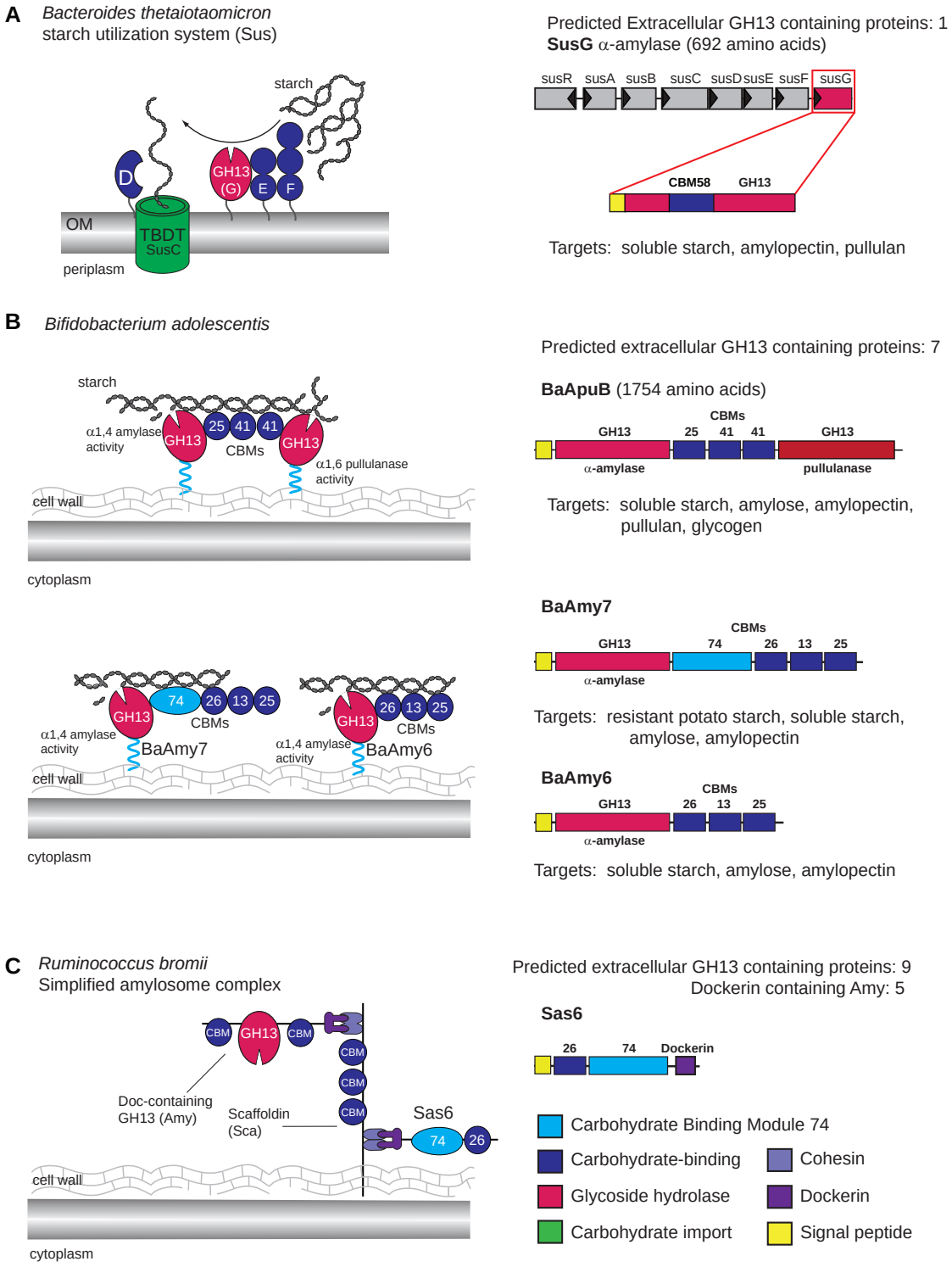


Figure 1.2: Starch breakdown paradigms

A. Starch Utilization System of *Bacteroides thetaiotaomicron*. **B.** Multi-modular enzymes encoded by *Bifidobacterium adolescentis*. BaAmy7 contains a CBM74. **C.** Example arrangement of *Ruminococcus bromii* amylosome. Sas6 is a CBM74-containing putative amylosome component. Modified from (36).

(121). SusG has a starch-binding CBM58 that interrupts the GH13 fold and structurally extends out from the GH13 (121). This CBM58 is critical for binding insoluble corn starch but not soluble starch or amylopectin (121). In contrast to SusE and SusF, SusG is mobile, roaming the cell surface until it encounters substrate and pauses (219). The products of SusG hydrolysis are scavenged by starch-binding protein SusD which then threads these maltooligosaccharides through the β-barrel of SusC, the TonB-dependent transporter that spans the outer membrane (85; 180). Transport is energized by pairing of a periplasmic domain of SusC to one of the 12 TonB proteins encoded by *B.thetaiotaomicron* (175). These are encoded throughout the genome and are not co-regulated with the sus locus (24). The cooperation between SusC-like and SusD-like proteins is highly conserved among all PULs regardless of substrate and the *susC/D* gene pair signature defines a PUL in Bacteroides genomes (213). Once maltooligosaccharides arrive in the periplasm, they are acted on by a neopullulanase, SusA (GH13_46) and α-glucosidase, SusB (GH97) which break down imported maltooligosaccharides to glucose (144; 37). *B. thetaiotaomicron* and other Bacteroides species co-regulate the production of distinct proteins for binding, breakdown, and import which interact at the cell surface for coordinated starch capture and utilization (69).

1.2.3.3 Multiple multi-modular cell surface enzymes

The most common form of coordination between starch binding and catalytic functions is to encode dedicated domains for each task together on the same polypeptide. Many bacteria encode multi-modular amylases that have a GH13 either at the N-terminus or C-terminus of the protein and encode one or more tandem CBMs along the length of the protein (41; 14) (**Fig. 1.2B**). A typical GH13 is about 500 amino acids and some of these amylases can be up to 1800 amino acids. With CBMs typically spanning about 100 amino acids, the size of these large amylases highlights the additional binding power that can be encoded along the polypeptide (87).

Another form of muti-modularity is the coordination of dual catalytic capabilities on the same enzyme. Within the GH13 family, some subfamilies target α1,4-linkages in starch (α-amylases), while others hydrolyze α1,6-branch points (pullulanases), and some accommodate both linkages (amylopullulanases). Some multi-modular enzymes encode two separate GH13s for hydrolyzing different linkages on the same polypeptide (36). This is the case for the Bifidobacterial α-amylase-pullulanase, ApuB. ApuB has a GH13_15 (amylase) at the N-terminus of the protein and a GH13_13 (pullulanase) at the C-terminus with dedicated starch-binding CBMs in between, which may allow it to bind at a branch point and clip the α1,6-linkage and a proximal α1,4-linkage in the same attachment event (167) (**Fig. 1.2B**).

It may be tempting to assume that once the different domains have been characterized with the help of AlphaFold for structural predictions that we understand how a multi-modular enzyme will work together as a catalytic machine. However, between these domains there are inert link-

ers, both structured and flexible, that generate some level of unpredictable wobble in the domain arrangement in space. There can be extended flexible linkers that allow for expansion and contraction between the domains upon substrate binding (35) or there can be stable linkers, typically annotated as Fibronectin-like III or Bacterial Ig-like domains, that determine the domain spacing and arrangement (222; 136; 73). These domains are not currently thought to participate directly in binding but have been shown to potentiate enzymatic activity and are speculated to participate in protein-protein interactions that may allow for multi-modular enzymes to work together at the cell surface (197). In one elegant example, a crystal structure and corresponding solution structure of a *Streptococcus pneumoniae* glycogen-degrading enzyme were solved and it was observed that two tandem N-terminal starch-binding CBM41s act as a dyad to bind starch and that the protein undergoes a conformational change that brings the dyad into alignment with the catalytic domain to form an extended catalytic site (125). This folding over of the CBMs has been described as a hammer descending on an anvil with a linker domain between them acting as a handle for the hammer (125). We will briefly touch on the proposed role of these linker domains in the proteins featured in Chapters 2 and 3. As we come to understand and dissect the domain architecture of multi-modular amylases, we still have much to learn about how the linkers between the domains allow the overall protein to fold into a molecular machine.

1.2.3.4 Amylosomes

A 2015 paper from Harry Flint's group first established the concept of the "amylosome", so named because it features an extracellular multi-protein assembly analogous to the cellulosome that enables degradation of cellulose (200; 240). The typical cellulosome architecture consists of one or more scaffoldin proteins, either secreted freely or cell surface attached, that encode discrete domains called cohesins (200) (**Fig. 1.2C**). Cohesins bind with high affinity to cognate domains called dockerins on other proteins (200). These cohesin-dockerin interactions are calcium mediated and involve extensive hydrogen bonding and hydrophobic interactions (163). This system allows for the assembly of dockerin-containing proteins with different starch binding and catalytic roles on a single scaffoldin protein.

R. bromii encodes nine extracellular GH13-containing proteins (Amys) based on the presence of a predicted signal peptide (240; 160). Seven of the nine are multi-modular with appended CBMs. Of those nine, five contain dockerin domains and are predicted to assemble into amylosomes: Amy 4, Amy 9, Amy10, Amy12, and Amy16 (160). We use the term amylosomes because there are five different scaffoldins (Sca proteins) to which these amylases may assemble, two of which are cell wall anchored (240). There is promiscuity in the cohesin-dockerin binding between most of the amylosome constituents meaning there are many different iterations of how these amylosomes can assemble (160). Sca1, or Amy4,has both a cohesin and a dockerin domain in addition to

the GH13_19 (240). The Sca1/Amy4 cohesin serves as an anchor for Amy9, Amy10, Amy16 as well as its own dockerin, leading to the potential for self-multimerization (240; 160). In this way Sca1/Amy4 may serve as a building block to multiply the possible binding interactions. Sca2 is one of the two cell wall anchored scaffoldins and its cohesin is bound by Amy4 and Amy9, which means by the transitive property of Amy4, it can also accommodate Amy10 and Amy16 (240; 160). The Sca3 and Sca4 cohesins are both also bound by the Amy4 and Amy9 dockerins recapitulating the possibility for multimeric assembly but are also bound by the dockerin domain from Amy16 (240; 160). It is unclear why the Amy16 dockerin is more selective than the dockerin domains from the other Amy proteins but all of them can bind to Sca3 and Sca4 (240). The last scaffoldin, Sca5 is cell-bound much like Sca2 (240). However, Sca5 is unique in that it has two cohesin domains (160). This presents an opportunity for two dockerin-containing proteins to bind adjacent to one another on the same polypeptide. The Sca5 cohesin binds the dockerin domains from all the Amys, except Amy16 (160).

In addition to the Amys, there are 22 other dockerin-containing proteins in the *R. bromii* genome that likely assemble by way of the six cohesin domains as well (35; 240). Some of these are important starch binding proteins, like Sas20, Sas6, and Sas22 so named for their roles in the starch adherence system (SAS) of the amylosome (35). These three proteins were identified in two separate proteomic studies as abundant in starch-grown *R. bromii* cells (240; 35). Sas20 is a multi-domain protein that binds raw starch by recognition of non-reducing chain ends by its first domain while the second domain binds the curvature of elongated maltooligosaccharides (35). Together, Sas20 likely specializes in binding the unpacked chains at the granule surface (35). Chapter 2 of this thesis will explore the structural basis of starch binding by Sas6. Beyond those proteins that have been characterized so far, there remain many other components of *R. bromii* amylosomes still to be characterized and much still to understand about how the arrangement of proteins upon binding to cohesins contributes to the overall function of amylosomes as molecular machines.

1.3 CARBOHYDRATE BINDING MODULES

1.3.1 CBM Structures

At their core, most CBMs exhibit a common β-sandwich fold (25). They are further classified into types based on the architecture of their binding sites (87; 25). Type A CBMs have a flat-hydrophobic surface for interacting with linear crystalline substrates like cellulose or chitin (25). Type B CBMs encompass most starch-binding CBMs (25). These CBMs have a binding cleft set into the surface for binding single polysaccharide chains like maltooligosaccharides, amylopectin, or amylose (25). Type C CBMs only bind mono-, di-, or tri-saccharides through the interaction at

a single sugar residue (25).

The binding sites of starch-binding CBMs share some conserved similarities. Typically, one or more aromatic residues (Trp, Tyr, Phe, His) provide CH-π stacking with the pyranose ring of glucose residues, resulting in strong van der Waals forces (87; 203). Starch binding sites are commonly made up of two aromatic residues oriented in such an angle from one another that they create a cradle for the curvature of the α1,4 linkage in maltooligosaccharides (106). The distance between these residues and their geometry on the protein surface can determine the affinity of the overall CBM for different starch substrates(106). Additional ligand binding contributions are provided by polar side chains around the binding site created by the aromatic residues. These provide hydrogen bonding to the hydroxyls of the glucose moieties (106). These polar side chains can also determine the ligand specificity as they protrude into the binding site and can select for maltooligosaccharide geometry or length (87). The third ligand binding determinant provided by the binding site is steric hindrance. Even within a CBM family, there can be a wide variation in protein loops near the binding site that can block the binding of one ligand but offer an ideal cleft for another (106). Thus, each new CBM structure provides insight into how a CBM family can bind various ligands or set a preference for one over another. A current review of the structurally determined starch-binding CBMs elucidates the structural similarities and differences among and between the different starch-binding CBM families (106).

1.3.2 CBM25

The first biochemical characterization of CBM25 as a discrete CBM was as part of α-amylase BAA from *Bacillus* sp. no 195 (207). The overall schematic of this amylase includes a GH13 followed by 2 tandem repeat CBM25s (GH13/CBM25-1/CBM25-2). The authors purified naturally occurring truncations of BAA that correspond to ΔCBM25-2 or ΔCBM25-1/2. The degree of binding to raw corn starch was decreased with the loss of each CBM25, resulting in no binding when both were deleted (207). The CBM25s bind raw starches from corn, wheat, rice, potato, and sweet potato, unlocking them as enzyme substrates without altering the kinetic parameters on soluble starch or a maltopentaose analog. Only the full-length enzyme has activity higher than porcine pancreatic amylase, a marker of a resistant starch degrading enzyme (207).

These results make it especially striking that when BAA is isolated from the culture supernatant of Bacillus sp. 195 it is proteolytically processed so that CBM25-2 is cleaved from the enzyme (207). This phenomenon was characterized in the supernatant, which leads to the alternative hypothesis that BAA retains the CBM25s when it is anchored at the cell surface and that the truncated isoforms are a feature of secreted BAA. The characterization of BAA and its CBM25 truncations highlights that CBM25s are raw starch binding domains and that they confer the ability

of an enzyme to access raw starch substrates (207).

The molecular basis for CBM25 binding was determined by the biochemical and structural characterization of a tandem CBM25 and CBM26 combination from a maltohexaose-forming amylase produced by Bacillus halodurans C-125 (26). CBM25s are B type CBMs with a central β-sandwich typical of starch-binding CBMs (106; 87). The binding site is composed of two tryptophan residues that form an aromatic platform with side chains from nearby loops providing hydrogen bonds to the ligand (26). CBM25s are commonly encoded as tandem repeats or adjacent to a CBM26 (106).

CBM25s have been used to target engineered proteins to the starch granule within the amyloplast during granule synthesis due to its ability to bind both amylose and amylopectin (102). Tandem CBM25s from the enzyme *Ma*AmyA are sufficient to give the enzyme the ability to form pores in the surface of wheat starch granules (221). Based on the roles of CBMs enumerated above, it is likely that these domains keep the enzyme anchored at a single region of catalysis. In this way, when the GH13 hydrolyzes the linkage it is working on, the enzyme doesn't completely release and reattach at a different site on the granule surface (216).

1.3.3 CBM26

Lactobacillus amylovorus Amy13A has a GH13_28 followed by 5 tandem repeats of CBM26s at the C-terminus (184). These CBMs contribute to stability across temperature and pH, but are required for binding raw corn starch (184). A follow-up study dissected the contribution of each of the CBM26s to raw starch binding and quantified the synergism between the domains (88). It was shown that the binding of more than two CBMs enhanced the binding of the next one but that likely in the context of the full enzyme all five do not bind concurrently (88).

The tandem repeat pattern observed in *B. halodurans*, *Butyrivibrio fibrisolvens*, and *L. amylovorus* is common for CBM26s (140). It is this avidity effect or synergism that likely gives CBM26-containing proteins the ability to bind raw starch (26). However, the presence of CBM26s does not appear to be an exclusive marker to raw starch degrading bacteria (www.cazy.org) (54). For example, *E. rectale* also has a cell surface amylase with five CBMs, two of which are tandem CBM26s, but the organism is unable to utilize raw corn or potato starch as a sole carbon source (17; 39). However, CBM26s frequently co-occur with CBM74s, domains that might be markers for raw starch degrading enzymes (223).

1.3.4 CBM74

CBM74 is a starch-binding CBM that was originally characterized as a discreet domain of a raw starch degrading amylase from the bacterium, *Microbacterium aurum* (193). *M. aurum* was

isolated from the waste sludge of a potato starch processing facility so it may be no surprise that *Ma*CBM74 binds raw potato, corn, and wheat starches as well as starch components, amylose and amylopectin (193; 223). *Ma*CBM74 has a higher affinity for potato derived starch over that from wheat and corn suggesting that the *Ma*CBM74 has a higher affinity for starches with A vs B-type crystallinity (223). The other CBM74-containing enzyme that has been characterized is a maltopentaose-forming amylase from an uncultured alkaliphilic bacterium (33). Deletion of the C-terminus of this enzyme eliminates binding to raw starch, though catalytic activity on raw starch was not compared to the full-length enzyme(33). The C-terminal region included a CBM26, 3 tandem repeats of CBM25, and a CBM74 separated by FN3 domains (33).

Aside from starch binding, CBM74 may play an additional role in raw starch degradation. The deletion of the CBM74 of MaAmyA resulted in smaller, but not fewer, pores on starch granules as shown by Scanning Electron Microscopy (223). This suggests that CBM74 may contribute to local polysaccharide chain disruption or may keep the enzyme localized to areas that have already been hydrolyzed.

CBM74 doesn't fit the paradigm of CBMs in that it consists of 300-350 amino acids as opposed to the canonical 100-150 amino acids (106). Prior to the current work, no CBM74 three-dimensional structure had yet been determined, so the molecular basis of binding was unclear (see Chapter 2. The most similar structure is that of CBM9 of a xylanase from *Thermotoga maritima* MSB8 (223). A tryptophan residue of CBM9 is conserved among all identified CBM74s and may be involved in starch binding (164; 223).

CBM74 is a starch-binding CBM and as such, when it is part of an enzyme the catalytic domain belongs to the GH13 family, predominantly subfamilies GH13_19, GH13_28, and GH13_32 (54). In some cases, CBM74 is not appended to any GH13 (223; 54). For example, Sas6 from *Ruminococcus bromii* has a CBM26, a CBM74, and a dockerin for assembly into amylosomes for coordinated starch adherence and breakdown (see Chapter 2).

In nearly every instance of a CBM74 it is encoded adjacent to a CBM25 or CBM26 (223). It is not currently known if CBM74 can participate in synergistic binding with these domains as observed when they are encoded as tandem repeats or if the co-existence of CBM74 and CBM25/26 opens a new mode of binding for CBM74-containing proteins.

CBM74 has potential as a marker for RS2-degrading bacteria. Genomes containing CBM74s increased when human fecal inocula were grown in a model colon with potato starch or high amylose corn starch as substrates (178). These newly sequenced genomes that reflected an increase on potato and high amylose corn starches each had a single CBM74 gene (178). In a metatranscriptomic study of pigs administered raw potato starch there is an increase in transcripts corresponding to CBM74, CBM21, and GH13s (238). The two established primary degraders of RS2, *B. adolescentis* and *R. bromii*, each encode a CBM74-containing protein (cazy.org) (54).

1.4 Conclusion

Starch is a polymer of a single monosaccharide, but has elaborate complexity conferred by the linkages between the glucose residues and the packaging of amylose and amylopectin into insoluble granules. Adjacent chains of glucose that extend from an α1,6 branch point have the propensity to form double helices. These helices are then packed together in crystalline layers blocking the action of most amylases. Only a few specialized gut microbes have the protein machinery to initiate RS2 breakdown but when they do, they provide a foothold for other beneficial members of the gut microbiota.

When searching for a molecular pattern that differentiates RS2 degraders, it is prudent to start with the starch-binding domains that set apart microbial amylases from human ones. The auxiliary binding domains are known as carbohydrate binding modules. CBM25, CBM26, and CBM74 are among the families of CBM that target granular starch (106). In particular, CBM74 stands out as a behemoth CBM that specializes in granular starch binding but has not yet been structurally characterized. CBM74s are encoded by each of the identified RS2-degrading bacterial species and strains from the gut microbiota so far and are not widely encoded among non-RS2 utilizing species as is the case for CBM25 and CBM26. CBM74 therefore may be a molecular marker for RS2 degrading ability in the gut. This thesis therefore seeks to understand the structural basis for why CBM74 is so unique among starch-binding CBMs.

1.5 Dissertation Outline

In light of the unique features of CBM74, this dissertation aims to investigate the molecular basis for starch binding by CBM74s and their significance in the process of starch degradation within gut bacteria.

In Chapter 2, we present the first 3D structure of a CBM74 family member, Sas6, found in *R. bromii*. Sas6 is a crucial component of the starch adherence system in *R. bromii* responsible for starch capture. It consists of an N-terminal CBM26 adjacent to a CBM74, with a dockerin domain at the C-terminus, facilitating its incorporation into *R. bromii* amylosomes. Our evaluation includes a crystal structure of Sas6 with a ligand bound to CBM26, as well as a structure of CBM74 with maltodecaose (a maltooligosaccharide with a degree of polymerization of 10) bound in a double helical conformation.

The arrangement of Sas6 reveals that the CBM26 and CBM74 binding sites are in close proximity, with limited conformational flexibility. Our findings indicate that CBM26 binds maltooligosaccharides consisting of at least 4 glucose residues, while CBM74 possesses an elongated binding groove along one face of the domain, requiring a ligand of at least 8 glucose residues with a preference for

double helices. By conducting separate binding assays for each domain, we observe that CBM74 primarily drives binding to granular starch, whereas CBM26 exhibits affinity for short oligosaccharides and amylopectin. To provide a comprehensive analysis of starch binding by CBM74, we compare the residues involved across all CBM74 family members, including those found in Bifidobacteria.

In Chapter 3, we extend our findings from Chapter 2 to explore the potential role of CBM74 in the breakdown of granular starch. Specifically, we focus on the CBM74-containing protein, BaAmy7, derived from *B. adolescentis*. BaAmy7 is an extracellular, multi-modular enzyme that resembles Sas6, as it features the domain motif CBM74-CBM26. Additionally, there are two other predicted CBMs, CBM13 and CBM25, positioned at the C-terminus, though only CBM25 is expected to specifically interact with starch. Surprisingly, our findings reveal that the C-terminal CBMs do not exhibit starch-binding capacity. Instead, it is CBM74 and CBM26 that are responsible for starch binding and are crucial for enzyme activity on granular potato starch. These results shed light on the initial stages of the RS2 degradation cascade at the cell surface of *B. adolescentis*.

Chapter 4 synthesizes the results of chapters 2 and 3 and presents a new model of the role of CBM74-containing proteins in granular starch capture and breakdown.

CHAPTER 2

The *Ruminococcus bromii* Amylosome Protein Sas6 Binds Single and Double Helical α-Glucan Structures in Starch

The contents of this chapter have been published as: Amanda L. Photenhauer, Rosendo Villafuerte-Vega, Filipe M. Cerqueira, Krista M. Armbruster, Filip Mareček, Tiantian Chen, Zdzislaw Wawrzak, Jesse B. Hopkins, Craig W. Vander Kooi, Štefan Janeček, Brandon T. Ruotolo, Nicole M. Koropatkin **The *Ruminococcus bromii* amylosome protein Sas6 binds single and double helical α-glucan structures in starch**; *Nature Structural and Molecular Biology* 2023

Krista M. Armbruster performed the Sas6 Western Blot that helped us answer the question: "Is Sas6 bound at the cell surface or secreted into the supernatant?"

Zdzislaw Wawrzak is a beamline scientist at Argonne National Laboratories and assisted with X-ray data collection from Sas6 protein crystals and processed the data to generate the initial electron density maps.

Filipe M. Cerqueira and Jesse B Hopkins performed the small angle X-ray scattering experiments that generated a low-resolution molecular envelope of Sas6 in solution. These experiments allowed us to determine the conformational variability Sas6 takes on in solution and how it compares to the crystal structure.

Tiantian Chen and Craig W. Vander Kooi performed hydrogen-deuterium exchange to map Sas6 peptides that are protected from deuterium exchange by ligand binding.

Rosendo Villafuerte-Vega and Brandon T. Ruotolo performed the native mass spectrometry experiments that allowed us to distinguish the mass/charge difference of Sas6 constructs bound to one or more molecules of maltodecaose. This allowed us to determine the proportion of molecules that are bound in a single helical or double helical conformation across increasing concentrations of ligand.

Filip Mareček and Štefan Janeček compared *Rb*CBM74 to 99 other CBM74 sequences and created a phylogenetic tree to group them by similarity.

2.1 ABSTRACT

Resistant starch is a prebiotic accessed by gut bacteria with specialized amylases and starch-binding proteins. The human gut symbiont *Ruminococcus bromii* expresses Sas6 (Starch Adherence System member 6) that consists of two starch-specific carbohydrate binding modules from family 26 (*Rb*CBM26) and family 74 (*Rb*CBM74). Here we present the crystal structures of Sas6 and of *Rb*CBM74 with a double helical dimer of maltodecaose bound. The *Rb*CBM74 starch-binding groove complements the double helical α-glucan geometry of amylopectin, suggesting this module selects this feature in starch granules. Isothermal titration calorimetry and native mass spectrometry demonstrate that *Rb*CBM74 recognizes longer single and double helical α-glucans, while *Rb*CBM26 binds short maltooligosaccharides. Bioinformatic analysis supports the conservation of the amylopectin-targeting platform in CBM74s from resistant-starch degrading bacteria. Our results suggest that *Rb*CBM74 and *Rb*CBM26 within Sas6 recognize discrete aspects of the starch granule, providing molecular insight into how this structure is accommodated by gut bacteria.

2.2 INTRODUCTION

The gut microbiota, the consortium of microbes that resides in the human gastrointestinal tract, influences many aspects of host physiology including digestive health (190). The composition of the gut microbiota is modulated by the human diet (82; 44; 12). After host nutrient absorption in the small intestine, indigestible dietary fiber transits the large intestine and becomes food for gut microbes (44). Bacterial fermentation of dietary carbohydrates produces beneficial short chain fatty acids including butyrate, a primary carbon source for colonocytes that also has systemic anti-inflammatory and anti-tumorigenic properties (82; 12).

Resistant starch, defined as starch that is resistant to digestion in the upper gastrointestinal tract, is a prebiotic fiber that tends to increase butyrate in the large intestine (239). Starch is a glucose polymer composed of amylopectin and amylose which is layered within granules (20; 172). Amylose is a predominantly α1,4-linked glucan with infrequent α1,6 branching, forming the amorphous layers of the granule (20). Amylopectin is an α1,4-linked polymer with α1,6-linked branches that allow the formation of parallel α1,4-glucan chains that form double helical structures that pack together in the crystalline region of the granule (20; 172). The distribution of amylopectin chain lengths varies with two distinct chain length populations, a short chain length (SCL) population with an average of 16-20 residues between branch points and a long chain length population (LCL) with an average of 48-73 residues, depending on botanical source (95). Raw, uncooked starch granules are resistant to host digestion due to their semi-crystalline structure and classified as resistant starch type 2 (RS2) (20). RS2 becomes food for gut bacteria that can adhere to and deconstruct granules,

releasing glucose and maltooligosaccharides that cross-feed other organisms (36).

Human gut bacteria that degrade RS2 *in vitro* include *Bifidobacterium adolescentis* and *Ruminococcus bromii* (241; 110; 212; 58; 17). *R. bromii* is a Gram-positive anaerobe that increases in relative abundance in the gut upon host consumption of resistant potato or corn starch (36; 226; 15). *R. bromii* is a keystone species for RS2 degradation because it cross-feeds butyrate-producing bacteria (241). *R. bromii* synthesizes multi-protein starch-degrading complexes called amylosomes via protein-protein interactions between dockerin and complementary cohesin domains (240; 200; 16). As many as 32 *R. bromii* proteins have predicted cohesin or dockerin domains including amylases, pullulanases, starch-binding proteins, and proteins of unknown function (240; 160). Many have carbohydrate-binding modules (CBMs) that presumably aid in binding starch and tether the bacterium to its food source (106).

CBMs are classified by amino acid sequence into numbered families and include members that bind only soluble starch and some that also bind granular starch (106; 223). One such family is CBM74 which was discovered as a discrete domain (MaCBM74) of a multimodular amylase from the potato starch-degrading bacterium, *Microbacterium aurum* (52). MaCBM74 binds amylose and amylopectin as well as raw wheat, corn, and potato starch granules (223). CBM74s are unique as they are ~300 amino acids, two to three times larger than most starch-binding CBMs (106). CBM74s are typically found in multimodular enzymes that include a glycoside hydrolase family 13 (GH13) for hydrolyzing starch and are flanked by a starch-binding CBM from family 25 or 26 (CBM25 or CBM26) (106; 223). Most CBM74 family members are encoded by gut microbes and 70% are found in Bifidobacteria (223). The genomes of *R. bromii* and *B. adolescentis* each encode one putative CBM74-containing protein. The prevalence of CBM74s encoded within the genomes of RS2-degrading bacteria, and its increased representation in metagenomic and metatranscriptomic analyses from host diet studies, suggest a role for this module in RS2 recognition in the distal gut (52; 178; 238).

The *R. bromii* starch adherence system protein 6 (Sas6) is a secreted protein of 734 amino acids that contains both a CBM26 and CBM74 followed by a C-terminal dockerin type 1 (242; 137). Here we present the biochemical characterization and crystal structure of Sas6, providing the first view of the CBM74 and its juxtaposition with the CBM26. We captured the structure of *Rb*CBM74 with a double helical dimer of maltodecaose, which mimics the architecture of double helical amylopectin in starch granules, revealing that this domain selects for this motif via an elongated binding groove. *Rb*CBM74 exclusively binds longer maltooligosaccharides ($\geq$ 8 glucose units), and native mass spectrometry suggests that both single and double helical α-glucans that adopt the geometry of double helical amylopectin are recognized. Our biochemical data demonstrate that CBM26 and CBM74 recognize different α-glucan moieties within starch granules leading to overall enhanced granule binding.

2.3 RESULTS

2.3.1 Modular Architecture of Sas6

Sas6 consists of five discrete domains: an N-terminal CBM26 (*Rb*CBM26), a CBM74 (*Rb*CBM74) flanked by Bacterial Immunoglobulin-like (BIg) domains, and a C-terminal type I dockerin (**Fig. 2.1A**) (137). Sas6 is encoded at the WP_015523730 locus (formerly RBR_14490 or Doc6, UniProt: A0A2N0UYM2) and includes a Gram-positive signal peptide (residues 1-30) that presumably targets the protein for secretion. *Rb*CBM74 spans residues 242-572 based on an alignment with annotated CBM74s (223). We used InterProScan to annotate the remaining sequence which added the Bacterial Immunoglobulin-like (BIg, Pfam 02368) domain A (BIgA), but did not predict BIgB, which we identified via structure determination (23).

2.3.2 Sas6 Cell Localization

Though Sas6 has a signal peptide it is unknown whether it is a constituent of a cell-bound amylosome, or part of a freely secreted complex (160). *R. bromii* synthesizes five scaffoldin (Sca) proteins that have cohesins for amylosome assembly; Sca2 and Sca5 are cell-bound and Sca1, Sca3, and Sca4 are freely secreted (160). The cognate cohesin for the Sas6 dockerin is unknown. Sas6 is detected in the cell-free supernatant of *R. bromii* cultures in stationary phase but also elutes from the surface of exponentially growing cells with EDTA which disrupts the calcium-dependent cohesin-dockerin interaction (240; 35). To determine the localization of Sas6, we grew cells to mid-log phase on potato amylopectin and performed a Western Blot with custom antibodies against recombinant Sas6 (**Fig. 2.1B**). Sas6 was detected in the cell fraction, not the cell-free culture supernatant and was visualized on the cell surface via immunofluorescence (**Fig. 2.1B,C**). Therefore, we conclude that Sas6 is a component of a cell-surface amylosome in actively growing cells. It is possible that Sas6 localization is dependent upon growth phase, as are cellulosome components in some organisms, explaining its previous detection in culture supernatant (240). Alternatively, *R. bromii*, like some cellulosome-producing bacteria, may release cell-surface amylosomes in stationary phase (26).

2.3.3 Sas6 Binds Starch

CBM26 and CBM74 are putative raw starch-binding families (223; 26). Plant sources of granular starch differ greatly in granule organization, including crystallinity (e.g., packing of the long helical chains), length of α1,4-linked chains, amylose location and organization, water content, and trace elements (20). We used a truncated construct of Sas6 (residues 31-665) lacking the C-terminal dockerin domain and N-terminal signal peptide, herein called Sas6T, to test Sas6 binding to starch

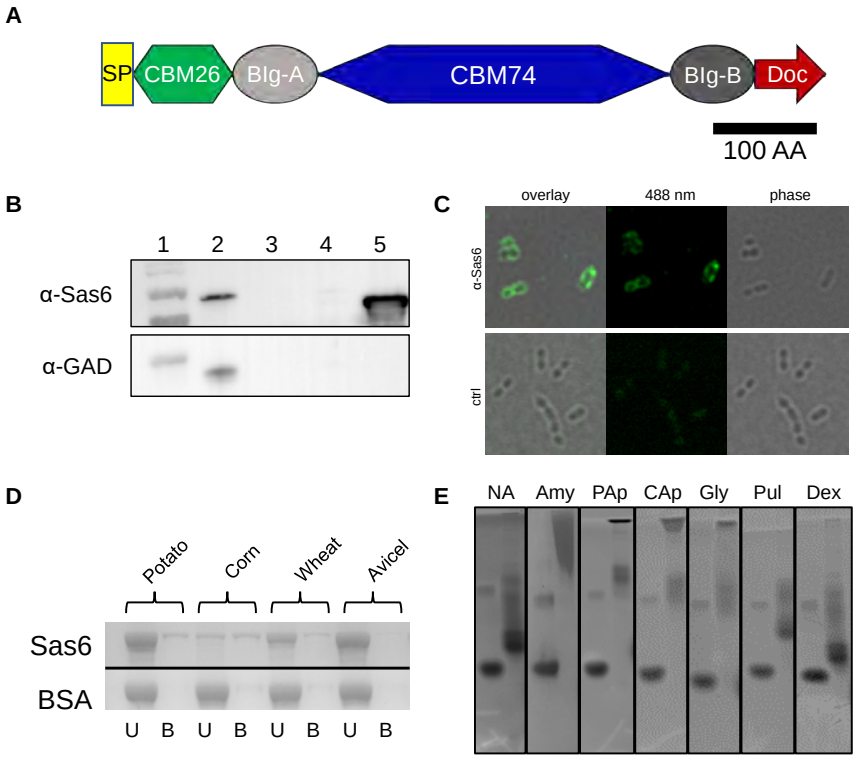


Figure 2.1: *Ruminococcus bromii* Sas6 is a starch-binding protein with two carbohydrate-binding modules. **A.** Domain architecture of Sas6 annotated according to the Carbohydrate Active Enzyme database (www.cazy.org) and the crystal structure. SP = Signal Peptide, CBM26 = Carbohydrate Binding Module family 26, BIg = Bacterial Immunoglobulin, CBM74 = Carbohydrate Binding Module family 74, Doc = Dockerin. Sas6T: Recombinantly expressed truncated version of Sas6 lacking the C-terminal dockerin. **B.** Top: Western blot with anti-Sas6 antibody showing localization of Sas6 in the cell fraction. Bottom: Parallel western blot with custom rabbit antiserum against glutamic acid decarboxylase to control for cell lysis. Lane 1: ladder, 2: *R. bromii* cell lysate, 3: cell-free culture supernatant, 4: TCA precipitated cell-free culture supernatant, 5: recombinant Sas6T. **C.** α-Sas6 immunofluorescent staining of fixed *R. bromii* cells grown in potato amylopectin. "ctrl" represents samples incubated with blocking buffer instead of α-Sas6 primary antiserum. **D.** SDS-PAGE gel from Sas6 adsorption to potato, corn, and wheat starch, and Avicel (cellulose) control. U=unbound protein, B=bound protein. **E.** Affinity PAGE with 0.1% of the indicated polysaccharide incorporated into the gel matrix. For each, left lane is bovine serum albumin, right lane is Sas6T. NA=native gel, Amy=potato amylose, PAp=Potato Amylopectin, CAp=corn amylopectin, Gly=Glycogen, Pul=Pullulan, Dex=Dextran.

polysaccharides. Sas6T binds potato, corn, and wheat starch granules, with the highest fraction of protein bound to corn starch, and no non-specific binding to Avicel (crystalline cellulose) (**Fig. 2.1D**). Of note, corn starch has a smaller granule size and therefore a larger surface area to mass ratio (172). We tested Sas6T binding to amylopectin and amylose, as well as glycogen and pullulan via affinity PAGE. Glycogen is similar to amylopectin with more frequent α1,6 branching, about every 6-15 residues for liver glycogen (148; 29). Pullulan is a fungal α-glucan composed of repeating α1,6-linked maltotriose units (199). Sas6T binds amylose, amylopectin (potato and corn), and glycogen but displays poor recognition of pullulan suggesting a preference for longer α1,4-linked regions within the polysaccharide (**Fig. 2.1E**). Sas6T does not bind dextran, a bacterially derived exopolysaccharide of α1,6-linked glucose (115), demonstrating its specificity for starch.

2.3.4 Sas6 is a Compact Globular Protein

The structure of Sas6T with α-cyclodextrin (ACX), was determined via single-wavelength anomalous dispersion of intrinsic sulfur-containing residues to a resolution of 1.6Å (Rwork=16.8%, Rfree=21.2%) (Table 1). The final model contained two molecules of Sas6T in the asymmetric unit, with four Ca2+ per chain and one molecule of ACX bound at the *Rb*CBM26. The Sas6T structure determined with ACX was used to phase a dataset from unliganded crystals (2.2Å, Rwork=19.7%, Rfree=25.5%) (Table 2.1). The overall crystal structure of Sas6T is compact, with *Rb*CBM26, BIgA and BIgB forming an arc over *Rb*CBM74 (**Fig. 2.2A**). The two chains in the asymmetric unit exhibit some flexibility resulting in different positioning between the *Rb*CBM26 binding site and the *Rb*CBM74 (**Fig. 2.2B**). Solution structure determination by small angle X-ray scattering (SAXS) suggests some conformational flexibility in the Sas6 structure between the *Rb*CBM26 and *Rb*CBM74, but the resulting MultiFoXS model shows that overall structure remains compact (**Fig. 2.2C-I**). This finding is supported by the extensive hydrogen bonding between BIgA (light grey) and BIgB (dark grey) which separate the *Rb*CBM26 and *Rb*CBM74 or *Rb*CBM74 and dockerin domains, respectively. The hydrogen bonding between the loops of BIgA and BIgB generate 354Å of buried surface area (**Fig. 2.2H,K**) (123). Ig-like or fibronectin-III domains act as spacers in multi-modular glycoside hydrolases and provide structural stability (222). In this case, the BIg domains may keep *Rb*CBM26 and *Rb*CBM74 properly oriented.

2.3.5 Unliganded Structure of *Rb*CBM74

*Rb*CBM74 (357 residues) has 21 β-strands and 13 short α-helices with a core β-sandwich fold of two sheets with five antiparallel β-strands (**Fig. 2.3A,B**). A third short β-sheet forms a convex face and two pairs of β-strands (residues 356-369 and 412-423) protrude from the region between the β-sandwich and the third β-sheet. In this structure, two short β-strands lie at the entrance and exit

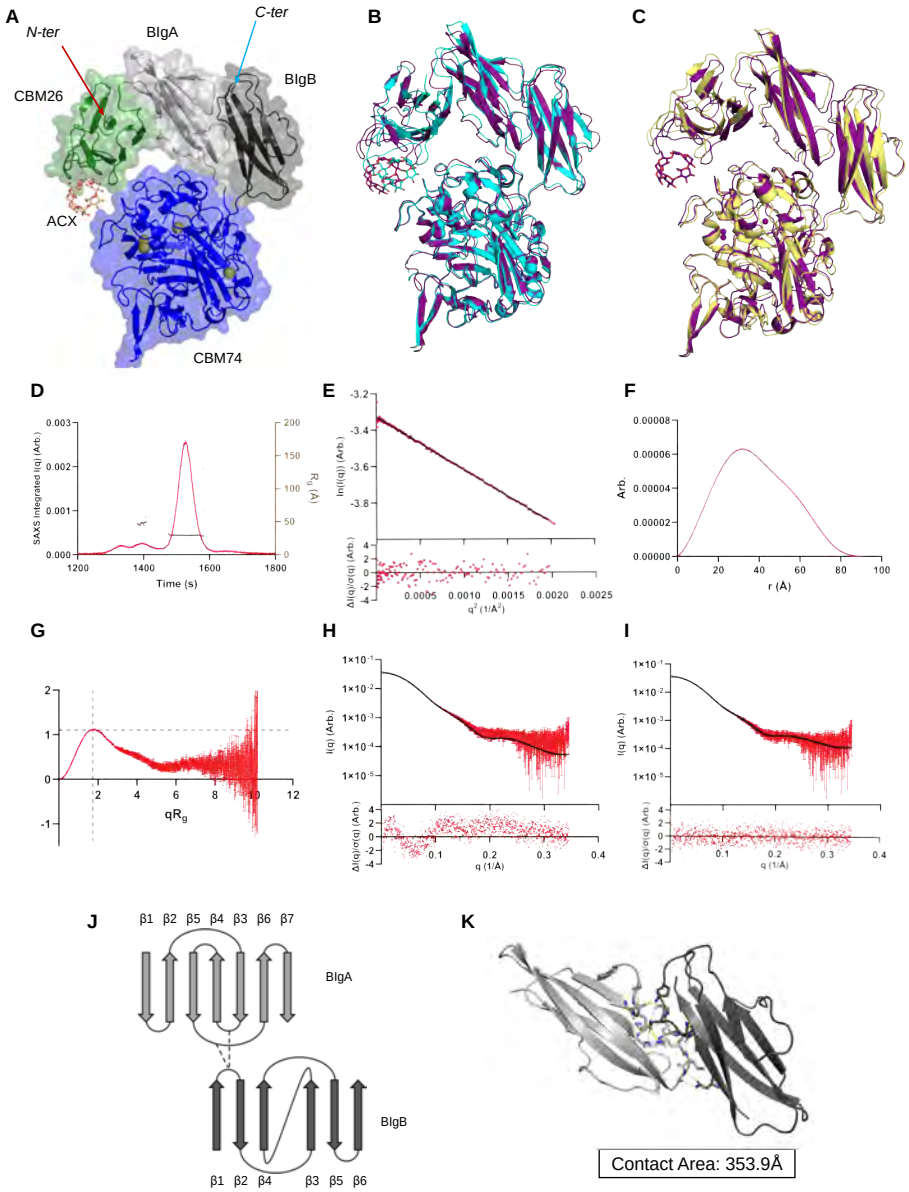
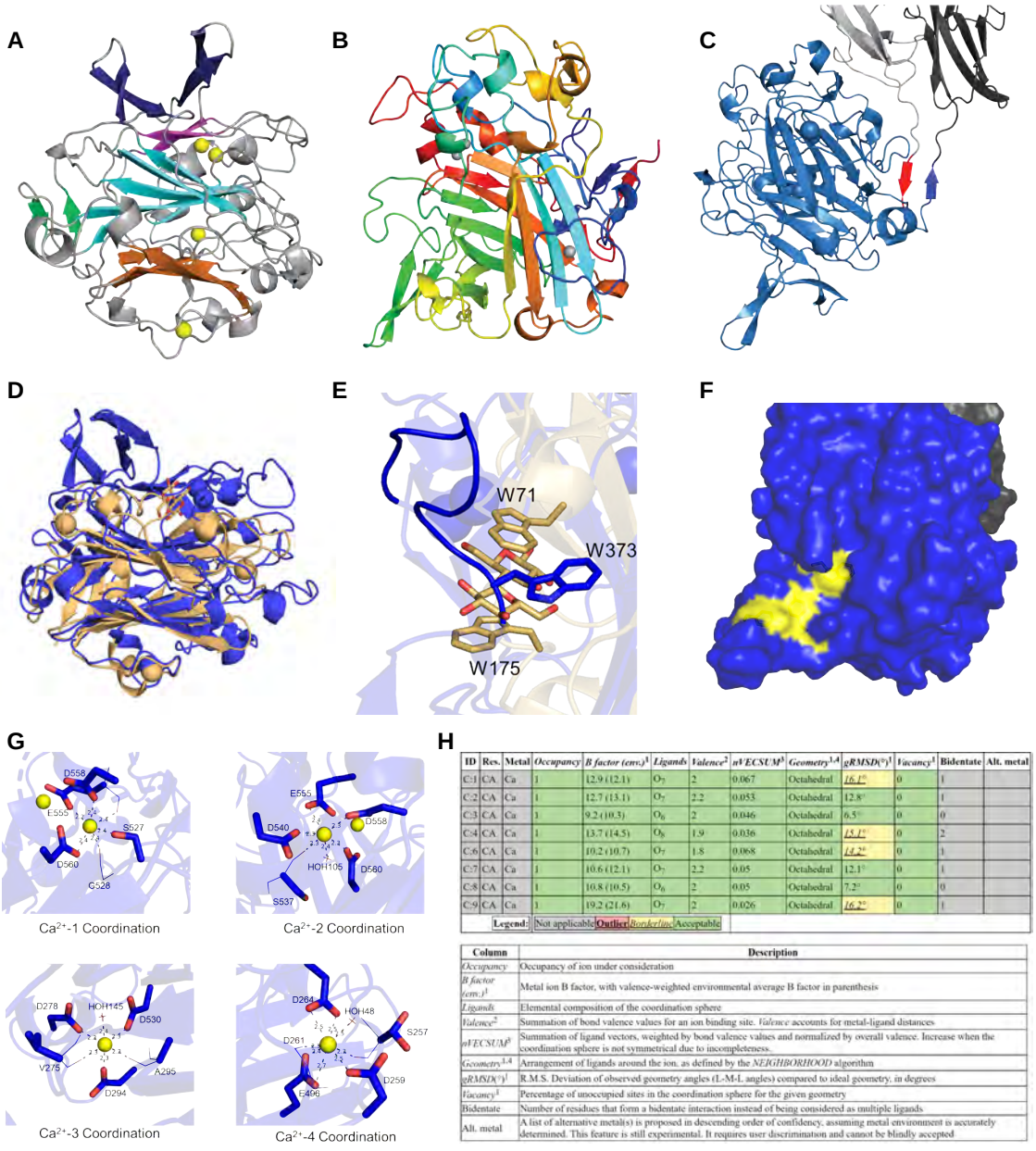


Figure 2.2: Sas6 is a compact protein with two BIg domains that orient *Rb*CBM26 and *Rb*CBM74. **A.** Semi-transparent surface rendition and cartoon of Sas6T (PDB 7uww) with *Rb*CBM26 in green, BIgA in light grey, *Rb*CBM74 in blue, and BIgB in dark grey. α-cyclodextrin (ACX) bound to *Rb*CBM26 is shown in wheat sticks and Ca^{2+} atoms are shown as yellow spheres. **B.** Overlay of Chain A (purple) and Chain B (cyan) within the asymmetric unit of 7uww, anchored on the CBM74, showing variation in the position of ACX relative to *Rb*CBM74. **C.** Overlay of Chain A of 7uww (purple) and SAXS-derived MultiFoXS model (yellow) anchored on the CBM74. RMSD of 1.2Å over 347 pruned atom pairs. **D.** Total subtracted scattering intensity (left y axis) and Rg (right y axis) as a function of time for the SEC-SAXS elution. **E.** Guinier fit analysis with normalized residual shown in the bottom panel. **F.** P(r) versus r normalized by I(0). **G.** Dimensionless Kratky plot; y=3/e and x=$\sqrt{3}$ as dashed gray lines to indicate where a globular protein would peak. **H.** FoXS and **I.** MultiFoXS fits (black) to the Sas6T SAXS data (red) with normalized residual shown in the bottom panel. **J.** Topology map of BIgA and BIgB domains illustrating the Greek key motif in BIgA and showing the loops that hydrogen bond with one another. **K.** A surface area analysis of the BIg domains using PISA in CCP4 gives a buried surface area of 353.9Å (123). Residues providing hydrogen bonding are represented by stick side chains and the hydrogen bonds are shown by dashed yellow lines.



ID	Res.	Metal	Occupancy	B factor (env.)[1]	Ligands	Valence[2]	nVECSUM[3]	Geometry[3,4]	gRMSD(°)[1]	Vacancy[1]	Bidentate	Alt. metal
C:1	CA	Ca	1	12.9 (12.1)	O_7	2	0.067	Octahedral	16.1°	0	1	
C:2	CA	Ca	1	12.7 (13.1)	O_7	2.2	0.053	Octahedral	12.8°	0	1	
C:3	CA	Ca	1	9.2 (10.3)	O_6	2	0.046	Octahedral	6.5°	0	0	
C:4	CA	Ca	1	13.7 (14.5)	O_8	1.9	0.036	Octahedral	15.1°	0	2	
C:6	CA	Ca	1	10.2 (10.7)	O_7	1.8	0.068	Octahedral	14.2°	0	1	
C:7	CA	Ca	1	10.6 (12.1)	O_7	2.2	0.05	Octahedral	12.1°	0	1	
C:8	CA	Ca	1	10.8 (10.5)	O_6	2	0.05	Octahedral	7.2°	0	0	
C:9	CA	Ca	1	19.2 (21.6)	O_7	2	0.026	Octahedral	16.2°	0	1	

Legend: Not applicable | Outlier | Borderline | Acceptable

Column	Description
Occupancy	Occupancy of ion under consideration
B factor (env.)[1]	Metal ion B factor, with valence-weighted environmental average B factor in parenthesis
Ligands	Elemental composition of the coordination sphere
Valence[2]	Summation of bond valence values for an ion binding site. Valence accounts for metal-ligand distances
nVECSUM[3]	Summation of ligand vectors, weighted by bond valence values and normalized by overall valence. Increase when the coordination sphere is not symmetrical due to incompleteness.
Geometry[3,4]	Arrangement of ligands around the ion, as defined by the NEIGHBORHOOD algorithm
gRMSD(°)[1]	R.M.S. Deviation of observed geometry angles (L-M-L angles) compared to ideal geometry, in degrees
Vacancy[1]	Percentage of unoccupied sites in the coordination sphere for the given geometry
Bidentate	Number of residues that form a bidentate interaction instead of being considered as multiple ligands
Alt. metal	A list of alternative metal(s) is proposed in descending order of confidency, assuming metal environment is accurately determined. This feature is still experimental. It requires user discrimination and cannot be blindly accepted

Figure 2.3: *Rb*CBM74 is a single globular domain. **A.** Side view of *Rb*CBM74 with the central β-sandwich sheets in orange and cyan. A third β-sheet is shown in magenta and the protruding pairs of β-strands and in dark blue. β-strands connecting the beginning and end of the *Rb*CBM74 are colored green. Ca^{2+} atoms are shown as yellow spheres. **B.** Structure of *Rb*CBM74 (PDB 7uww) colored from N-terminus (blue) to C-terminus (red). **C.** Short β-strands leading into and out of *Rb*CBM74 are colored in red and blue. **D.** Overlay of *Tm*CBM9 (gold) (PDB 1i82-A) and *Rb*CBM74 (blue). The DALI server calculated an RMSD of 3.2Å and sequence identity of 17%. **E.** Close-up view of *Tm*CBM9 binding site showing the two *Tm*CBM9 Trp residues involved in binding cellobiose (gold) and W373 of *Rb*CBM74 (blue) which lies in the same region but is occluded from the surface by a loop containing residues 374-384. **F.** Y367 and Y424 (yellow) are surface exposed aromatic residues in the same vicinity as W373 (yellow). **G.** Zoomed in view of calciums coordinated in the *Rb*CBM74 with side chains shown in sticks, main chain shown in lines and Ca^{2+} ions by yellow spheres. Atomic distances are shown in Å and residues are labeled. Residues are colored by element with oxygen shown in red. **H.** Ion validation by web server CheckMyMetal (244) for Sas6T (PDB 7uww).

of the CBM74, marking the domain boundaries (**Fig. 2.3C**).

The central fold of *Rb*CBM74 resembles CBM9 from *Thermotoga maritima* Xylanase10A (PDB ID: 1I82-A, Z-score: 9.8, RMSD: 3.2Å, identity: 17%) (97; 164) (**Fig. 2.3D**). *Tm*CBM9 binds glucose, cellobiose, cello- and xylo- oligomers at the reducing ends, and amorphous and crystalline cellulose (164). *Tm*CBM9 (189 residues) is larger than most CBMs which range from 80-120 amino acids (164). The ligand binding site of *Tm*CBM9 is formed by two tryptophan residues that create an aromatic clamp around cellobiose. In *Rb*CBM74, W373 is conserved with one of these tryptophans and lies within an extended, shallow channel partially covered by residues 374-384 that form a flexible loop only resolved in one monomer (**Fig. 2.3E**). In addition to W373, there are surface-exposed two Tyr residues, Y397 and Y424, that we hypothesized might form a binding site (**Fig. 2.3F**).

There are three putative structural Ca^{2+} in the *Tm*CBM9 structure and four in *Rb*CBM74, one of which (Ca^{2+}-4) aligns with a Ca^{2+} in *Tm*CBM9 (**Fig. 2.3G**). We modeled these as Ca^{2+} based upon coordination geometry and atomic distances (**Fig. 2.3H**) (155; 244). Ca^{2+}-1 and Ca^{2+}-2 are separated by 3.8Å and share three coordinating residues but only Ca^{2+}-2 is surface exposed. Like *Tm*CBM9, the Ca^{2+} ions in the *Rb*CBM74 structure may be important for structural stability (206).

2.3.6 Molecular Basis of *Rb*CBM26 Binding

The N-terminal *Rb*CBM26 displays a β-sandwich consistent with other members of the CBM26 family (106). In both chains of the asymmetric unit, CH/π stacking with ACX is provided by W63 and Y55 with hydrogen bonding mediated by Y53, K101, Q103, and the peptidic oxygen of A107 (**Fig. 2.4A**). In chain A only, K97 provides hydrogen bonding with O3 of Glc6. In chain B, ACX lies 3.2Å from S286 of the CBM74 and hydrogen bonds with O2 and O3 of Glc3. In contrast, S286 is 9.5Å from ACX in chain A. The top structural homologs of *Rb*CBM26 from DALI are the CBM25 and CBM26 of α-amylase G-6 from *Bacillus halodurans* C-125 (*Bh*CBM25 and *Bh*CBM26), and *Er*CBM26b of Amy13K from *Eubacterium rectale* (97; 96). In all three structures, the overall fold and starch-binding platform are conserved (**Fig. 2.4B**) (26; 39). *Rb*CBM26, in contrast to *Er*CBM26 and *Bh*CBM26, has a longer loop containing K97 and K101 that provide additional hydrogen bonding with ACX. Unlike *Bh*CBM26, *Rb*CBM26 does not undergo a conformational change upon ligand binding (**Fig. 2.4C**) (26). A sequence alignment with CBM26 members *Bh*CBM26, *Er*CBM26 and the *Lactobacillus amylovorus* α–amylase CBM26 (*La*CBM26), demonstrates conservation of the aromatic platform but more variation in the hydrogen-bonding network (**Fig. 2.4D**). Sas6 W63 corresponds to *La*CBM26 W32 that, when mutated, results in complete loss of binding (185). The *R. bromii* protein Sas20 has a CBM26-like domain that shares 26% sequence identity with *Rb*CBM26, yet *Rb*CBM26 shares more structural similarity with *Bh*CBM26

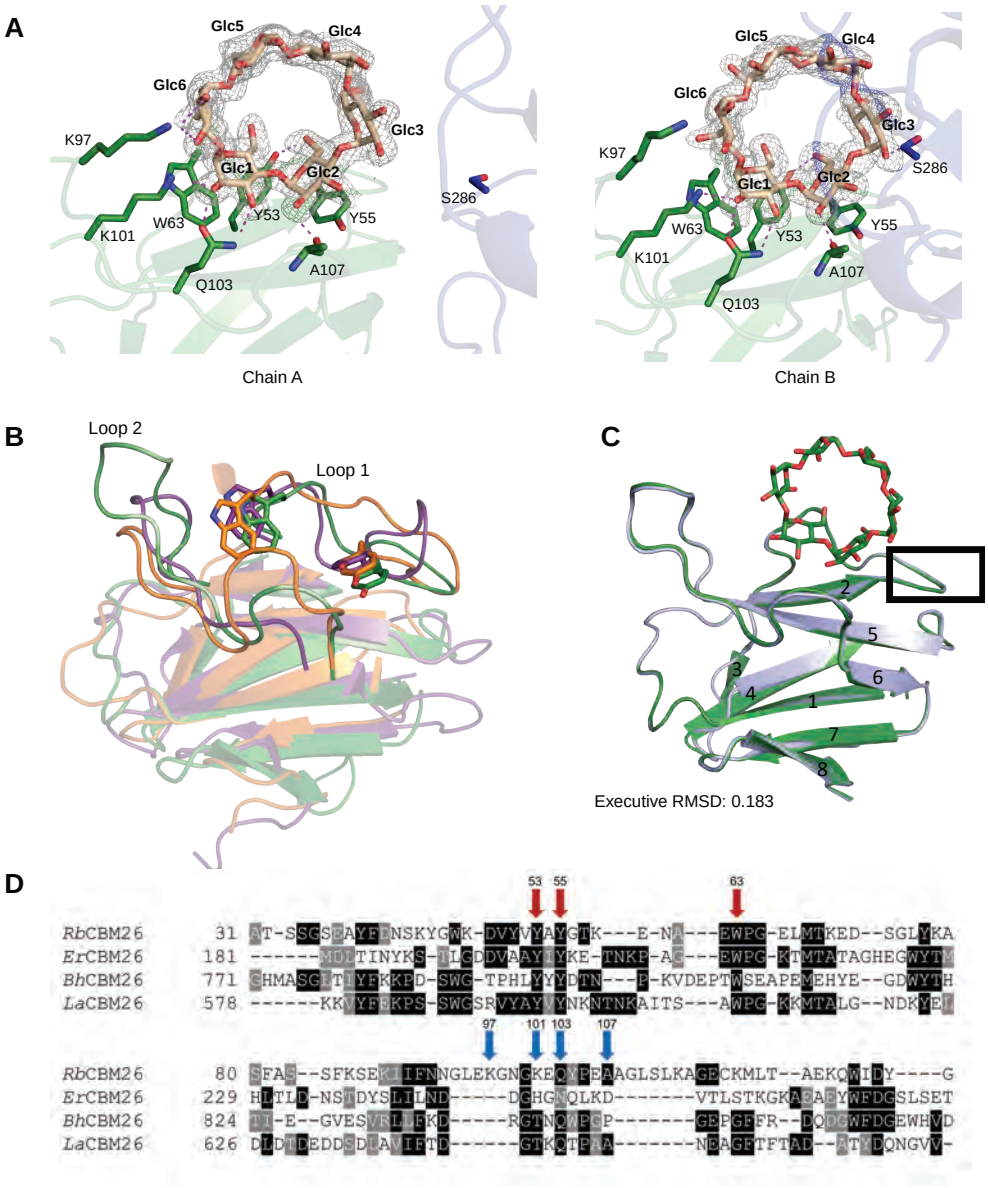


Figure 2.4: *Rb*CBM26 shares a conserved binding site with other CBM26 family members. **A.** ACX bound at *Rb*CBM26 (green) in chain A (left) and chain B (right), demonstrating minor conformational flexibility that places S286 from *Rb*CBM74 (blue) within the binding site. Side chains involved in ligand binding are shown as green sticks with a hydrogen bond cutoff of 3.2 Å. ACX is displayed as wheat sticks. Omit map is contoured to 2.0σ and carved within 1.6 Å of ACX ligand. **B.** The top structural homologs of *Rb*CBM26 from DALI are the CBM25 from Bacillus halodurans C-125 (*Bh*CBM25) from α-amylase G-6 (PDB ID: 2C3V-A, Z-score: 12.4, RMSD 1.9 Å, identity: 16%) and CBM26 (*Bh*CBM26) from the same enzyme (PDB ID: 6B3P-B, Z-score: 12.1, RMSD 1.9 Å, identity: 20%) (96; 26). Another top DALI result is *Er*CBM26b of Amy13K from *Eubacterium rectale* (PDB ID 2C3H-B, Z-score: 10.8, RMSD 1.7 Å, identity: 19%). Overlay of *Rb*CBM26 (green) with *Bacillus halodurans* CBM26 (PDB 2c3h, orange), and *Eubacterium rectale* Amy13K CBM26 (PDB 6b3p, purple). **C.** Overlay of unliganded *Rb*CBM26 (blue) and ACX-bound *Rb*CBM26 (green) showing that loop 1 does not move upon ligand binding. β-strands are numbered for reference. **D.** Sequence alignment of *Rb*CBM26, *Er*CBM26 (ERE_20420), *Bh*CBM26 (BH0413), and *La*CBM26 (Q48502). Conserved binding site residues are indicated by a red arrow while variable residues are indicated by a blue arrow and provide hydrogen bonding.

and *Er*CBM26 (35). While the previously studied CBM26 modules have not been assayed for binding to ACX, many bind α-cyclodextrin, which has a similar geometry, within a K_d range of ~350 – 600μM (26; 39).

2.3.7 Contributions of *Rb*CBM26 and *Rb*CBM74 to Overall Sas6 Binding

The tandem CBM26 modules of the *L. amylovorus* amylase enhance degradation of starch granules (88). Therefore, we performed an adsorption depletion (i.e., pulldown) assay of Sas6T and its constituent CBMs for binding to starch granules. Sas6T bound corn starch (K_d =0.95μM ± 0.15; Bmax=0.102μmol/g ± 0.001) with modestly better affinity than potato starch granules (K_d =1.66μM ± 0.40; Bmax=0.026μmol/g ± 0.004) (**Fig. 2.5A**). The binding capacity (Bmax) for corn starch is ~4-fold higher. We found that BIg-*Rb*CBM74-BIg bound both corn starch (K_d =1.57μM ± 0.40; Bmax=0.114μmol/g ± 0.01) and potato starch (K_d =2.69μM ± 1.59; Bmax=0.031μmol/g ± 0.006) starch at similar affinity to Sas6T, with no statistically significant difference, while *Rb*CBM26 did not display measurable binding to either (**Fig. 2.5A**). These results support that *Rb*CBM74 drives insoluble starch binding by Sas6.

We screened Sas6T, *Rb*CBM26 and BIg-*Rb*CBM74-BIg for binding to amylose and amylopectin via affinity PAGE demonstrating that *Rb*CBM26 displays poor recognition of both polysaccharides based upon the relatively small change in migration (**Fig. 2.5B**). Using isothermal titration calorimetry (ITC), we found that Sas6T and BIg-*Rb*CBM74-BIg bound amylopectin with sub-micromolar affinity whereas binding was not detectable for *Rb*CBM26 (**Fig. 2.5C; Table 2.3**) (1). Sas6T binds maltotriose (G3), maltoheptaose (G7), maltooctaose (G8) with a K_d in the hundreds of μM but exhibits a K_d of 6.2 ± 2 μM for maltodecaose (G10). Titrations with higher concentrations of G10 allowed for full saturation of both the *Rb*CBM74 site (K_d = 8.6 ± 2.7 μM) and *Rb*CBM26 (K_d = 730 ± 59 μM). Binding of G10 to *Rb*CBM26 when this domain was expressed separately resulted in modestly better binding (K_d = 252 ± 128 μM) perhaps due to better access in the binding site (**Table 2.3; Fig. 2.5C**). Interestingly, *Rb*CBM26 binds G7 and α-cyclodextrin, while BIg-*Rb*CBM74-BIg had no detectable affinity for these sugars (**Table 2.3; Fig. 2.5C**). None of the constructs bound glucosyl-α1,6-maltotriosyl-α1,6-maltotriose, an oligosaccharide of pullulan, suggesting that the α1,6 linkages are not specifically recognized by either domain. We determined that BIg-*Rb*CBM74-BIg binds exclusively longer α-glucans of at least 8 residues. Notably, α1,4-linked glucose polymers form double helices at 10 glucose units due to internal hydrogen bonding, so we hypothesized that *Rb*CBM74 accommodates a double helical structure (172).

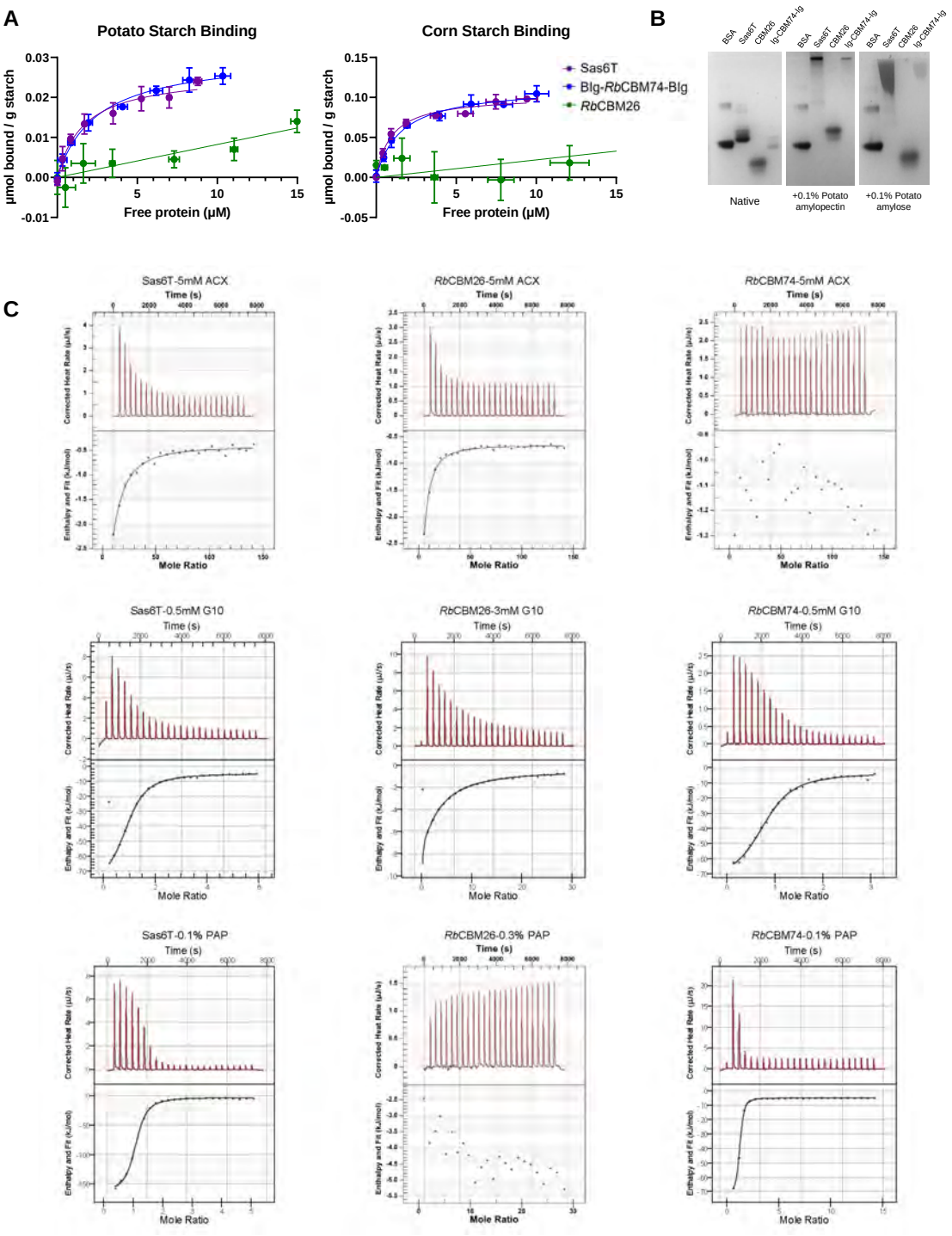


Figure 2.5: **A.** *Rb*CBM74 drives binding to granular potato and corn starch as determined by adsorption depletion. The μmoles of protein bound per gram of starch was plotted against free protein to determine dissociation constants (K_d) and binding maxima (B_{max}) using a one-site specific binding model in GraphPad prism. Graphs show non-linear fit of 3 experiments with points indicating the mean and standard deviation. **B.** Affinity PAGE of Sas6T or individual domains, *Rb*CBM26 and BIg-*Rb*CBM74-BIg, with 0.1% polysaccharide. BSA = bovine serum albumin. **C.** Representative ITC graphs for Sas6 constructs bound to α-cyclodextrin (ACX), maltodecaose (G10), or potato amylopectin. Note that exothermic heat release is denoted with an upward peak on this machine.

2.3.8 Molecular Basis of *Rb*CBM74 Binding

We co-crystallized BIg-*Rb*CBM74-BIg with G10 to 1.70Å resolution (Rwork=17.9%, Rfree=19.9%) (**Fig. 2.6A**). Remarkably, we observed two molecules of G10 as an extended double helix of ~42Å along the face of *Rb*CBM74 extending from S286 (reducing ends) to W373 (non-reducing ends). There was strong electron density for 12 glucoses in one molecule, and nine glucoses in the other chain. This was most likely the result of slight variation in the occupancy of the G10 chain along the binding cleft among different monomers in the unit cell; the electron density thus provides us with composite of the placement of ligand in the binding cleft (**Fig. 2.6B**). An overlay of the unliganded and G10 bound structures demonstrates little global change in the CBM74 upon binding with the exception of G374 to K381 (**Fig. 2.6C**). In the unliganded structure this loop occludes surface exposure of W373 and in the G10 bound structure the loop opens to create a continuous binding surface (**Fig. 2.6C,D**). Additionally, Ca^{2+}-4 is exchanged for Na^{+}, representing flexibility in ion identity at that site.

G10 is arranged as two parallel left-handed helices (G10A and B) stabilized by hydrogen bonds within each chain (most as Glcn O2 – Glcn+1 O3) and between chains (most as O2 to O6) (**Fig. 2.6E,F**). The Φ (O5-C1-O4'-C4'), Ψ (C1-O4'-C4'-C5') angles of G10A and B approximate those observed in crystal structures of double helical A (88.8°± 3, -149.2°± 4) and B type starch (84.1°± 0.3, -144.4°± 0.3) (**Table 2.2**) (104; 105). The Φ, Ψ angles vary along the length of the G10 double helix as predicted in double helical models of amylopectin (170) (**Fig. 2.6F**) (**Table 2.2**). The average pitch, the length for one complete helical turn, of G10A and G10B is 17.1 ± 1.7 Å, which deviates from the ~8 Å pitch of left-handed single helical V-type amylose that features Φ = 91 -115° and Ψ = - 97-131° (80) (**Fig. 2.6G**). The substantially different dimensions of single helical α-glucan prevent its selection within the *Rb*CBM74 binding site leading us to conclude that this domain selects for the elongated double helical geometry.

Each G10 molecule interacts with protein as a stretch of three Glcs at a time, before the natural helical curvature brings the chain out of the contact with the protein (**Fig. 2.6H**). For example, at the non-reducing end, Glc 1-3 of G10A fit into the ligand-binding groove, while Glcs 4-6 of G10A are solvent exposed and Glc 1-3 of G10B then fill the cavity. Along the length of the cavity, from the non-reducing end to the reducing end, Glcs 1-3 and 7-9 of both G10A and G10B alternate to fill this binding site.

The binding cleft features a network of residues that hydrogen bond to the hydroxyl groups of glucose (**Fig. 2.6H**). At the non-reducing end, Glc A1 hydrogen bonds with the indole nitrogen of W373. Glc A2 stacks with W373 with hydrogen bonding provided by G374 and N403. Glc A3 hydrogen bonds with S338. G10B contacts the next part of the binding groove and is anchored by hydrogen bonding of Glc B3 by R336 and Y524. Where the first molecule turns back into the binding groove, Glc A8 hydrogen bonds with E290, D549, and K556. Glc A9 hydrogen bonds

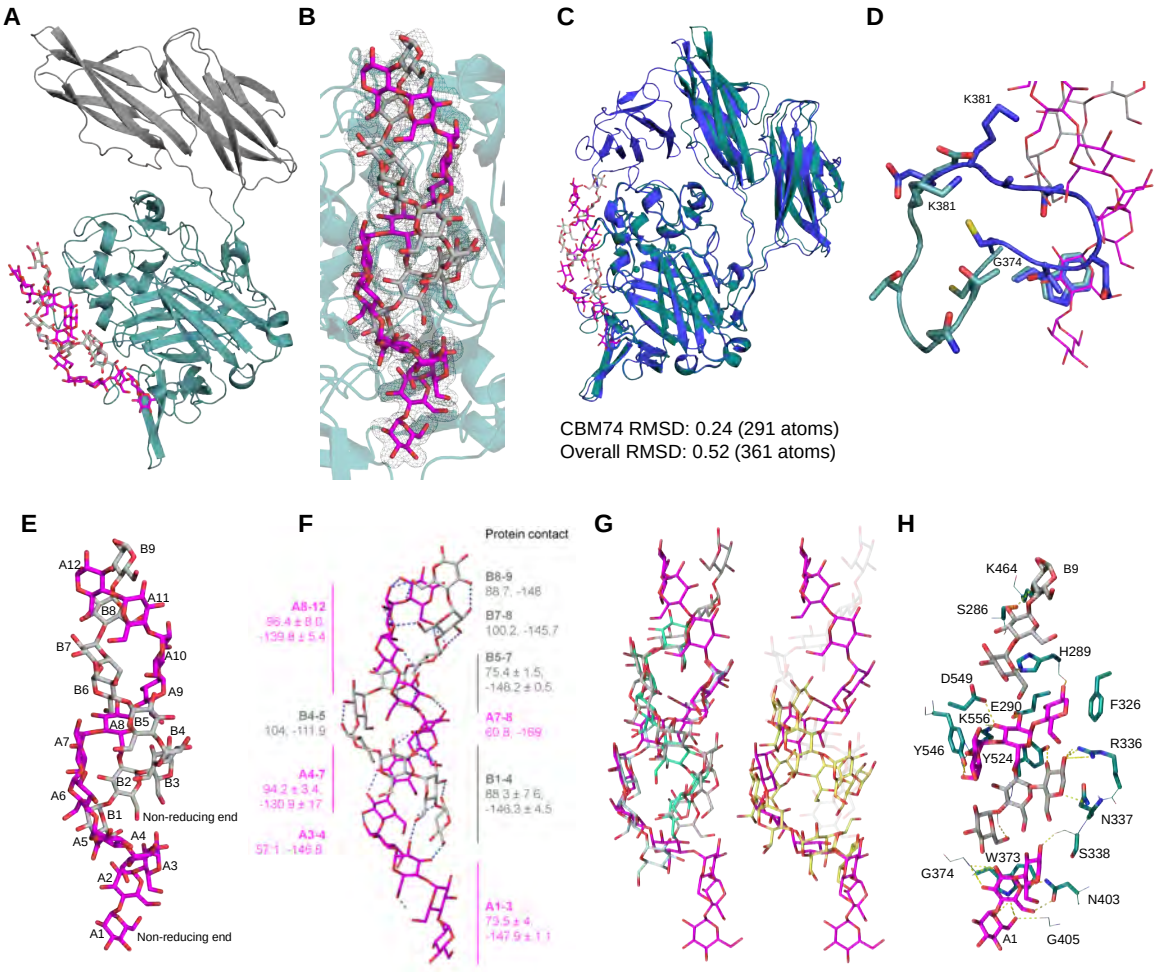


Figure 2.6: **A.** The BIg-*Rb*CBM74-BIg (PDB 7uwv) starch-binding site is an extended groove that spans nearly the length of the domain. A cartoon representation of BIg domains in light grey, CBM74 in teal with two chains of maltodecaose (G10) wrapped around one another shown in magenta and grey sticks. **B.** *Rb*CBM74 is co-crystallized with G10 in a double helical conformation. Electron density for G10 demonstrated by an omit map contoured to 2.0σ and carved to 1.6$\mathring{A}$ with one chain of modeled Glc in magenta and the other in grey. **C.** Overlay of *Rb*CBM74 from Sas6T structure (PDB 7uww) in blue with *Rb*CBM74 from BIg-*Rb*CBM74-BIg co-crystal structure (PDB 7uwv) in deep teal. **D.** Loop from G374-G382 demonstrating that the unliganded loop (blue) occludes W373 but moves to allow access to W373 in the ligand-bound structure (deep teal). **E.** Double helical G10 structure with Glc residues labeled from non-reducing to reducing ends. Chain G10A (A1-12) shown in magenta and chain G10B in grey (B1-9) sticks. **F.** Intramolecular hydrogen bonds (3.6$\mathring{A}$ cutoff for ideal geometry and 3.2$\mathring{A}$ with minimal acceptable geometry) within and between G10 chains are shown in slate. Φ (O5-C1-O4'-C4') and Ψ (C1-O4'-C4'-C5') angles of the Glc linkages in the G10 double helix ligand are labeled with G10A in magenta and G10B in grey. **G.** The geometry of the G10 ligand more closely resembles that of double helical B starch (cyan) than single helical cycloamylose (yellow, 1c58) (80). Models were manually aligned in PyMOL to compare the angles, pitch, and period of the helical turns. **H.** Corresponding hydrogen-bonding network (3.2$\mathring{A}$ cutoff) between *Rb*CBM74 and G10. Side chains involved in hydrogen bonding are shown in teal sticks with nitrogens indicated in blue and oxygens in red. Hydrogen bonds are indicated by yellow dashed lines and G10 residues directly involved in binding are shown in magenta (G10A) and grey (G10B) sticks.

with the backbone of H289 and pi stacks with F326. The H289 side chain hydrogen bonds with Glc B7 and provides aromatic character for pi stacking with Glc B8. Near the region of *Rb*CBM74 that lies adjacent to *Rb*CBM26, K464 and S286 hydrogen bond with Glc B9.

To define the starch-binding properties of *Rb*CBM74 in solution, we employed Hydrogen–Deuterium eXchange Mass Spectrometry (HDX-MS). The conformational dynamics of BIg-*Rb*CBM74-BIg alone and in the presence of G10 were measured over a 4-log timescale (**Fig. 2.7A**). The overall conformational dynamics of the apo protein were consistent with the determined crystal structure, in terms of well-ordered domains and associated loops or flexible regions. The flanking BIg domains showed higher exchange rates than the core CBM74. Intriguingly, the linker regions between domains do not show differentially high dynamic exchange, as would be expected for flexibly tethered independent domains, further supporting the integral nature of BIg-*Rb*CBM74-BIg motif.

The binding of G10 to *Rb*CBM74 was explored by differential protection from exchange in the absence and presence of G10. Significant protection was observed in the presence of G10, while no significant increases in exchange were observed (**Fig. 2.7B**). This is consistent with the minimal global conformation changes between the two states of the protein. The protected regions upon G10 binding were highly localized to a single surface binding region (**Fig. 2.7C**). This protected region constitutes a single extended surface, which directly overlaps with the G10 binding site observed in the co-crystal structure (**Fig. 2.7D**). With the exception of the peptide from A314-Y318 (ANTTY), each of the protected peptides identified by HDX-MS contains at least one key binding residue identified from the co-crystal structure (**Fig. 2.7C,D**). These data provide a comprehensive picture of the structural dynamics of *Rb*CBM74 binding to long maltooligosaccharides via an extended starch binding cleft.

2.3.9 CBM74 Conservation

An alignment of all 99 CBM74 sequences demonstrates that the CBM74s fall into 6 distinct clades (**Fig. 2.12**). *Rb*CBM74 (No. 28) is in a distinct cluster of proteins (blue) that invariably include a dockerin domain as part of the full-length protein. However, there are other CBM74s originating from dockerin-containing proteins found in three more groups (green, cyan, and magenta) (**Fig. 2.11**). The prototypical CBM74 of the subfamily GH13_32 α-amylase from *Microbacterium aurum* (No. 52) bins into a clade (cyan) with its GH13_32 counterpart from *Sanguibacter* sp. (No. 54) and the CBM74-containing α-amylase from *Clostridium bornimense* (No. 58). A similar GH13_28 α-amylase from *Streptococcus suis* (No. 68) is in the adjacent cluster (magenta) very close to the CBM74s from two other hypothetical dockerin-containing proteins from *Ruminococcus bovis* (No. 67) and Ruminococcaceae bacterium (No. 70). Most CBM74s appended to α-amylases from the

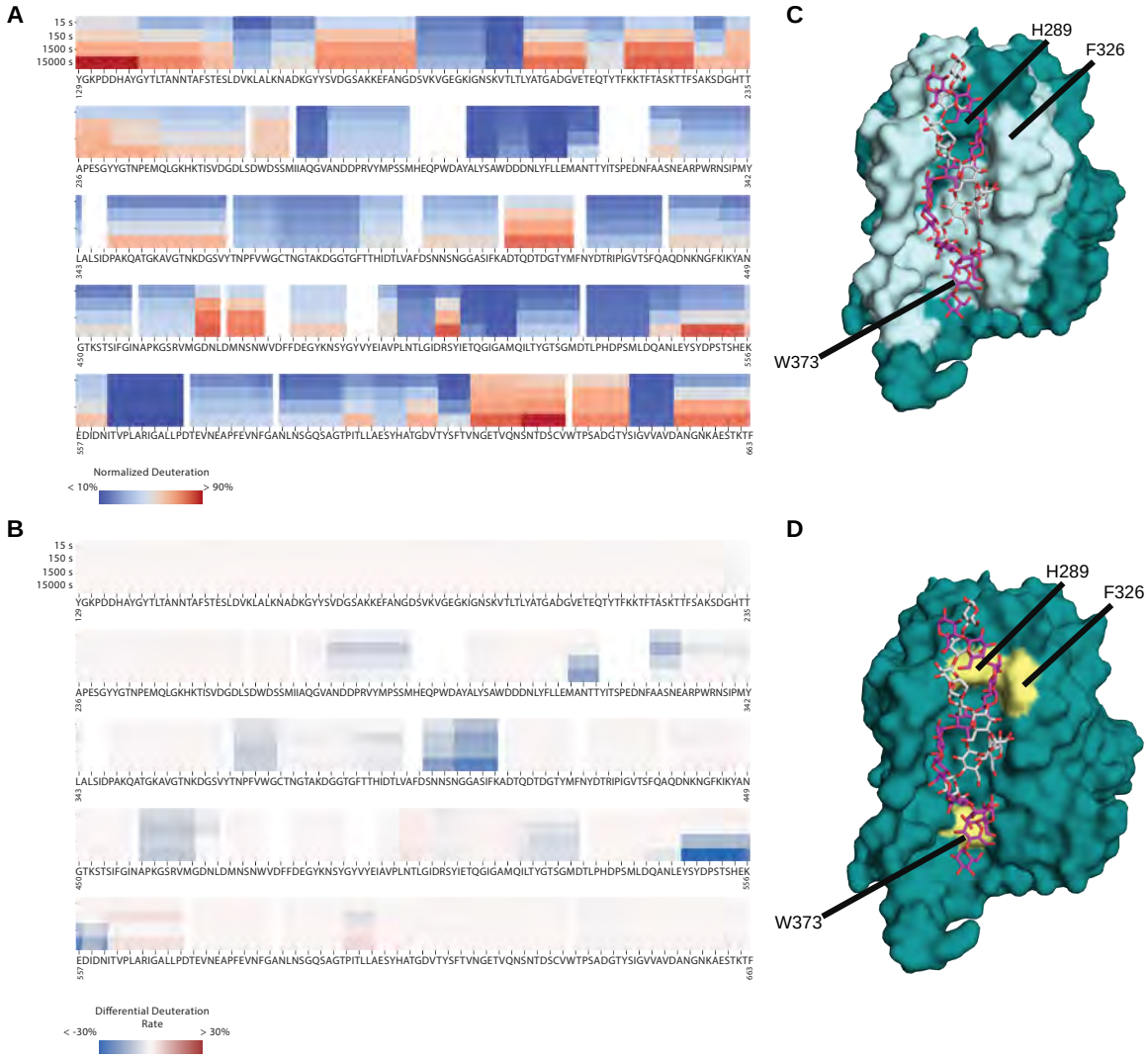


Figure 2.7: HDX-MS analysis of *Rb*CBM74

A. Heatmap of exchange dynamics of BIg-*Rb*CBM74-BIg. All values are the average of three replicates. **B.** Heatmap of the differential exchange dynamics of BIg-*Rb*CBM74-BIg in the absence and presence of G10. Blue represents lower exchange (protection) in the G10 bound form and red higher exchange in the G10 bound form. All values are the average of three replicates. **C.** Surface representation of RbCBM74 with peptides protected from deuterium exchange in the presence of G10 colored in light cyan as determined by hydrogen-deuterium exchange mass spectrometry. **D.** A surface representation of CBM74 (teal) with aromatic residues that provide CH-π stacking colored in yellow and G10 represented by magenta and grey sticks.

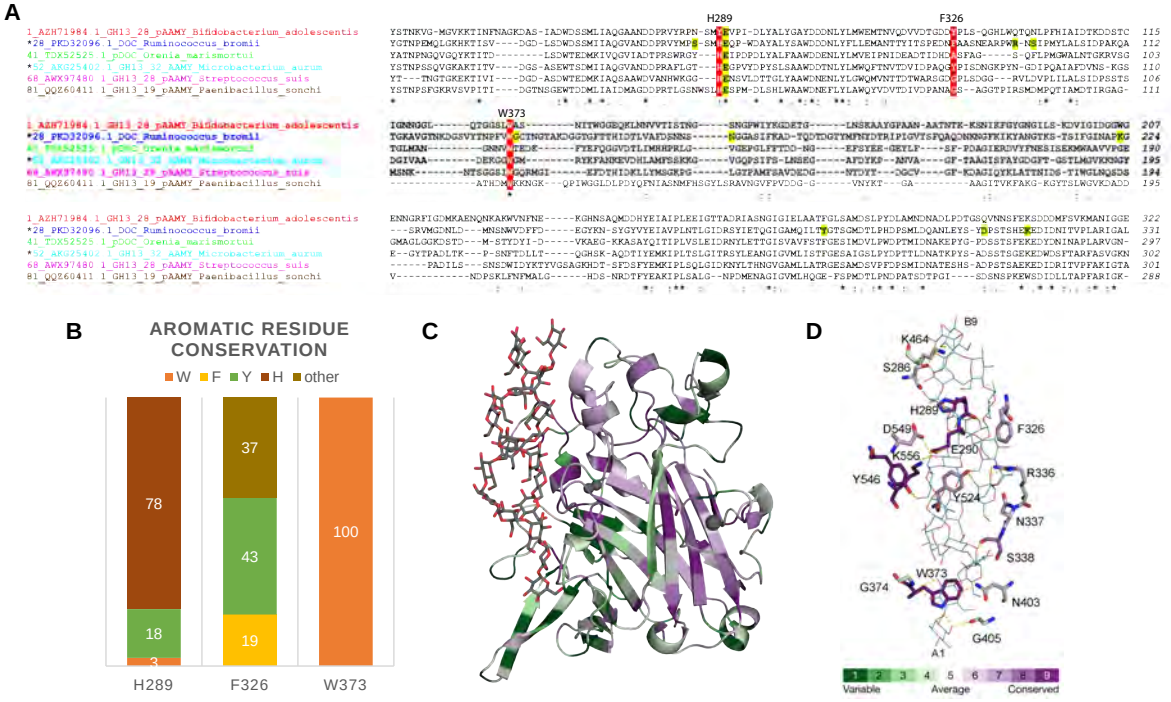


Figure 2.8: Conservation of binding residues among select CBM74 family members

A. Sequence alignment of 6 representative sequences, one from each of the 6 clades of the CBM74 family. The residues responsible for stacking interactions are highlighted in red and those involved in hydrogen bonding with glucose moieties of the bound α-glucan are highlighted in yellow, based upon the *Rb*CBM74 structure (blue). Identical and similar positions are signified by asterisks and dots/semicolons under the alignment blocks. **B.** Conservation of H289, F326 and W373 among the 99 CBM74 members is displayed as a bar graph. **C.** *Rb*CBM74 (PDB 7uwv) colored by conservation score from least conserved (green) to most conserved (purple) generated using CONSURF **D.** Binding residues of *Rb*CBM74 (PDB 7uwv) colored by conservation score.

subfamily GH13_28, predominantly from Bifidobacteria, group together in a separate cluster (red). Finally, the sixth cluster (walnut) covers CBM74s found in GH13_19 α-amylases. In total, CBM74s occur in α-amylases from several subfamilies or non-catalytic dockerin-containing proteins and are widely represented among Bifidobacteria.

We aligned representative sequences from each clade to highlight the similarities within these binding sites (**Fig. 2.8A**). Here as well as within the full alignment, W373 from *Rb*CBM74 is 100% conserved among all 99 family members (**Fig. 2.8B**). H289 is shared with 78 sequences or substituted with a Tyr (18/99) in Bifidobacteria and Candidatus scatavimonas (No. 25) and a Trp (3/99) in Pseudoscardovia species (**Fig. 2.8B**). F326 is perhaps the most variable, sharing sequence identity or similarity with 3 of the 6 clades (F-19/99, Y-43/99), while the other clades feature a glycine or alanine in this position (36/99). We mapped the conservation of all 99 CBM74 family members onto our structure using CONSURF (10; 9; 34) (**Fig. 2.8C**). The binding site features an elaborate network of residues that provide hydrogen bonding with the ligand. The residues at the center of the cleft including K556 (80/99), D549 (63/99), and E290 (99/99) exhibit the highest conservation (**Fig. 2.8D**). The hydrogen bonding residues at the ends of the cleft are more varied, including S286 (22/99) which interacts with the *Rb*CBM26 ligand. Intriguingly, in a large proportion of the sequences there is an aromatic residue at the site of K556 (W-19/99) and Y524 (Y-12/99, F-45/99) that could provide pi stacking in those CBM74s (**Fig. 2.8A,B**). This moderate variability in the composition of the putative binding site may suggest that CBM74 family members have different affinities for starch.

2.3.10 *Rb*CBM74 Mutational Studies

Because most CBM binding is mediated by aromatics, we hypothesized that mutation of W373, F326, or H289 would dramatically decrease or eliminate binding. We therefore mutated each of these residues to alanine separately within the BIg-*Rb*CBM74-BIg construct. We tested maximum binding of each of the aromatic mutants to insoluble corn (1%) and potato starch (5%). The W373A and H289A constructs lost the ability to bind to insoluble corn starch while binding of the F326A construct was greatly reduced (**Fig. 2.9A**). This trend was somewhat different for potato starch, in which a lower percentage of H289A bound compared to the F326A and W373A mutants. By affinity PAGE, neither the W373A nor the F326A mutant lost appreciable binding to amylopectin while the H289A mutant had a modest decrease in binding to potato amylopectin (**Fig. 2.9B**). When we quantified binding via ITC, W373A lost all binding for G10 while H289A and F326A had a ~10-20-fold decrease in affinity (**Fig. 2.9C; Table 2.3**). On potato amylopectin, F326A had a 10-fold reduction in affinity while H289A and W373A exhibited a ~20-fold reduction (**Fig. 2.9C, Table 2.3**). That single mutations do not eliminate binding is perhaps not surprising given

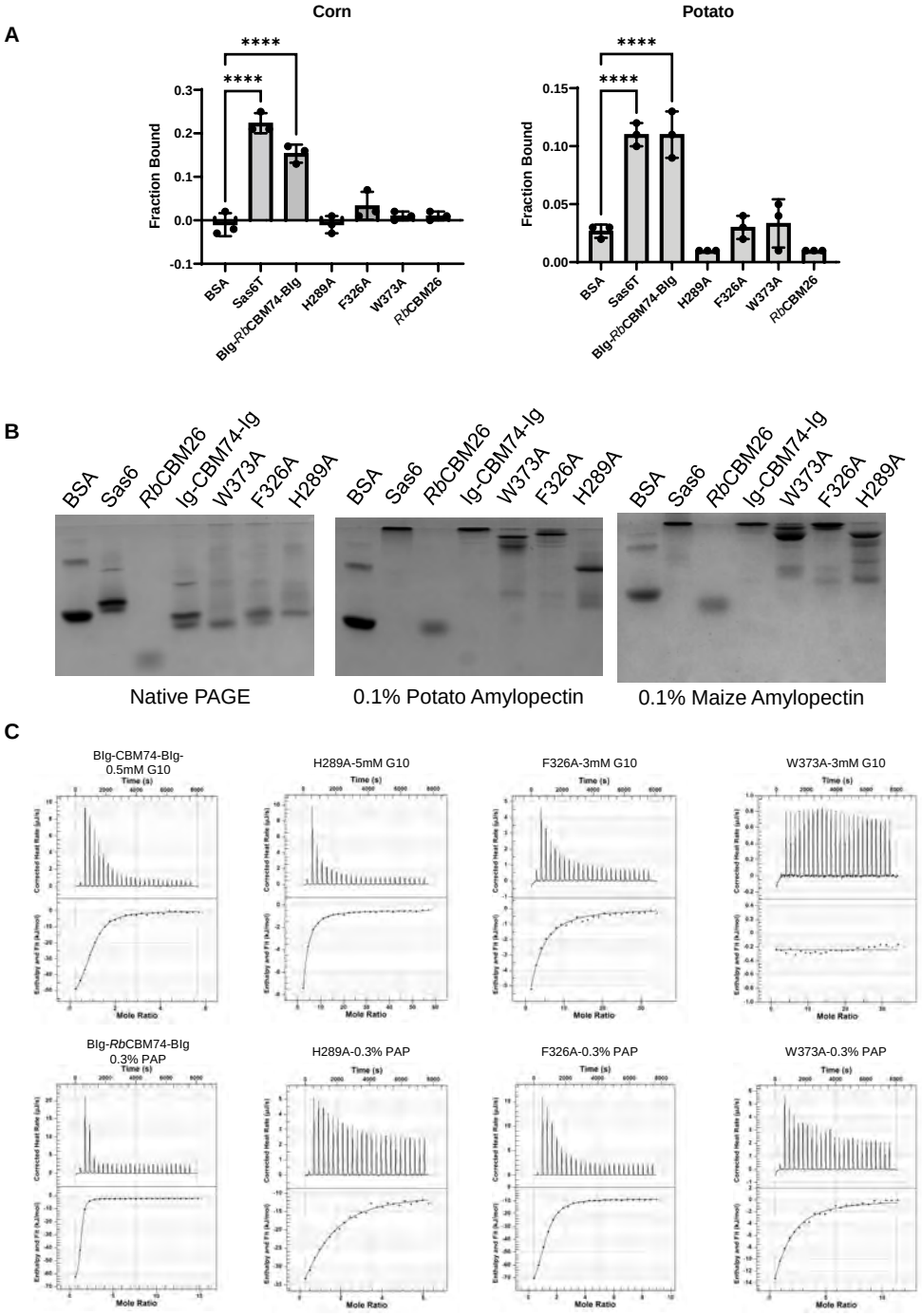


Figure 2.9: W373A, F326A, and H289A mediate starch binding by *Rb*CBM74

A. Binding to insoluble starch is eliminated or greatly reduced when W373, H289 or F326 is mutated to alanine. The amount of protein bound to starch granules was determined by quantitation of protein remaining in solution after binding (n = 3). **B.** Mutation of aromatic residues decreases but does not eliminate binding to amylopectin. Affinity PAGE with 0.1% potato amylopectin or maize amylopectin added to the gel matrix. Binding is indicated by reduced migration through the gel. **C.** BIg-*Rb*CBM74-BIg, H289A, F236A, and W373A mutations binding to maltodecaose (G10) and potato amylopectin. Note that exothermic heat release is denoted with an upward peak on this machine.

the extensive binding platform. Moreover, the enhanced affinity of these mutants to amylopectin over G10 further suggests that productive interactions with the protein extend beyond a 10-glucose unit footprint. Indeed, the somewhat staggered double helical G10 bound in our crystal structure suggests that at least 12 glucose units contribute to binding (**Fig. 2.6H**).

2.3.11 Native mass spectrometry

ITC revealed a binding stoichiometry of 1:1 between BIg-*Rb*CBM74-BIg and G10, while the co-crystal structure demonstrates that two molecules of G10 are accommodated. To better determine the stoichiometry of this binding event, and the proportion of single vs double helical maltooligosaccharide in solution, we employed native mass spectrometry (MS) in the presence of varying concentrations of G10 (**Fig. 2.10**). Each observed state differed by ~1639 Da, which agrees with the theoretical mass of G10. To obtain binding affinities, we summed the peak intensities of all abundant charge states in our spectra and analyzed these intensity values as described previously (202). The K_d for BIg-*Rb*CBM74-BIg was determined to be 2.16 ± 0.53 μM, which agrees with our ITC data (**Fig. 2.10A,B**). As the concentration of ligand is increased, ligand molecules can bind nonspecifically during the nESI process, generating artifactual peaks in the mass spectra corresponding to a two ligand-bound complex. This step is given by Kn, which corresponds to the dissociation constant for the nonspecific binding step during the nESI process; this variable also captures multimers of the ligand itself or nESI artifacts that encompass high concentrations of ligand trapped within individual droplets. Our Kn of 922.7 ± 259.9 μM suggests that an additional binding site on BIg-*Rb*CBM74-BIg is highly unlikely (**Fig. 2.10A**).

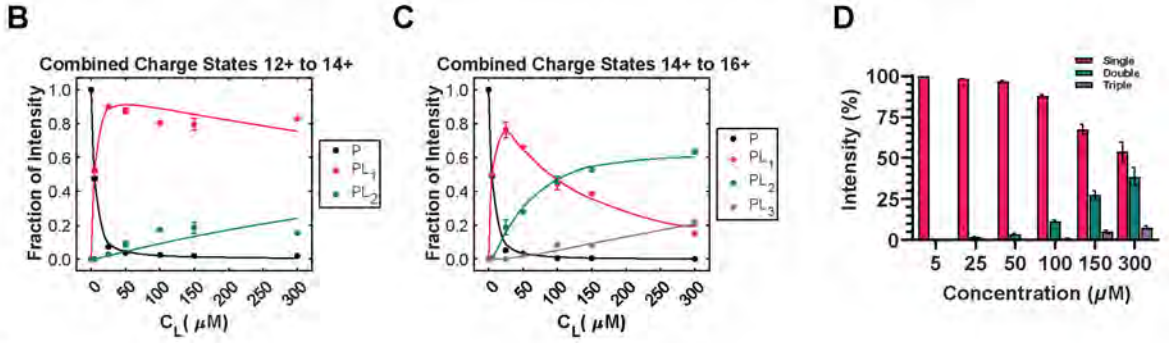
For Sas6T the binding state distribution was markedly different (**Fig. 2.10A,C**). At low G10 concentrations, there is a mix of 1-bound and 2-bound states, and as G10 increases the 2-bound fraction dominates. K_d values for 1:1 and 1:2 protein: ligand complexes were calculated to be 2.30 ± 0.25 μM and 104.64 ± 8.63 μM, respectively, in reasonable agreement with ITC data for BIg-*Rb*CBM74-BIg and *Rb*CBM26 alone (**Fig. 2.10A**). These data best support a model whereby *Rb*CBM26 and *Rb*CBM74 each bind one molecule of G10 independently. Since longer maltooligosaccharides form double helices in solution, we performed high resolution MS with G10 which showed a wide range of charged state distributions corresponding to single, double, and triple helical structures depending on concentration (**Fig. 2.10D**). G10 forms double as well as triple helices at high concentrations (300 μM), the latter of which may be an artifact of the ESI process, or by double helix formation from overlapping G10 molecules . While we could not resolve peaks from higher concentrations of G10, we can conclude that at 1mM, the concentration of G10 used in crystallization experiments, most of the ligand forms double helices. However, as *Rb*CBM74 does not absolutely require a double helix but rather α-glucan that adopts the correct geometry, it is not

A

G10

P	Charge States	K_{d1} (μM)	K_{d2} (μM)	K_n (μM)[a]	SSR[b]
BIg-*Rb*CBM74-Big	12+ to 14+	2.16 ± 0.53	N/A	922.73 ± 259.93	0.0375
Sas6	14+ to 16+	2.30 ± 0.25	104.64 ± 8.63	737.48 ± 142.57	0.0094

[a]K_n - dissociation constant for nonspecific binding step during nESI. Values reported with 95% confidence interval. [b]SSR - sum of squared residuals



E

G14

P	Charge States	K_{d1} (μM)	K_{d2} (μM)	K_n (μM)[a]	SSR[b]
BIg-*Rb*CBM74-Big	12+ to 14+	1.29 ± 0.10	N/A	125.50 ± 5.13	0.0030
Sas6	14+ to 16+	0.17 ± 0.03	64.44 ± 4.83	494.80 ± 63.10	0.0083

[a]K_n - dissociation constant for nonspecific binding step during nESI. Values reported with 95% confidence interval. [b]SSR - sum of squared residuals

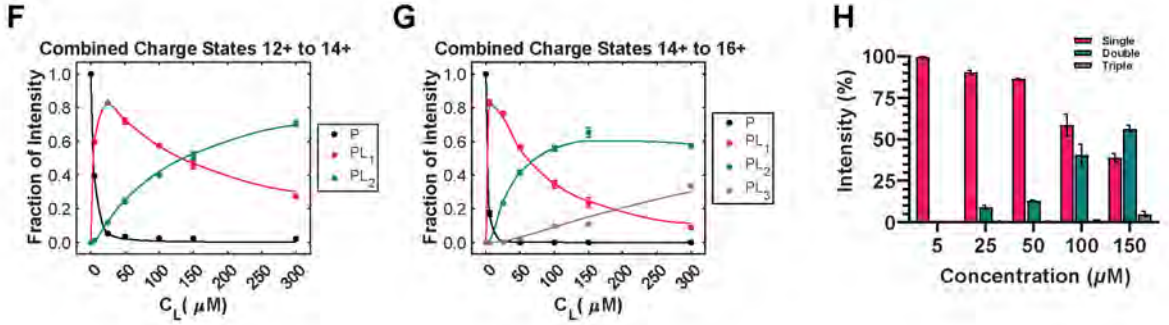


Figure 2.10: RbCBM74 and RbCBM26 bind separate molecules of G10 in solution

Native MS of 5μM protein and ligand. Intensities of each species, combined across multiple charge states, were extracted from the mass spectra and used to calculate the fractional abundance of unbound and bound states at equilibrium (n=3). **A.** Binding affinities (K_d) calculated from the fractional intensity of each species for G10. **B.** Nonlinear least-squares fitting of fractional abundance of unbound and bound states for 0-300μM G10 with BIg-*Rb*CBM74-BIg and **C.** Sas6T **D.** Isotopic distribution of single, double, and triple helices over different concentrations of G10. **E.** Binding affinities (K_d) calculated from the fractional intensity of each species for G14. **F.** Nonlinear least-squares fitting of fractional abundance of unbound and bound states for 0-300μM G14 with BIg-*Rb*CBM74-BIg and **G.** Sas6T. **H.** Isotopic distribution of single, double, and triple helices over different concentrations of G14.

surprising to see our high affinity binding site saturated by a single G10 via both ITC and native mass as this is the more abundant species at low concentrations.

To investigate *Rb*CBM74 binding to longer species, as this protein binds solubilized potato amylopectin with ~10-fold better affinity than G10, we performed native MS with G14. Here, the binding state distribution for BIg-*Rb*CBM74-BIg demonstrated that the 2-bound state becomes the dominant species at higher G14 concentrations (**Fig. 2.10F**). BIg-*Rb*CBM74-BIg exhibits a modestly higher affinity for G14 (K_d = 1.29 ± 0.10 μM) (**Fig. 2.10E,F**). For Sas6T, we observed a binding state distribution similar to that with G10 (**Fig. 2.10G**). Notably, we observed higher affinities for G14 for both 1-bound and 2-bound states (K_d values of 0.17 ± 0.03 μM and 64.44 ± 4.83 μM, respectively) (**Fig. 2.10E**). This suggests that the full binding platform of *Rb*CBM74 extends beyond what we see in our crystal structure. Our lower K_n for both constructs with G14 may be an artifact of the nESI process, as described above, rather than an additional binding site, although we cannot completely rule this out. The higher affinity observed for Sas6T over the *Rb*CBM74 may be because binding by the longer ligand is aided by the juxtaposition of the CBM26, which then becomes saturated at higher concentrations. Thus, it is possible there is some synergy in binding at the two sites that occurs with longer ligands, though further work is needed to fully investigate this possibility. High resolution mass spectrometry of G14 alone demonstrates both single and double helical populations, with more than half of the ligand forming double helices at 150 μM (**Fig. 2.10H**). As these states are in equilibrium, we cannot test binding to a single vs double helical structure. However, the *Rb*CBM74 site clearly selects for maltooligosaccharides that adopt the geometry found in low energy double helical α1,4-linked glucose.

2.4 DISCUSSION

CBMs are distinct protein domains that assist with substrate breakdown by specifically binding polysaccharide targets. These domains are especially important for binding to insoluble substrates like crystalline cellulose and semi-crystalline starch granules. The CBM74 family binds insoluble starch and its constituents, amylose and amylopectin. CBM74s are frequently (81/99 sequences) encoded adjacent to another starch-binding CBM family, either a CBM25 or CBM26 (223). Sas6 includes both a CBM26 and a CBM74 that have different affinities for maltooligosaccharides but work together to bind granular starch. *Rb*CBM26 has a canonical binding platform that accommodates motifs found in linear and circular maltooligosaccharides. In contrast, *Rb*CBM74 has an extended ligand binding groove that requires at least 8 glucose residues and accommodates the geometry specific to double helices found in amylopectin. Although our data suggests binding to single helices as well, the dimensions of the binding platforms preclude binding of single helices that adopt the wider V-amylose geometry (**Fig. 2.6G**). Because it is on the cell surface, the CBM74

of Sas6 may target *R. bromii* to the crystalline amylopectin regions of starch granules that are not easily accessible to human or other bacterial amylases.

Sas6 is a putative *R. bromii* amylosome component and likely cooperates with amylases and pullulanases via the interaction of its dockerin domain with a cohesin from a scaffoldin protein (160). It could bind cell anchored scaffoldins Sca2 or Sca5, associate with Sca1/Amy4, or bind to an as-of-yet unknown scaffoldin (160). Breakdown of starch by *R. bromii* relies on the coordinated effort of approximately 40 distinct proteins, of which Sas6 may play an integral part by specifically targeting the helical regions of starch (160). Amylopectin within starch granules is so tightly packed that multiple hydrogen-bonding interactions stabilize interactions between adjacent double helices (172; 104; 105). The binding cleft on *Rb*CBM74 is quite shallow (Fig 3A) such that we speculate *Rb*CBM74 recognition of these double helices could occur without disrupting the crystalline architecture. Our data using maltooligosaccharides as long as 14 glucose units suggests that individual molecules bind at *Rb*CBM26 and *Rb*CBM74. However, it is possible that longer α-glucan chains, such as amylopectin α1,4-linked regions which can span up to 60 glucose units, or even amylose chains that adopts the proper geometry, would allow these modules to dock to the same helical structure (95). While the CBM26 and CBM74 do not recognize α1,6 branch points, these motifs make up a relatively small proportion of amylopectin and thus are unlikely to interfere with protein docking.

Whether Sas6 aids in localizing the organism and its enzymatic machinery to the granule or if the protein has a more integral role in aiding catalysis by unwinding/disrupting the crystalline structure of the granule is unknown. At this moment we favor the idea that the primary function of this domain is simply targeting or docking to starch granules, as *Rb*CBM74 seems to recognize the native shape of the double helical amylopectin. However, it is possible that as the protein docks to the granule this results in local disruption of the crystalline network that aids in starch degradation.

Unlike *R. bromii*, resistant starch-utilizing Bifidobacteria encode CBM74-containing multimodular extracellular amylases (36). A recent study looked at the amylases that were differentially encoded between Bifidobacterial strains that could bind and degrade starch granules and those that could not (111). Resistant Starch Degrading enzyme 3 (RSD3) was differentially encoded in the resistant starch-binding strains. It contains a CBM74 and has high activity on high amylose corn starch. RSD3 has an N-terminal GH13 followed by CBM74, CBM26, and a CBM25. The CBM74-CBM26 motif is present in RSD3 so the structural and functional insights we have gleaned from Sas6 may suggest how these CBMs structurally assist the enzyme with granular starch hydrolysis.

Although starch is a polymer composed solely of glucose, there is massive variation in granule structure (20; 172). This is a function of primary structure (i.e. α1,4 or α1,6 linkages), secondary structure (single or double helices) and tertiary structure (helical packing and amylose content), making granules an exquisitely complex substrate (179). This complexity is unlocked by only a few

specialized gut bacteria, making granular starch a targeted prebiotic (36; 226; 15). CBM74s might serve as a molecular marker for the ability to break down resistant starch in metagenomic samples (223). Furthermore, CBM74s might make attractive additions to engineered enzymes for enhanced starch degradation on the industrial scale, or as an adjunct to starch prebiotics. The structural and functional picture of *Rb*CBM74 here will accelerate the targeted use of this domain for various health and industrial applications.

2.5 METHODS

2.5.1 Recombinant Protein Cloning and Expression

We used a previously described cloning and expression protocol to generate each of the recombinant protein constructs used in this study (31). Genomic DNA was isolated from *R. bromii* strain L2-63 and the constructs for Sas6 without the signal peptide with overhangs complementary to the Expresso T7 Cloning & Expression System N-His pETite vector (Lucigen). The forward primers were engineered to include the 6x His sequence that complemented the vector plus a TEV protease recognition site for later tag removal. PCR was performed with Flash PHUSION polymerase (ThermoFisher). The amplified products and the linearized N-his pETite vector were transformed in HI-Control10G Chemically Competent Cells (Lucigen) and plated on LB plates supplemented with 50 μg/ml kanamycin (Kan). Transformants were screened for the insertion of Sas6 and validated via sequencing. The Sas6-pETite plasmids were transformed into chloramphenicol (Chl)-resistant *E. coli* Rosetta (DE3) pLysS cells and plated on LB plates supplemented with 50μg/ml Kan and 20μg/mL Chl. *E. coli* cells were grown at 37°C to OD600 0.6-0.8 in Terrific Broth supplemented with 50μg/ml Kan and 20μg/ml Chl after which time the temperature was lowered to 20°C and 0.5mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) was added. After 16 hours of growth, 1L of cells was centrifuged, resuspended in 40mL of Buffer A (20mM Tris pH 8.0, 300mM NaCl) and lysed by sonication. Cell lysate was separated from cell debris by centrifugation for 30min at 30,000xg. 3mL of Ni-NTA resin was packed into Econo-Pac Chromatography Columns (BioRad) and equilibrated with Buffer A. Lysate was passed through the packed columns and washed with 70mL of Buffer A. Proteins were eluted from the columns via stepwise increase in Buffer B (20mM Tris pH 8.0, 300mM NaCl, 500mM imidazole). Proteins eluted in 10-25% Buffer B fractions. TEV protease (1mg) was added to each protein to initiate cleavage of the His-tag and the mix was dialyzed overnight using dialysis tubing (SpectraPor) in 1L of storage buffer (20mM HEPES pH 7, 100mM NaCl). The dialyzed protein-TEV mixture was applied to Ni-NTA resin and the flow-through was collected and concentrated using a VivaSpin 20 concentrator (Fisher Scientific).

2.5.2 Sas6 Immunofluorescence

Custom α-Sas6T antiserum was generated by rabbit immunization with purified recombinant Sas6T protein (Lampire Biological Laboratories). The resulting antiserum was used for western blotting and cell staining. *R. bromii* cells were grown to mid-log phase on RUM media (240) with 0.1% potato amylopectin and 2mL of the cell culture was collected for immunostaining and western blotting. For immunostaining, 1mL of *R. bromii* culture was centrifuged for 1min at 13,000xg and washed 3 times with 1X phosphate buffed saline pH 7.4 (PBS). 2μL of cells were then spread on a glass slide and fixed with 10% formaldehyde in PBS. Slides were washed 3x in PBS to remove fixative but were not permeabilized. Cells were blocked for 30min with 10% goat serum (Jackson ImmunoResearch). α-Sas6T antiserum was diluted 1:1000 in 10% goat serum and applied for 1hr to cells at room temperature (ctrl samples incubated in blocking buffer instead). The primary antiserum was removed, and slides were washed 3 x 5min in PBS before the application of 1:500 goat α-rabbit AlexaFluor488 antibody (ThermoFisher) for 30min. Slides were washed 3 x 5min in PBS and preserved with Prolong Gold Antifade reagent and dried overnight before imaging. Cells were imaged at the University of Michigan Microscopy Core on a Leica Stellaris Light Scanning Confocal microscope with a 100X objective.

2.5.3 Western Blotting

R. bromii was grown to mid-log phase overnight on RUM media containing 0.1% potato amylopectin (240). 1mL of cells was pelleted and washed twice in phosphate buffered saline (PBS) pH 7.4, then resuspended to a final volume of 50μL in 5mM Tris-HCl pH 8.5. The culture supernatant was passed through a 0.2μm filter and 50μL was reserved for analysis. Proteins were precipitated from the remaining supernatant by the addition of $\frac{1}{4}$ volume of 100% trichloroacetic acid (TCA) and incubated 30 min on ice. The precipitate was collected via centrifugation and washed twice with 200μL cold acetone. The resulting pellet was dried and resuspended in 50μL of 5mM Tris-HCl pH 8.5. Samples were separated by SDS-PAGE on two 10% Tris-glycine gels, then transferred to polyvinylidene difluoride (PVDF) membrane. Blots were blocked in EveryBlot Blocking Buffer (BioRad) for 30min then washed with PBS pH 7.4 + 0.05% Tween 20 (PBST). To detect Sas6, one membrane was incubated with custom rabbit α-Sas6 antiserum (Lampire) diluted 1:500 and the other with custom rabbit α-glutamic acid decarboxylase from *R. bromii* (Lampire) diluted 1:10,000 in PBST + 5% non-fat dry milk (PBST-milk) for 1hr. Blots were washed in PBST and incubated in horse radish peroxidase-conjugated goat α-rabbit antibody (ThermoFisher) diluted 1:5,000 in PBST-milk and the signal was detected by ECL chemiluminescence (ThermoFisher).

2.5.4 Granular starch binding assays and adsorption depletion

Granular starch-binding assays were conducted with potato starch (Bob's Red Mill), corn starch (Sigma), wheat starch (Sigma), or Avicel (Fluka). Prior to use, all polysaccharides were washed 3x with an excess of assay buffer (20mM HEPES pH 7.0, 100mM NaCl) to remove soluble starch and oligosaccharides and prepared as a 50mg/mL slurry. 1mg (corn) or 5mg (potato) of starch slurry was aliquoted into 0.2mL tubes in triplicate, centrifuged at 2,000xg for 2 min and the supernatant was carefully removed. 100μL of protein ranging from 0.5μM-10μM protein was added to each starch and the tubes were agitated by end-over-end rotation at room temperature for 1hr. After centrifugation at 2,000xg for 2min, 20μL of the supernatant was removed for unbound protein concentration determination by absorbance at A280 using a ThermoFisher NanodropOne with three replicate measurements per sample. The remaining 80μL of supernatant was removed and set aside for SDS-PAGE gel analysis. The concentration of unbound protein remaining in the supernatant was used to determine the μmoles of protein bound per gram of starch which was plotted against the concentration of initial (free) protein to generate a binding curve (26). Overall affinity (K_d) and binding maximum (Bmax) was determined via a one-site binding model (specific binding) using GraphPad Prism version 9.2.0 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com) (26).

To assess the remaining starch granules for bound protein, the granules were washed three times with an excess of assay buffer by mixing and centrifugation, the final wash supernatant was removed, and 100μL of Laemmli buffer containing 1M urea was added to the starch pellet to denature any bound protein but keep the original volume consistent. To qualitatively determine the amount of unbound and bound protein, 10μL each of the wash supernatant and solubilized pellet fraction were run separately via SDS-PAGE. Bovine serum albumin was used as a negative control and to confirm unbound protein was sufficiently washed from the starch granules.

2.5.5 Polysaccharide Affinity PAGE

Non-denaturing polyacrylamide gels with and without potato amylopectin (Sigma), corn amylopectin (Sigma), potato amylose (Sigma), bovine liver glycogen (Sigma), pullulan (Sigma), or dextran (Sigma) to a final concentration of 0.1% polysaccharide were cast. All polysaccharides were autoclaved and amylose was solubilized by alkaline solubilization with 1M NaOH and acid neutralization to pH 7 with HCl (94). Sas6 protein samples were mixed with 6X loading dye lacking SDS. Gels were run concurrently for 4 hours on ice and subsequently stained with Coomassie (0.025% Coomassie blue R350, 10% acetic acid, and 45% methanol). Gels were imaged on a Bio-Rad Gel Doc Go imaging system. The distance between each band and the top of the separating gel were measured using ImageJ (195). The ratio of the distance migrated by each band was

determined to the distance the BSA band traveled. Binding was considered positive if the ratio was less 0.85 as previously described (39).

2.5.6 Isothermal Titration Calorimetry

All ITC experiments were carried out using a TA Instruments standard volume NanoITC. For each experiment, 1300μL of 25μM protein was added to the sample cell and the reference cell was filled with distilled water. The sample injection syringe was loaded with 250μL of the appropriate ligand concentration (0.5mM - 5mM) to fully saturate the protein by the end of 25 injections of 10μls. Titrations were performed at 25°C with a stirring speed of 250 rpm. The resulting data were modeled using TA Instruments NanoAnalyze software employing the pre-set models for independent binding and blank (constant) to subtract the heat of dilution. Note that exothermic heat release is denoted with an upward peak on this machine. For interactions with high affinity (c-value at 25μM protein greater than 5), no alterations were made to the model. If the calculated c value of an interaction fell below 5, the n value was set to 1 as indicated in the figure legend following the guidance for modeling low affinity interactions (217). For polysaccharide titrations, curves were modeled by varying the substrate concentration until n=1 such that the K_d represents the overall affinity for the construct (1).

2.5.7 Protein Crystallization

Crystallization conditions for α-cyclodextrin (2mM) bound (pdb 7UWW) and unliganded (pdb 7UWU) crystals of Sas6T were screened via 96-well sparse matrix screen (Peg Ion HT, Hampton Research) in a sitting drop vapor diffusion experiment at room temperature. Screens were set up using an Art Robbins Gryphon robot with 20mg/mL protein in a 3-well tray (Art Robbins 102-0001-13) using protein-to-well solution ratios of 2:1, 1:1, and 1:2. Small crystals were observed in 0.2M Potassium thiocyanate pH 7.0, 20% w/v Polyethylene glycol 3,350 (condition B2) and were further optimized by varying pH, PEG 3350 percentage, and potassium thiocyanate concentration. Crystals were microseeded with a crystal seeding tool (Hampton) in a sitting drop setup of 1.5μL drops with 2:1, 1:1, or 1:2 protein:well solution ratios. The optimal crystallization solution contained 0.3M Potassium thiocyanate pH 7.0, 24% PEG 3350 and 1mM Anderson-Evans polyoxotungstate $[TeW_6O_{24}]^{6-}$ (TEW) (Jena Biosciences) to improve crystal diffraction. Prior to data collection, crystals were cryoprotected in a mixture of 80% crystallization solution supplemented with 20% ethylene glycol then plunged into liquid nitrogen.

Crystallization conditions for maltodecaose-bound *Rb*CBM74 structure (pdb 7UWV) were generated from the construct lacking the CBM26 (BIg-*Rb*CBM74-BIg, residues 134-665) using 96-well sparse matrix screens. A crystalline mass observed in 60% v/v Tacsimate pH 7.0, 0.1M BIS-TRIS

propane pH 7.0 (Hampton Salt-Rx HT-well H12) was used to microseed an optimized solution containing 30% Tacsimate, 0.1M HEPES pH 7.0 and 2mM maltodecaose (CarboExpert). No additional cryo-protection was required prior to plunge freezing into liquid nitrogen.

2.5.8 Structure Determination and Refinement

X-ray data were collected at the Life Sciences Collaborative Access Team (LS-CAT) at Argonne National Laboratory's Advanced Photon Source (APS) in Argonne, IL. Data were processed at APS using autoPROC with XDS for spot finding, indexing, and integration followed by Aimless for scaling and merging (228; 112; 64). Intrinsic sulfur SAD phasing was used to determine the structure of Sas6T/α-cyclodextrin (7UWW) using AutoSol in Phenix (59; 214). Those coordinates were then used for molecular replacement in Phaser to determine the unliganded Sas6T (7UWU) and BIg-*Rb*CBM74-BIg/G10 (7UWV) structures (149). All three structures were refined via manual model building in Coot and refinement in Phenix.refine (60; 3). Metal ion identities were validated using the web-based CheckMyMetal (CMM) tool (243) (https://cmm.minorlab.org/). Carbohydrate models were validated using Privateer (4).

2.5.9 SEC-SAXS experiment

SAXS was performed at Biophysics Collaborative Access Team (BioCAT, beamline 18ID at APS) with in-line size exclusion chromatography (SEC-SAXS) to separate the sample from aggregates and other contaminants. Sample was loaded onto a Superdex 200 Increase 10/300 GL column (Cytiva), which was run at 0.6ml/min by an AKTA Pure FPLC (GE) and the eluate after it passed through the UV monitor was flown through the SAXS flow cell. The flow cell consists of a 1.0mm ID quartz capillary with ~20μm walls. A coflowing buffer sheath is used to separate the sample from the capillary walls, helping prevent radiation damage (119). Scattering intensity was recorded using a Pilatus3 X 1M (Dectris) detector which was placed 3.6m from the sample giving a q-range of 0.003Å-1 to 0.35Å-1. 0.7 s exposures were acquired every 1s during elution and data was reduced using BioXTAS RAW 2.1.1 (99). Within RAW, the Volume of Correlation (VC), molecular weight, and oligomeric state were determined (177; 174). Buffer blanks were created by averaging regions flanking the elution peak and subtracted from exposures selected from the elution peak to create the I(q) vs q curves used for subsequent analyses. The molecular weight was calculated by comparison to known structures (Shape&Size) (72). P(r) function was determined using GNOM (208). GNOM and Shape&Size are part of the ATSAS package (version 3.0) (143). High resolution structures were fit to the SAXS data using FoXS and flexibility in the high-resolution structures was modeled against the Multi-FoXS data (196).

2.5.10 Hydrogen–Deuterium eXchange Mass Spectrometry (HDX-MS)

HDX-MS experiments were performed using a Synapt G2-SX HDMS system (Waters), similar to previously reported (161). Deuteration reactions were incubated at 20°C for 15s, 150s, 1500s, and 15,000s in triplicate. 3μL of BIg-*Rb*CBM74-BIg alone or in the presence of G10 were diluted with 57μL of deuterated labeling buffer. Nondeuterated data were acquired by dilution with protonated buffer and fully deuterated data were prepared by dilution in 99% D2O, 1% (v/v) formic acid) for 48h at room temperature. Samples were measured in triplicate using automated handling with a PAL liquid handling system (LEAP), using randomized sequential collection with Chronos.

Following incubation, deuteration was quenched by mixing 50μL of the solution with 50μL of 100mM phosphate, pH 2.5 at 0.3°C. Immediately after the samples were quenched, 95μL of the sample was loaded onto an Acquity M-class UPLC (Waters) with sequential inline pepsin digestion (Waters Enzymate BEH Pepsin column, 2.1mm × 30mm) for 3min at 15°C followed by reverse phase purification (Acquity UPLC BEH C18 1.7μm at 0.2°C). Sample was loaded onto the column equilibrated with 95% water, 5% acetonitrile, and 0.1% formic acid at a flow rate of 40μL/min. A 7min linear gradient (5%–35% acetonitrile) followed by a ramp and 2min block (85% acetonitrile) was used for separation and directly continuously infused onto a Synapt XS using Ion Mobility (Waters). [Glu1]-Fibrinopeptide B was used as a reference.

Data from nondeuterated samples were used for peptide identification with ProteinLynx Global Server 3.0 (Waters). Full coverage of the protein was obtained, with the exception of the region from residues 289-296, where peptides were not detected. The filtered peptide list and MS data were imported into HDExaminer (Sierra Analytics) for deuterium uptake calculation using both retention time and mobility matching. Representative peptides were utilized for a final cumulative sequence coverage of 91.4%. Normalized deuterium uptake data was calculated for protein alone and with G10, and differential protection, defined as those regions with an average of 5% difference in deuteration between states over the 150-15000s timepoints, were mapped onto the crystal structure using PyMOL (Schrodinger).

2.5.11 Native Mass Spectrometry (MS)

Stock solutions of BIg-*Rb*CBM74-BIg and Sas6 were de-salted and solvent exchanged into 200mM ammonium acetate (pH 6.8 – 7.0) using Amicon Ultra-0.5mL centrifugal filters (MilliporeSigma) with a 10kDa molecular weight cut-off. Ten consecutive washing steps were performed to achieve sufficient desalting. The final concentrations of each protein stock solution after desalting were estimated via UV absorbance at 280nm. A stock solution of G10 was prepared by dissolving a known mass in 200mM ammonium acetate to achieve a final concentration of 200μM. For native MS titration experiments used to quantify K_d values, the concentration of protein was fixed at

5μM, and enough G10 was added to achieve final concentrations of 0, 5, 25, 50, 100, and 150μM. Protein-G10 mixtures were then incubated at 4°C overnight to achieve equilibration prior to native MS analysis. All native binding experiments were performed using a Q Exactive Orbitrap MS with Ultra High Mass Range (UHMR) platform (Thermo Fisher Scientific) (224). Each sample (~3μM) was transferred to a gold-coated borosilicate capillary needle (prepared in house), and ions were generated via direct infusion using a nano-electrospray ionization (nESI) source operated in positive mode. The capillary voltage was held at 1.2kV, the inlet capillary was heated to 250°C, and the S-lens RF level was kept at 80. Low m/z detector optimization and high m/z transfer optics were used, and the trapping gas pressure was set to 2. In-source trapping was enabled with the desolvation voltage fixed at -25V for improved ion transmission and efficient salt adduct removal. Transient times were set at 128ms (resolution of 25,000 at m/z 400), and 5 microscans were combined into a single scan. A total of ~50 scans were averaged to produce the presented mass spectra. All full scan data were acquired using a noise threshold of 0 to avoid pre-processing of mass spectra. A total of three measurements for each ligand concentration were performed. Data were then processed and deconvoluted using UniDec software (147).

2.5.12 K_d Measurements by Native MS.

We performed titration experiments for both BIg-*Rb*CBM74-BIg and Sas6T using G10 and acquired modeled titration curves. Each bound state differed by ~1639 Da, which agrees with the theoretical mass of G10. To obtain the binding constants, we summed the peak intensities of all abundant charge states in our mass spectra. K_d values were calculated using the relative intensities of unbound protein and each ligand bound species from the mass spectra as previously described (89). Briefly, the protein-ligand binding equilibrium of BIg-*Rb*CBM74-BIg with G10 in solution can be described by the following reversible reaction:

$$\begin{array}{ccc} \mathbf{P} & \overset{L}{\rightleftharpoons} & \mathbf{PL} \\ \uparrow\downarrow L & & \uparrow\downarrow L \\ \mathbf{P}l & & \mathbf{PL}l \end{array} \tag{2.1}$$

where L is the ligand and **P** and **PL** are the free protein and protein with one specifically bound ligand, respectively. BIg-*Rb*CBM74-BIg possesses one ligand-binding site, *Rb*CBM74. As the concentration of ligand is increased, ligand molecules can bind nonspecifically during the nESI process, generating artifactual peaks in the mass spectra corresponding to a two ligand-bound complex. As the concentration of ligand is increased, ligand molecules can bind nonspecifically during the nESI process, generating artifactual peaks in the mass spectra corresponding to a two

ligand-bound complex. Here, we presume that nonspecific binding arises equally for free protein and that which possesses one specifically bound ligand represented by $\mathbf{P}l$ and $\mathbf{PL}l$ in Eq.~(2.1). Based on these assumptions, the equations of mass balance and binding states can be described the following system of equations:

$$c_p = [\mathbf{P}] + ([\mathbf{PL}] + [\mathbf{P}l]) + [\mathbf{PL}l] \tag{2.2a}$$

$$c_L = [L] + ([\mathbf{PL}] + [\mathbf{P}l]) + 2[\mathbf{PL}l] \tag{2.2b}$$

$$K_d = \frac{[\mathbf{P}][L]}{[\mathbf{PL}]} \tag{2.2c}$$

$$K_n = \frac{[\mathbf{P}][L]}{[\mathbf{P}l]} = \frac{[\mathbf{PL}][L]}{[\mathbf{PL}l]} \tag{2.2d}$$

where c_P and c_L represent the total concentrations of protein and ligand, respectively, and concentrations in brackets represent those at equilibrium. K_d and K_n represent the dissociation constants for specific and nonspecific binding steps, respectively. If 1) the peak intensities of free protein and ligand-bound complexes are proportional to the abundances of those in solution and 2) the spray and detection efficiency of all species is the same, then the fractional intensities of each species can be determined:

$$F_i = \frac{\sum_n I(\mathbf{PL}_i^{n+})/n}{\sum_{i=0}^{2}\sum_n (\mathbf{PL}_i^{n+})/n} \tag{2.3}$$

Here, the fractional intensities are calculated as the sum of the intensities of main peak ions at all charge states. Since a Fourier transform MS method is utilized, signal intensities are proportional to both ion abundance and charge state. Therefore, the ion intensities are normalized for each charge state, n (232; 13). These fractional intensities can be calculated from the titration experiment at each ligand concentration and can then be related to the equilibrium constants:

$$F_0 = \frac{K_d K_n}{K_d K_n + [L](K_d + K_n) + [L]^2} \tag{2.4a}$$

$$F_1 = \frac{[L](K_d + K_n)}{K_d K_n + [L](K_d + K_n) + [L]^2} \tag{2.4b}$$

$$F_2 = \frac{[L]^2}{K_d K_n + [L](K_d + K_n) + [L]^2} \tag{2.4c}$$

$[L]$ can also be determined from nESI-MS titration data:

$$[L] = c_L - c_P(F_1 + 2F_2) \tag{2.5}$$

$[L]$ was then obtained at each ligand concentration and applied to the Eqs.~(2.4a)-(2.4c). Equations (2.4a)-(2.4b) were then fitted to experimental fractional intensities using nonlinear least-squares curve fitting using the lsqnonlin.m. function in MATLAB. A more detailed derivation of these equations is provided elsewhere (89), along with the approach utilized for Sas6 which possesses two sites for specific binding (*Rb*CBM74 and *Rb*CBM26) and exhibits a third nonspecific bound state as shown here:

$$\begin{array}{ccccc} \mathbf{P} & \overset{L}{\rightleftharpoons} & \mathbf{PL} & \overset{L}{\rightleftharpoons} & \mathbf{PL_2} \\ \uparrow\downarrow L & & \uparrow\downarrow L & & \uparrow\downarrow L \\ \mathbf{P}l & & \mathbf{PL}l & & \mathbf{PL_2}l \end{array} \tag{2.6}$$

2.5.13 High-resolution MS

Stocks of G10 or G14 were diluted to 5, 25, 50, 100, 150, and 300 μM with 200 mM ammonium acetate. Similar tuning parameters to those in native MS experiments were used with a few exceptions. All experiments were performed in negative mode, and low m/z detector optimization and low m/z transfer optics were used. We chose negative mode for our experiments since positive mode spectra were heavily adducted with common cations present in solution. In-source trapping was 0 and -25 V for G10 and G14, respectively. Here, G14 required more activation to assist in sufficient desolvation. We noticed that higher energies would generate excessive fragments of both G10 and G14. Transient times were set at 1024 ms (resolution of 200,000 at m/z 400), and ~6 scans were averaged to produce the presented mass spectra.

An overlap of monoisotopic m/z peaks corresponding to single, double, and triple helices was observed at high concentrations. To approximate the relative abundance of each oligomeric state, we first simulated the isotopic distribution of each state using enviPat and then calculated the theoretical proportion of the monoisotopic species with respect to the proceeding peak that corresponds to a difference in carbon-13 composition (138). We then manually utilized these proportion factors to approximate the intensities of each oligomeric state in our experimental data. Calculations were performed for 6 individual scans and averaged.

2.5.14 Sequence collection

Amino acid sequences of CBM74 modules were collected according to information in the CAZy database (http://www.cazy.org/) yielding 29 sequences (CAZy update: March 2022) (54). This set was subsequently completed with sequences of hypothetical CBM74s based on protein BLAST searches (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using the CBM74 sequences from Sas6 of *Ruminococcus bromii* (GenBank Acc. No.: PKD32096.1) and the GH13_32 α-amylase from *Mi-*

crobacterium aurum (GenBank Acc. No.: AKG25402.1) as queries (160; 221; 6). In total, three searches with each query sequence were performed, limiting the searched databases to taxonomy kingdoms of Bacteria, Archaea and Eucarya (with no relevant results for the latter two). To capture a wide spectrum of organisms harboring a CBM74 module, one non-redundant amino acid sequence was selected to represent each species and/or bacterial strain. The BLAST searches thus yielded 93 additional CBM74 sequences of bacterial origin; the last sequence taken being the CBM74 module of a putative α-amylase from uncultured *Eubacterium* sp. (GenBank: SCJ65691.1; E-value: 3e-39). That preliminary set of 122 sequences was reduced by eliminating 23 sequences due to their redundancy and/or incompleteness of the CBM74 module. The final set of CBM74 modules consisted of 99 sequences. All sequences were retrieved from the GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and/or UniProt (https://www.uniprot.org/) databases (194; 220).

2.5.15 Sequence comparison and evolutionary analysis

The alignment of 99 CBM74 modules from the final set was performed using the program Clustal-Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (198). Only a subtle manual tuning of the computer-produced alignment was necessary to perform to maximize sequence similarities. The evolutionary tree of these 99 sequences was calculated by a maximum-likelihood method (on the final alignment including the gaps) using the WAG substitution model and the bootstrapping procedure with 500 bootstrap trials implemented in the MEGA-X package (233; 66; 124). The calculated tree file was displayed with the program iTOL (133) (https://itol.embl.de/). From both the alignment and the tree of all 99 sequences, a sample of 33 representative CBM74s was selected for a simplified alignment and tree. The structural comparison was created using the above-mentioned alignment in conjunction with the web-based CONSURF tool (10; 9; 34).

2.5.16 Data availability

The X-ray structures and diffraction data reported in this paper have been deposited in the Protein Data Bank under the accession codes 7UWU, 7UWV and 7UWW. The SAXS data are deposited in the small angle X-ray scattering database (SASDB) under the accession code SASDPE2 (116). All mass spectrometry data will be made available upon request.

Table 2.1: Sas6 Structure Statistics

Construct	Sas6T + α-cyclodextrin	Sas6T unliganded	BIg-*Rb*CBM74-BIg + G10
PDB Accession	7UWW	7UWU	7UWV
Wavelength	0.979	0.979	0.979
Resolution range	35 - 1.61 (1.67 - 1.61)	44.77 - 2.19 (2.268 - 2.19)	62.48 - 1.70 (1.76 - 1.70)
Space group	P 21 21 21	P 21 21 21	P 21 21 2
Unit cell	69.45 82.53 213.47 90 90 90	69.17 82.37 213.32 90 90 90	69.69 160.071 67.858 90 90 90
Total reflections	1690626 (29309)	1044914 (104226)	512772 (51008)
Unique reflections	122358 (2727)	63042 (6261)	84071 (8261)
Multiplicity	13.8 (10.7)	16.6 (16.9)	6.1 (6.2)
Completeness (%)	76.8 (24.8)	99.34 (99.97)	99.9 (100.0)
Mean I/sigma(I)	16.0 (1.1)	22.66 (14.34)	17.2 (2.2)
R-merge	0.092 (1.799)	0.09561 (0.192)	0.052 (0.75)
R-meas	0.095 (1.889)	0.09871 (0.1979)	0.057 (0.81)
R-pim	0.025 (0.563)	0.02433 (0.04781)	0.023 (0.33)
CC1/2	0.999 (0.549)	0.998 (0.993)	0.999 (0.822)
Reflections used in refinement	122315 (3895)	63244 (6262)	84061 (8260)
Reflections used for R-free	6093 (201)	3200 (309)	4056 (427)
R-work	0.168 (0.238)	0.197 (0.276)	0.179 (0.279)
R-free	0.212 (0.281)	0.255 (0.348)	0.199 (0.281)
Number of non-hydrogen atoms	11425	10523	4954
macromolecules	9721	9527	4043
ligands	253	38	237
solvent	1451	964	674
Protein residues	1294	1246	531
RMS(bonds)	0.013	0.001	0.013
RMS(angles)	1.43	0.44	1.7
Ramachandran favored (%)	96.3	96.1	97
Ramachandran allowed (%)	3.6	3.9	3
Ramachandran outliers (%)	0.08	0	0
Rotamer outliers (%)	0	1.57	0.46
Clashscore	3.26	4.98	0.24
Average B-factor	22.2	25.59	33.2
macromolecules	20.7	25.35	31.8
ligands	35.8	29.93	31.5
solvent	30	27.79	41.8

No.[a]	Protein[b]	CAZy[c]	GH13[d]	Organism[e]	Source material[f]	GenBank[g]	UniProt[h]	Length[i]	CBM74[j]	Additional CBMs[k]
1	pAAMY	Yes	GH13_28	*Bifidobacterium adolescentis*	Korean adult feces; HG	AZH71984.1	A0A809K2S0	1462	657-978	2×CBM25; CBM26
2	pAAMY		GH13_28	*Bifidobacterium adolescentis*	Human feces; HG	AJE06470.1	A0A0B5BPU9	1432	621-942	2×CBM25; CBM26
3	pAAMY			*Bifidobacterium ruminantium*	Human feces; HG	MBU9112168.1		1200	636-957	CBM26#
4	pAAMY		GH13_28	*Bifidobacterium angulatum*	Human feces; HG	AMK57563.1	A0A126SUD8	1527	629-950	2×CBM20; CBM26
5	pAAMY			*Bifidobacterium merycicum*		SHE84896.1	UPI0009218F34	1186	657-978	CBM26#
6	pAAMY			*Bifidobacterium tsurumiense*	Hamster dental plaque; RoG	KFJ06122.1	A0A087EEC1	1361	650-971	CBM25#
7	pAAMY		GH13_28	*Bifidobacterium pseudocatenulatum*		BAR04166.1	UPI0001847757	1434	621-942	CBM20; CBM26
8	pAAMY	Yes	GH13_28	*Bifidobacterium pseudolongum*	*Sus scrofa* cecum; PG	UBZ05046.1		1530	644-956	CBM25
9	pAAMY		GH13_28	*Bifidobacterium choerinum*	Rumen fluid of *Bos taurus*	ATU19782.1	A0A2D3D2U4	1613	644-956	CBM25
10	pAAMY			*Bifidobacterium gallicum*	Huaman adult intestine; HG	KFI59086.1	A0A087AJY6	1612	648-960	CBM25#
11	pAAMY			*Bifidobacterium dolichotidis*	*Dolichotis patagonum* feces; RoG	RSX55520.1	A0A430ERN0	1228	635-946	CBM25#
12	pAAMY			*Bifidobacterium castoris*	*Castor fiber* feces; RoG	RSX43996.1	A0A430F482	1439	633-944	CBM25#
13	pAAMY			*Pseudoscardovia suis*		PJI62656.1	UPI000CB6C43C	1455	621-934	CBM26#
14	pAAMY			*Pseudoscardovia radai*	*Sus scrofa* digestive tract; PG	OZG53567.1	A0A261F364	2041	624-937	CBM26#
15	pAAMY			*Bifidobacterium minimum*	Wastewater	KFI73654.1	A0A087BRK4	1457	631-943	2×CBM25; CBM26#
16	pAAMY			*Bifidobacterium thermophilum*	Swine feces; PG	KFJ07018.1	UPI000505012E	1642	604-917	CBM26#
17	pAAMY			*Bifidobacterium goeldii*	Feces of tamarin; PrG	RSX51755.1	A0A430FFT3	1522	632-946	2×CBM20; CBM25#
18	pAAMY			*Bifidobacterium pullorum*		WP_204464541.1	UPI001956DCA0	1251	626-939	CBM25#
19	pAAMY			*Gallilscardovia ingluvei*		GGI14461.1	UPI001666D4DC	1473	650-963	3×CBM25#
20	HYPO			*Prevotella* sp.	HG	MBD9271786.1	UPI001DD49E24	677	373-677	CBM26#
21	pAAMY			*Prevotella hominis*	Human feces; HG	TFH82515.1	A0A4Y8VPY7	1095	791-1095	CBM26#
22	pAAMY			*Prevotella* sp. CAG:386	HG	CDC29131.1	R6QE37	1082	777-1082	CBM26#
23	pAAMY			*Prevotella copri*	Human feces; HG	MBV3413028.1		1092	787-1092	CBM26#
24	pDOC			Oscillospiraceae bacterium	Equine fecal microbiome	MBQ0097748.1		575	45-344	CBM26#
25	pDOC			*Candidatus Scatavimonas merdigallinarum*	*Gallus gallus*; CG	HIQ81347.1		714	235-533	CBM26#
26	pDOC			*Eubacterium* sp. CAG:202	HG	CDC03229.1	R6N4I8	671	175-505	
27	pDOC			*Eubacterium* sp. OM08-24	Human feces; HG	RGM19132.1	A0A374UMY1	737	241-571	CBM26#
28*	DOC	Yes		*Ruminococcus bromii*	HG	PKD32096.1	A0A2N0UYM2	734	242-572	CBM26
29	pDOC	Yes		*Ruminococcus bovis*	Rumen fluid of *Bos taurus*; RuG	QCT06902.1	A0A4P8XW45	730	240-570	CBM26
30	pDOC			*Candidatus Scatovicinus merdipullorum*	*Gallus gallus*; CG	HIR02602.1		949	435-759	CBM26#
31	pDOC			Clostridia bacterium	Goat gastroinestinal tract; RuG	MBQ2687794.1		985	431-752	CBM26#
32	pAAMY			*Succinimonas* sp.	Goat gastrointestinal tract; RuG	MBQ3681472.1		1799	543-836	CBM26#
33	pAAMY			*Succinimonas amylolytica*		WP_019000730.1	UPI000380E766	1786	533-826	3×CBM26#
34	pAAMY			*Ruminobacter* sp.	Roe deer gastrointestinal tract; RuG	MBR1924924.1		2051	808-1101	CBM26#
35	pAAMY			Succinivibrionaceae bacterium	Dairy cattle gastrointestinal tract; RuG	MBQ5525197.1		2157	941-1232	
36	pAAMY			*Ruminobacter amylophilus*		SFP12024.1	A0A662ZIG8	1959	721-1015	CBM26#
37	pAAMY			*Vibrio vulnificus*		MBN8105319.1	UPI0019D43A75	2050	694-982	CBM26#
38	pAAMY			*Vibrio navarrensis*	Cattle; RuG	MBE4579468.1	UPI001869786D	2476	687-975	CBM26#
39	pAAMY			*Vibrio cincinnatiensis*		WP_238130312.1		2465	694-983	CBM26#
40	pDOC			Fusobacteria bacterium	Anaerobic digestion of organic wastes under variable temperature conditions and feedstock	NLK63198.1	A0A7X8J1U3	599	307-599	
41	pDOC			*Orenia marismortui*		TDX52525.1	A0A4R8IH020	804	508-804	
42	HYPO			Spirochaetales bacterium	Sheep gastrointestinal tract; RuG	MBR2317321.1		838	533-838	
43	HYPO			*Treponema* sp.	Water deer gut; RuG	MBP3562042.1	UPI001B74B407	879	573-879	

Figure 2.11: CBM74 Sequences

44	pAAMY		*Microbacterium ugarici*			PFG29411.1		A0A2A9DSR8	1398	1098-1398	2×CBM25#
45	pAAMY		*Microbacterium lindanitolerans*			TQO22419.1		UPI00117OBA3D	1398	1098-1398	2×CBM25#
46	pAAMY		*Microbacterium lindaniningi*			WP_166982566.1		UPI0014IDC7DD	1401	1101-1401	2×CBM25#
47	pAAMY		*Arthrobacter pigmenti*			NIC24392.1		A0A846RV66	1402	1102-1402	2×CBM25#
48	pAAMY		*Microbacterium* sp. A18JL241			WP_194420059.1		UPI0018893ADD	1124	825-1124	2×CBM25#
49	pAAMY		*Agromyces atrinae*			NYD67366.1		A0A4Q2M4N3	1310	1010-1310	CBM25#
50	HYPO		*Microbacterium* sp. BWT-G7		Larvae of *Trypoxylus dichotomus* (Japanese rhinoceros beetle)	MCC2033550.1			361	61-361	
51	pAAMY		*Arthrobacter stackebrandtii*			MBP2412725.1		UPI001D5C3EE1	2095	982-1283	3×CBM25#
52*	AAMY	Yes	GH13_32	*Microbacterium aurum*	Wastewater treatment plant from a potato starch factory	AKG25402.1		A0A0G2T4B5	1417	1116-1417	2×CBM25
53	HYPO			Aeromonadaceae bacterium	Wastewater	MBP8172431.1		UPI001B401EC6	701	400-701	
54	pAAMY	Yes	GH13_32	*Sanguibacter* sp. HDW7	*Hydrophilus acuminatus* (dark diving beetle)	QIK84188.1		A0A6G7Z5A5	1153	853-1153	2×CBM25
55	HYPO			*Streptomyces* sp. NBRC 109706		WP_062214241.1		UPI0007823FAB	355	51-353	
56	HYPO			Porphyromonadaceae bacterium	Anaerobic digestion of organic wastes under variable temperature conditions and feedstock	NLO70176.1		UPI0016A25B17	613	23-335	CBM26#
57	pDOC			Flavobacteriales bacterium	Activated sludge from Hong Kong Shatin wastewater treatment plant	MCB9201997.1			604	30-336	CBM26#
58	pAAMY	Yes	GH13_28	*Clostridium bornimense* M2/40		CDM67953.1		W6RUH9	1725	778-1086	CBM26
59	HYPO			*Clostridium saudiense*	*Gallus gallus* gut; CG	MBM6820797.1		UPI00195A2FDA	545	54-359	
60	HYPO			Lachnospiraceae bacterium		HAU84760.1		A0A349YHJ5	709	75-374	3×CBM25#
61	pAAMY			*Candidatus Mediterraneibacter cottocaccae*	*Gallus gallus* gut; CG	HIA65501.1			1527	848-1145	CBM26#
62	pAAMY			*Lachnoclostridium* sp. An76		WP_162611201.1		UPI0013A610D6	1424	829-1126	CBM26#
63	pAAMY			*Eubacterium rectale*	Human feces; HG	RGW40301.1		A0A413BHP0	913	225-523	CBM26#
64	HYPO			*Clostridium* sp. MSJ-8	*Macaca fascicularis* feces; PrG	MBU5488563.1			720	46-344	
65	pAAMY			*Pseudoruminococcus massiliensis*		WP_106763003.1		UPI000D106684	1475	787-1092	CBM26#
66	pAAMY			uncultured *Eubacterium* sp.		SCJ65691.1		A0A1C6I798	1518	877-1180	CBM26#
67	pDOC	Yes		*Ruminococcus bovis*	Rumen fluid of *Bos taurus*; RuG	QCT07491.1		A0A4P8XWH7	648	225-528	CBM26
68	pAAMY	Yes	GH13_28	*Streptococcus suis*	Tonsil scrape of *Sus scrofa*; PG	AWX97480.1		A0A2Z4PIT7	1636	763-1063	4×CBM26
69	HYPO			*Lachnoclostridium* sp. MSJ-17	*Macaca fascicularis* feces; PrG	MBU5462054.1			921	39-341	
70	pDOC			Ruminococcaceae bacterium P7		SCX01552.1		A0A1G4V5K7	925	39-340	
71	HYPO			*Clostridium* sp.	Human feces; HG	MBS6600272.1			811	51-347	
72	HYPO			Muribaculaceae bacterium	Porcine feces; PG	MBS7352324.1			524	28-333	CBM26#
73	pAAMY			*Clostridium* sp. CAG:411	HG	CDE46637.1		R7I7Z7	1517	833-1132	CBM26#
74	pAAMY			*Barnesiella* sp.	*Myodes glareolus* feces; RoG	MBD5235098.1		UPI0019A69732	1110	755-1059	
75	pDOC			Paludibacteraceae bacterium	Sheep gastrointestinal tract; RuG	MBR3872341.1			570	208-513	
76	pAAMY			*Candidatus Amulumruptor caecigallinarius*	*Gallus gallus*; CG	RXI23900.1		A0A4Q0U9L8	1131	753-1054	
77	pAAMY			*Bacteroides* sp.	*Myodes glareolus* feces; RoG	MBD5349186.1		UPI0019C8840C	1269	905-1217	2×CBM26#
78	HYPO			*Breznakibacter* sp.	Wastewater	MBP8849315.1		UPI001B6D11F8	454	138-454	
79	HYPO			*Alistipes* sp.		WP_026957379.1		UPI0004073BE6	994	688-994	CBM26#
80	HYPO			*Alkalitalea saponilacus*		WP_026970432.1		UPI00047A61D5	994	688-994	CBM26#
81	pAAMY	Yes	GH13_19	*Paenibacillus sonchi*	Soil	QQZ60411.1		A0A7U1I386	1583	1296-1583	2×CBM25; CBM26
82	pAAMY			*Paenibacillus jilunlii*		SDL85044.1		A0A1G9NF80	1585	1298-1585	2×CBM25; CBM26#
83	pAAMY			*Paenibacillus borealis*	Milk	OMD46076.1		A0A1R0YDQ8	2351	2064-2351	3×CBM25; CBM26#
84	pAAMY			*Paenibacillus zeisolI*	Soil	RUT36321.1		A0A433XQE6	2641	2354-2641	2×CBM25; CBM26#
85	pAAMY	Yes	GH13_19	*Paenibacillus donghaensis*	Marine sediment	ASA20374.1		A0A222K6E7	1578	1291-1578	2×CBM25; CBM26
86	pAAMY			*Paenibacillus anaericanus*	*Aporrectodea caliginosa* (earthworm) gut	RUT47280.1		A0A3S1DTY1	2567	2280-2567	2×CBM25; CBM26#

Figure 2.11: CBM74 Sequences

87	pAAMY		*Fontibacillus panacisegetis*	Soil	SDG38409.1	A0A1G7TUK6	2442	2153-2442	2×CBM25; 2×CBM26#
88	pAAMY		*Paenibacillus* sp. MMS18-CY102	Soil	MWC30933.1	A0A7X3GNM1	1675	1388-1675	3×CBM25; CBM26#
89	pAAMY		*Paenibacillus curdlanolyticus*	Soil	EFM08800.1	E0IF24	1677	1390-1677	3×CBM25; CBM26#
90	pAAMY	Yes	GH13_19	*Cohnella* sp. KS 22	Human blood	QMV43648.1	A0A7G5C364	1587	1300-1587	2×CBM25; CBM26
91	pAAMY	Yes	GH13_19	*unidentified bacterium* UGO163		CAA37453.1	Q03658	1684	1397-1684	3×CBM25; CBM26
92	HYPO			Bacilli bacterium	anaerobic digestion of organic wastes under variable temperature conditions and feedstock	HHU20570.1	A0A7V6LK53	426	139-426	
93	pAAMY			Lentisphaerae bacterium GWF2_57_35	Rifle well CD01 at time point 6/F; 5m depth; 0.2 filter	OGV41563.1	A0A1G0Y6L0	1484	1194-1484	CBM25#
94	HYPO			Kiritimatiellae bacterium	Wastewater	MBP9573123.1	UPI001B4DD1FE	871	586-871	
95	pAAMY			Verrucomicrobia bacterium	Rare earth elements-acid mine drainage contaminated river water	MBU6182917.1	UPI001C2983BD	1518	1213-1518	
96	HYPO			Myxococcales bacterium	Activated sludge from Hong Kong Shatin wastewater treatment plant	MCA9546954.1	UPI001DE396C8	419	19-305	
97	HYPO			Myxococcota bacterium		MBU1432708.1		471	20-309	
98	HYPO			Candidatus Sumerlaeia bacterium	Bioreactors innoculated with microbial mats from alkaline soda lake	MCC5876423.1		1755	181-462	CBM25; 2×CBM53#
99	HYPO			Deltaproteobacteria bacterium	Deep marine sediment from hydrothermal vent	MBW2735654.1		401	103-401	

[a] All CBM74s originate from bacterial proteins/enzymes. The two experimentally characterized CBM74s – the Nos 28 and 52 – are marked by an asterisk; the CBM74 from *Ruminococcus bromii* (No. 28) is central to the present study since its three-dimensional structure has been determined here.

[b] The abbreviations of proteins and enzymes: AAMY, α-amylase; pAAMY, putative α-amylase; DOC, dockerin-containing protein; pDOC, putative dockerin-containing protein; HYPO, unknown hypothetical protein.

[c] "Yes" means that the particular CBM74 has already been classified in the CAZy database.

[d] The GH13 subfamily affiliation for the source enzyme, if available.

[e] Organism – the source of the protein/enzyme containing the CBM74.

[f] Sources of isolation (if available): CG, chicken gut; HG, human gut; PG, pig gut; PrG, primate gut; RoG, rodent gut; RuG, ruminant gut.

[g] GenBank database Accession No.

[h] UniProt database Accession No. (if available).

[i] Length of the source protein.

[j] The boundaries of the particular CBM74 in the source protein/enzyme.

[k] Additional CBMs according to: (i) CAZy; or (ii) Pfam and/or InterPro (marked by #).

Figure 2.11: CBM74 Sequences

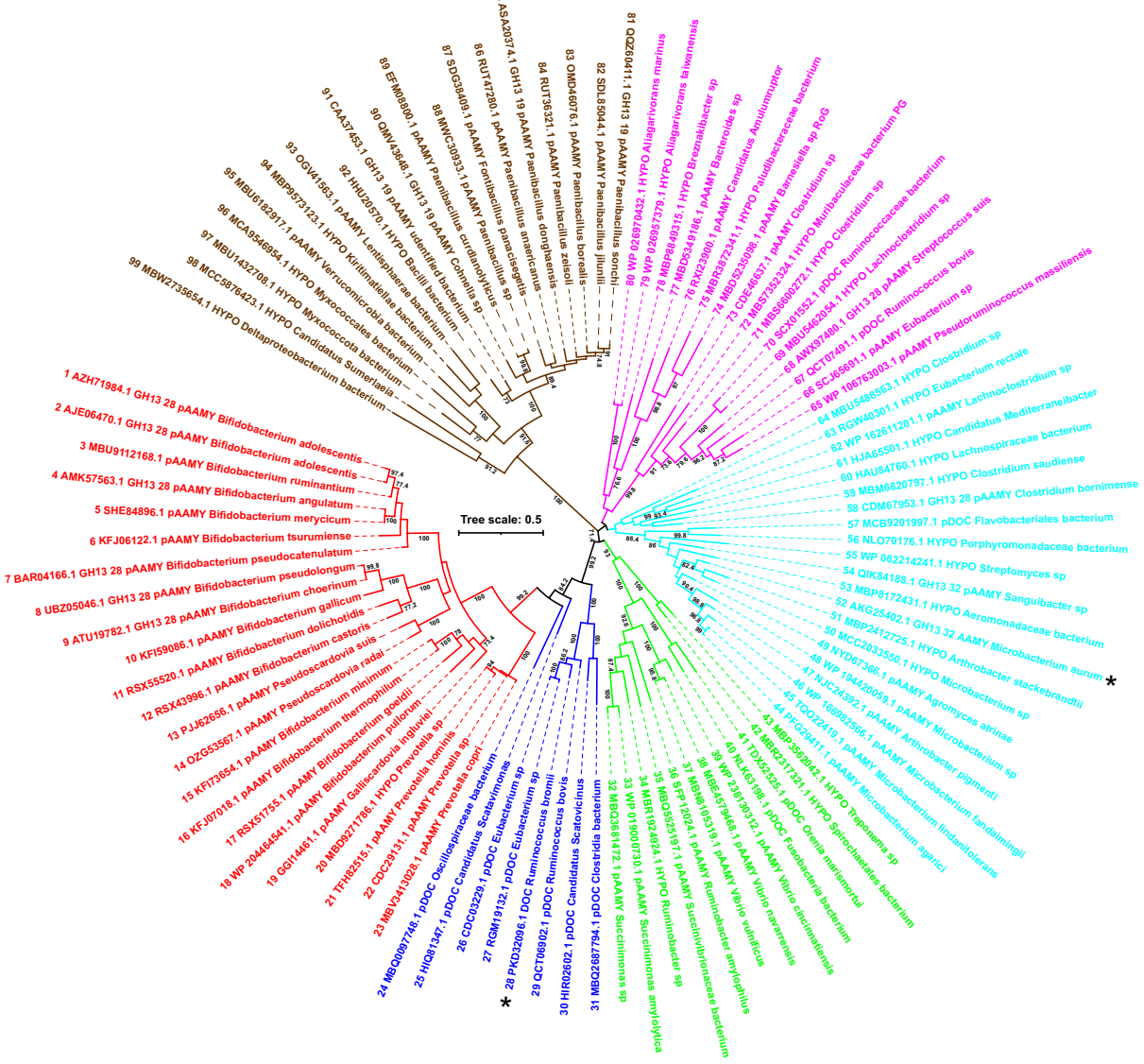


Figure 2.12: CBM74 Tree

Table 2.2: Phi-Psi Angles of G10 ligand

	*RB*CBM74		B starch		V amylose	
	φ	Ψ	φ	Ψ	φ	Ψ
GlcA1-2	76.3	-148.7	83.9	-144.2	102.2	-143.4
GlcA2-3	70.6	-147.1	84.2	-144.0	112.1	-108.9
GlcA3-4	57.1	-146.8	84.0	-144.3	98.3	-143.9
GlcA4-5	91.3	-142.2	84.0	-144.0	105.8	-146.3
GlcA5-6	97.9	-111.3	83.9	-144.2	103.1	-124.7
GlcA6-7	93.3	-139.3			101.5	-121.9
GlcA7-8	60.8	-160.0			104.0	-123.8
GlcA8-9	99.3	-134.8			105.0	-125.9
GlcA9-10	102.4	-136.1			105.7	-144.5
GlcA10-11	99.2	-146.3			103.0	-142.2
GlcA11-12	84.5	-142.1			109.0	-116.6
Average	84.8	-141.3	84.0	-144.1	104.5	-131.1
GlcB1-2	97.1	-141.7	83.9	-144.2		
GlcB2-3	84.7	-146.4	84.3	-144.0		
GlcB3-4	83.1	-150.7	83.9	-144.2		
GlcB4-5	104.0	-111.9	84.0	-144.0		
GlcB5-6	76.4	-147.9	83.9	-144.2		
GlcB6-7	74.3	-148.6				
GlcB7-8	100.2	-145.7				
GlcB8-9	88.7	-148.0				
Average	88.9	-142.7	84.0	-144.1		

Table 2.3: Sas6 Construct ITC parameters

	G3		ACX		G7		G8		G10		Potato AP
	n	Kd (μM)	n	Kd (μM)	n	Kd (μM)	n	Kd (μM)	n	Kd (μM)	Kd (μM)
Sas6T	1*	880 ± 25	1.0	178 ± 26	1*	332 ± 15	1*	496 ± 260	0.9	5.5 ± 1.9	0.3 ± 0.08
RbCBM26	-	NB	0.8	169 ± 16	1	310 ± 34	1*	285 ± 84	0.7	252 ± 128	NB
BIg-*Rb*CBM74-BIg	-	NB	-	NB	-	NB	1*	393 ± 136	0.85 ± 0.2	5.2 ± 1.1	0.7 ± 0.05
W373A									-	NB	13.4 ± 5.4
H289A									0.54 ± 0.09	73.1 ± 7.7	21.4 ± 3.1
F326A									0.70 ± 0.13	100 ± 11	3.9 ± 1.4

CHAPTER 3

BaAmy7 Is a Multi-Modular Resistant Starch-Degrading Amylase From *Bifidobacterium adolescentis*

3.1 INTRODUCTION

Bifidobacteria play a vital role in maintaining a healthy human microbiota (166). In particular, *Bifidobacterium adolescentis* has been shown to inhibit various enteric viruses (117; 134; 169; 130), possess anti-carcinogenic properties (65; 127), contribute to immune system development and regulation (76; 93; 247; 154; 135), lower serum cholesterol levels (235; 209), and alleviate antibiotic-associated diarrheal symptoms (92; 86; 189). Dietary fibers known as prebiotics have been identified as promoters of beneficial bacteria growth, with certain prebiotics specifically stimulating the growth of Bifidobacterial species (77). Resistant starch (RS) has shown promise as a bifidogenic prebiotic (7). RS refers to the portion of starch that remains undigested by human salivary and pancreatic amylases (36; 63). Type 2 resistant starches (RS2) are raw starches derived from plants like corn, wheat, and potato, characterized by their compact and indigestible granular structure (61; 21). Only two known microorganisms, *Ruminococcus bromii* and *Bifidobacterium adolescentis*, have been found to degrade RS2 in the human gut, and their relative abundance increases following RS2 consumption (241; 17; 226; 15; 68; 178). However, the specific protein machinery employed by *B. adolescentis* to break down RS2 remains unclear.

Granular starch, a glucose storage polymer produced by plants, consists of insoluble granules ranging from a few to 100 micrometers in diameter (187). These granules require boiling or acid treatment to disrupt their internal structure and become soluble in water. The insolubility arises from the alternating layers of amorphous and crystalline regions within the granules (20; 172). Starch is primarily composed of approximately 70% amylopectin and 30% amylose (141). Amylopectin is a polysaccharide with α-1,4 linkages and α-1,6 linkages every 15-80 residues, creating branching points that result in adjacent chains (95). When these chains reach a sufficient length (around

10 residues), they can form a double helix, leading to the tight packing of double helices in the crystalline regions and the banding pattern observed in the granule (20). On the other hand, amylose is almost entirely composed of α-1,4 linkages and forms a single helix with a complete turn every six glucose residues (80).

To degrade RS2, *B. adolescentis* utilizes enzymes that can bind to and hydrolyze both types of starch linkages (111). Carbohydrate binding modules (CBMs) mediate starch binding, while glycoside hydrolase family 13 (GH13) domains catalyze starch hydrolysis (157). Gut bacterial enzymes capable of accessing RS2 feature a CBM family 74 (CBM74) (223). CBM74 is a raw starch binding domain that is three times the size of most CBMs and typically co-occurs with a CBM family 25 (CBM25) or 26 (CBM26) (223). Previous studies have determined the crystal structure of a binding protein comprising a CBM74 and CBM26, revealing the ability of CBM74s to bind starch double helices found in the crystalline regions of starch granules, while CBM26 can bind short maltooligosaccharides from nearby regions (see Chapter 2).

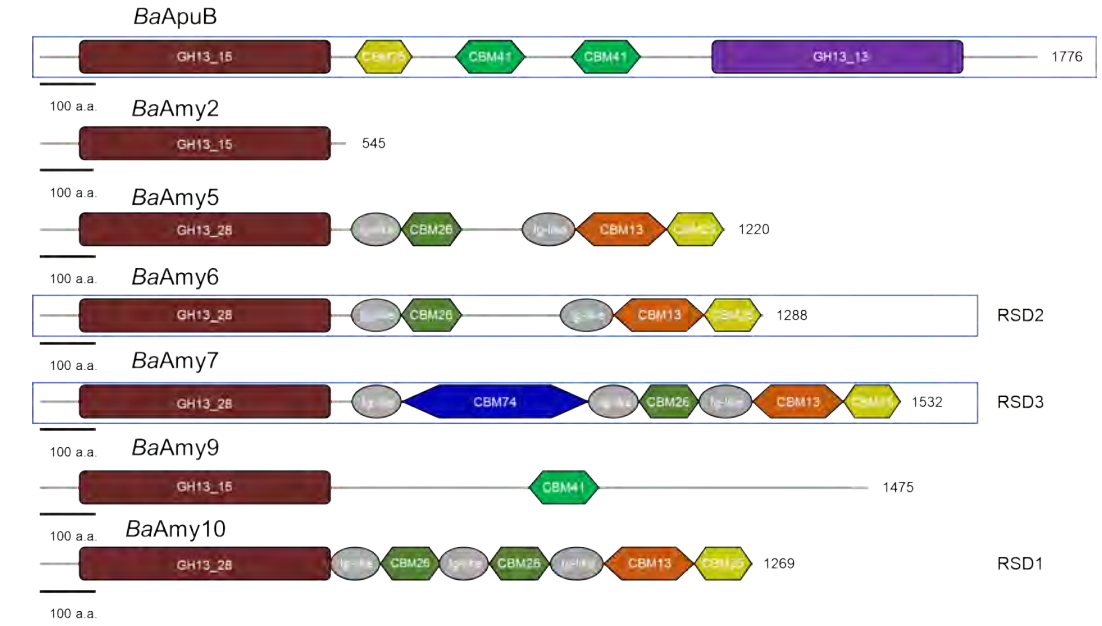
A similar motif is present in an extracellular amylase found in *B. adolescentis*, known as BaAmy7. BaAmy7 consists of four CBMs attached to the GH13. In this study, we recombinantly express BaAmy7, investigate its substrate preferences, and evaluate its activity on RS2. Furthermore, we analyze the binding contribution of each of the four CBMs to elucidate their roles in RS2 breakdown. These findings will enhance our understanding of how CBMs coordinate with GH13s to interact with granular starch and contribute to the design of amylases for industrial or biotechnical applications.

3.2 RESULTS

3.2.1 *Bifidobacterium adolescentis* L2-32 encodes 7 extracellular GH13s

B. adolescentis L2-32 encodes 12 GH13-containing proteins, here numbered BaAmy1-12. Of those 12, seven are predicted to be extracellular (**Fig. 3.1A**). BaAmy1 is a homolog of ApuB, a bifunctional α-amylase/pullulanase from *Bifidobacterium breve* UCC2003 (167). ApuB has a GH13 subfamily 15 (GH13_15) at the N-terminus which hydrolyzes α1,4-linkages in starch (amylase) while the C-terminal GH13_13 targets α1,6 branch points (pullulanase) (167). In the region between the two catalytic domains there are three CBMs: one CBM family 25 (CBM25) and two CBM41s (167). CBM41s come as tandem pairs and are typically associated with pullulanases (106). Both CBM25 and CBM41s bind starch, amylose, and amylopectin though CBM41, unlike CBM25, accommodates α1,6 in the canonical binding site (106). ApuB is active on amylopectin and pullulan as expected (118). However, the activity is variable on raw starches. ApuB binds to raw corn and potato starch but only breaks down corn starch and *soluble* potato starch but not raw

A



B

```
BaAmy5    961   YAELDALAKAHASDLEDGTYAVSTALKDGMVLDVADGSRKDGANVRLWSSNGTKAQQWTVSHDSKGYV
BaAmy6   1029   YAELDALAKAHASDLEDGTYAVSTALKDGMVLDVADGSRKDGANVRLWSSNGTKAQQWTVSHDSKGYV
BaAmy7   1272   YAELDALAKAHASDLEDGTYAVSTALKDGMVLDVADGSRKDGANVRLWSSNGTKAQQWTVSHDSKGYV
BaAmy10  1010   YAELDALAKAHASDLEDGTYAVSTALKDGMVLDVADGSRKDGANVRLWSSNGTKAQQWTVSHDSKGYV

BaAmy5   1029   TLRNVNSGKALDVKDGKAANGSNVHQYAPNSYRSQKWVAVRSGSVYKLVSALSPSMALDVKDGKAANG
BaAmy6   1097   TLRNVNSGKALDVKDGKAANGSNVHQYAPNSYRSQKWVAVRSGSVYKLVSALSPSMALDVKDGKAANG
BaAmy7   1340   TLRNVNSGKALDVKDGKAANGSNVHQYAPNSYRSQKWVAVRSGSVYKLVSALSPSMALDVKDGKAANG
BaAmy10  1078   TLRNVNSGKALDVKDGKAANGSNVHQYAPNSYRSQKWVAVRSGSVYKLVSALSPSMALDVKDGKAANG

BaAmy5   1097   SNVQIYTANGYRSQQWTFKTVAQPLKSATVWYRPSSAQSRVRVQWRVYGSPDTTGGMEMTQACGGWWK
BaAmy6   1165   SNVQIYTANGYRSQQWTFKTVAQPLKSATVWYRPSSAQSRVRVQWRVYGSPDTTGGMEMTQACGGWWK
BaAmy7   1408   SNVQIYTANGYRSQQWTFKTVAQPLKSATVWYRPSSAQSRVRVQWRVYGSPDTTGGMEMTQACGGWWK
BaAmy10  1146   SNVQIYTANGYRSQQWTFKTVAQPLKSATVWYRPSSAQSRVRVQWRVYGSPDTTGGMEMTQACGGWWK

BaAmy5   1165   ATVPSAGSTRVGLSFSYGSTTDDNGGKLYDVKGESAAVSGGQAVTDVTPNCAVTNK
BaAmy6   1233   ATVPSAGSTRVGLSFSYGSTTDDNGGKLYDVKGESAAVSGGQAVTDVTPNCAVTNK
BaAmy7   1476   ATVPSAGSTRVGLSFSYGSTTDDNGGKLYDVKGESAAVSGGQAVTDVTPNCAVTTKQ
BaAmy10  1214   ATVPSAGSTRVGLSFSYGSTTDDNGGKLYDVKGESAAVSGGQAVTDVTPNCAVTNK
```

Figure 3.1: *B. adolescentis* L2-32 encodes 7 extracellular GH13s
A. A schematic of 7 extracellular GH13-containing proteins. The domain sizes correspond to the provided scale. The three resistant starch degrading enzymes identified as differentially encoded among Bifidobacterial strains that adhere to granular starch are indicated by RSD1, 2, and 3 (111). The three boxed enzymes are included in the current work.
B. Sequence alignment of the C-terminal CBM13 and CBM25s from BaAmy5/6/7/10.

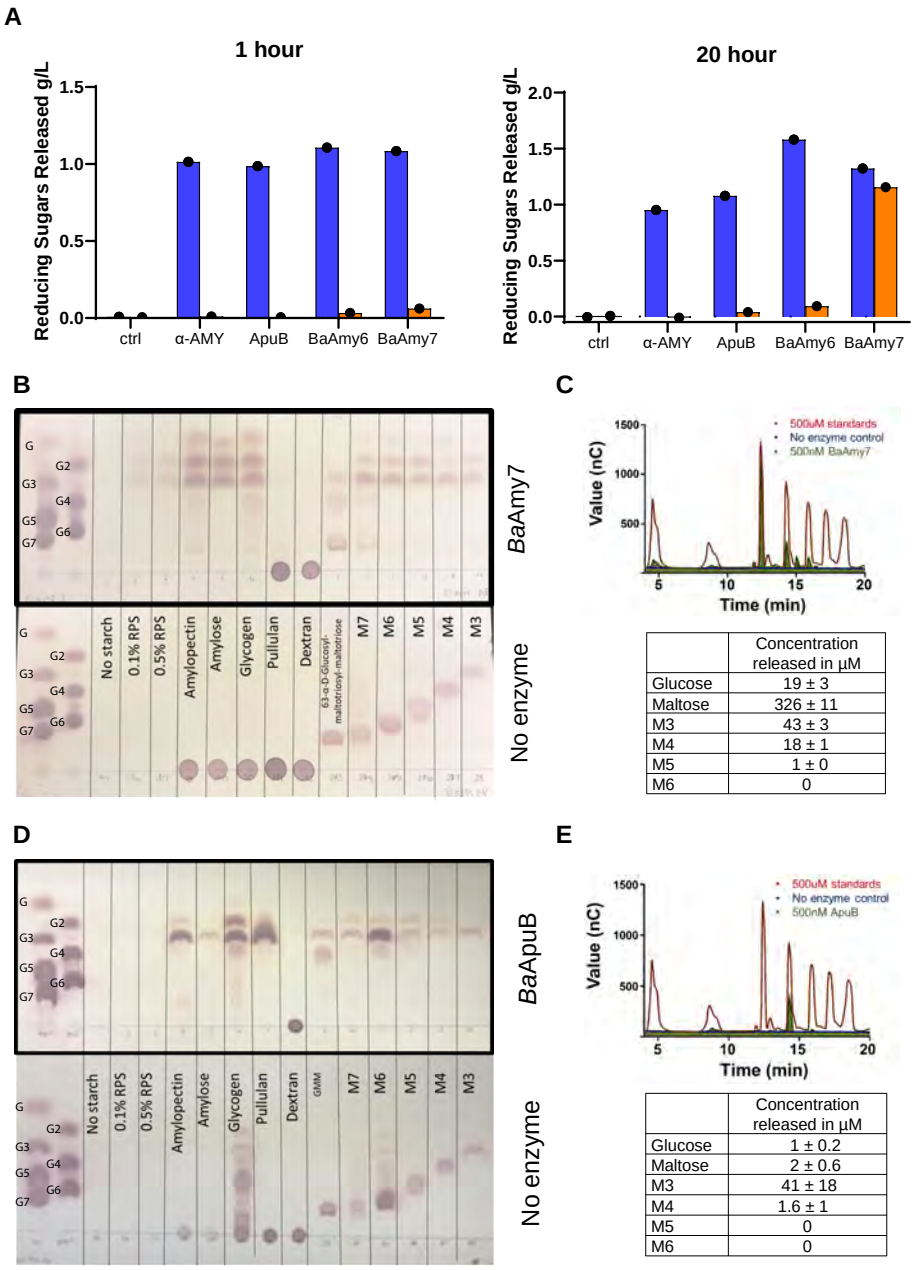
potato starch (118). ApuB is required for growth on starch by *B. breve* UC2003 highlighting its role in the starch utilization process by Bifidobacterial species though it is unlikely to be the critical enzyme for utilization of raw, granular starches (167).

Four of the extracellular amylases, BaAmy5, BaAmy6, BaAmy7, and BaAmy10 share similarities in their overall domain architecture. All four GH13s are from subfamily 28 (GH13_28), a subfamily that is prevalent within the RS-degrading Bifidobacterial species (www.cazy.org) (54). The number of GH13_28s in Bifidobacterial genomes ranges from 1-5 copies (54). Most *B. adolescentis* strains have 2-3 GH13_28 containing proteins suggesting that one of these four proteins may be absent in other *B. adolescentis* strains (54). Each of these four amylases contains at least one CBM26 and share 100% identity at the C-terminus (**Fig. 3.1B**). They differ in the CBMs between the GH13_28 and the C-terminal CBM13/25. BaAmy6, BaAmy7, BaAmy10 are particularly of interest as RS-active enzymes, as they are differentially encoded in Bifidobacterial strains that adhere to granular starch (RSD2, RSD3, and RSD1 respectively) (111). BaAmy5 and BaAmy6 encode a single CBM26, BaAmy10 encodes two CBM26s, and BaAmy7 is the longest of the four due to the presence of a CBM74. CBM74 is a unique domain that specializes in binding granular starch by accommodating starch double helices that are enriched in the crystalline regions (see 2.3.8). Based on the previously determined role of CBM74 in starch binding, we here hypothesize that CBM74 enhances the activity of BaAmy7 on raw starches.

3.2.2 Comparison of Amylase Activity - ApuB, BaAmy6, and BaAmy7

To better understand how CBM74 contributes to RS-degrading activity we first compared the activity of CBM74-containing BaAmy7, BaAmy6 which lacks a CBM74, ApuB, and porcine pancreatic α-amylase on potato amylopectin and potato starch. Potato amylopectin is an accessible substrate that can be broken down by α-amylases generally. In contrast, raw potato starch is a resistant starch that is recalcitrant to all but a few specialized α-amylases. In a preliminary experiment (n=1), we measured the products released on both substrates at 1 hour and at 20 hours. As expected, the activity of all four enzymes was similar at 1 hour on potato amylopectin with very little detectable product from the potato starch reaction by BaAmy6 and BaAmy7 (**Fig. 3.2A**). At 20 hours, when the soluble portion of the raw potato starch has been digested, the products released represent RS activity (19; 28). Despite BaAmy6 exhibiting more potato amylopectin breakdown than BaAmy7, BaAmy7 is the only enzyme to exhibit appreciable breakdown of potato starch. Out of these three extracellular amylases investigated, BaAmy7 is the primary enzyme capable of raw potato starch breakdown.

We next analyzed the sugars released by BaAmy7 hydrolysis of a panel of starch substrates using thin layer chromatography. By determining the breakdown products released from hydrolysis



	Concentration released in µM
Glucose	19 ± 3
Maltose	326 ± 11
M3	43 ± 3
M4	18 ± 1
M5	1 ± 0
M6	0

	Concentration released in µM
Glucose	1 ± 0.2
Maltose	2 ± 0.6
M3	41 ± 18
M4	1.6 ± 1
M5	0
M6	0

Figure 3.2: BaAmy7 is active on potato starch

A. Reducing sugars released from enzyme reactions at 37°C with 0.1% potato amylopectin (blue bars) or 1% potato starch (orange bars) at 1 hour or 20 hours. 5 Units of each enzyme were added to initiate the reaction (see methods). Note an n=1, awaiting replicates. **B.** Products generated by reaction of BaAmy7 with indicated polysaccharides after 24 hour at 37°C as determined by thin layer chromatography. Maltooligosaccharide standards ranging from glucose-maltoheptaose are shown on the left of the plate. **C.** Concentration of each maltooligosaccharide released after 3 days of BaAmy7 reaction with potato starch at 37°C determined by High-Performance Anion-Exchange chromatography with Pulsed Amperometric Detection (HPAEC-PAD). Concentrations of Mean±SD of 2 experiments (n=2) **D.** TLC plate of products released from BaApuB reaction with listed polysaccharides. **E.** HPAEC-PAD analysis of products released at 3 days from BaApuB potato starch reaction of n=2 experiments.

of polysaccharides and oligosaccharides we can gain insight into . BaAmy7 is active on raw potato starch, potato amylopectin, potato amylose, and glycogen but is not active on pullulan or dextran (**Fig. 3.2B**). BaAmy7 is able to hydrolyze oligosaccharides as short as 3 glucose units (maltotriose), including a 7 residue oligomer of maltotriose motifs linked by α1,6-branch points (6^3-α-D-Glucosyl-maltotriosyl-maltotriose). The primary hydrolysis products of oligosaccharides are maltose and maltotriose. The products of polysaccharide breakdown are glucose, maltose, and maltotriose. To quantify the different products of potato starch degradation, we allowed the reaction to proceed for 3 days to allow complete breakdown and used high performance anion exchange chromatography to detect specific products. The most abundant product from this reaction was maltose with lower concentrations of glucose, maltotriose (M3), and maltotetraose (M4) (**Fig. 3.2C**).

In contrast, BaApuB only breaks down maltooligosaccharides of at least four residues (M4). BaApuB is active on potato amylopectin, amylose, glycogen, and pullulan with the predominant product being maltotriose (**Fig. 3.2D**). At the 20 hour timepoint investigated by TLC, no products were detected from reaction with potato starch. At 3 days, using HPAEC which is a more sensitive approach, a small amount of maltotriose was detected, likely reflecting hydrolysis of the more accessible regions at the surface of granular potato starch (**Fig. 3.2E**). The TLC results for BaAmy7 and BaApuB on corn starch followed the same patterns as potato starch (data not shown). Together these results show that BaAmy7 has a specialized ability to degrade granular potato starch.

3.2.3 BaAmy7 CBM Binding

BaAmy7 has four predicted CBMs separated by inert bacterial Ig-like domains presumably for structural stability. We first sought to determine whether the predicted CBMs of BaAmy7 bind starch substrates. The two most N-terminal CBMs are from families CBM74 and CBM26 that are capable of granular and soluble starch binding and often appear together as a pair (165; 106; 26; 173). The region spanning amino acids 1272-1432 was annotated by the HMMER tool in dbCAN2 as a "CBM13" family (242) but remained unannotated in the CAZy database (54). If truly a CBM13, its presence in a GH13-containing enzyme would be unusual as CBM13 family members typically target xylan and galactan, but can accommodate monosaccharides including glucose (27). The most C-terminal domain is annotated as a CBM25, a granular and soluble starch binding family that often appears as a tandem repeat or adjacent to CBM26s (245). We performed a polysaccharide macroarray to qualitatively determine binding of His-tagged CBM constructs to several starch substrates (glycogen, potato amylopectin, maize amylopectin, amylose, pullulan) and dextran as a non-starch control. We expressed a catalytically inactive version of BaAmy7 with Glu311 mutated to a Gln (E311Q). E311 aligns with E208, part of the catalytic triad of the *Bacillus subtilis* α-

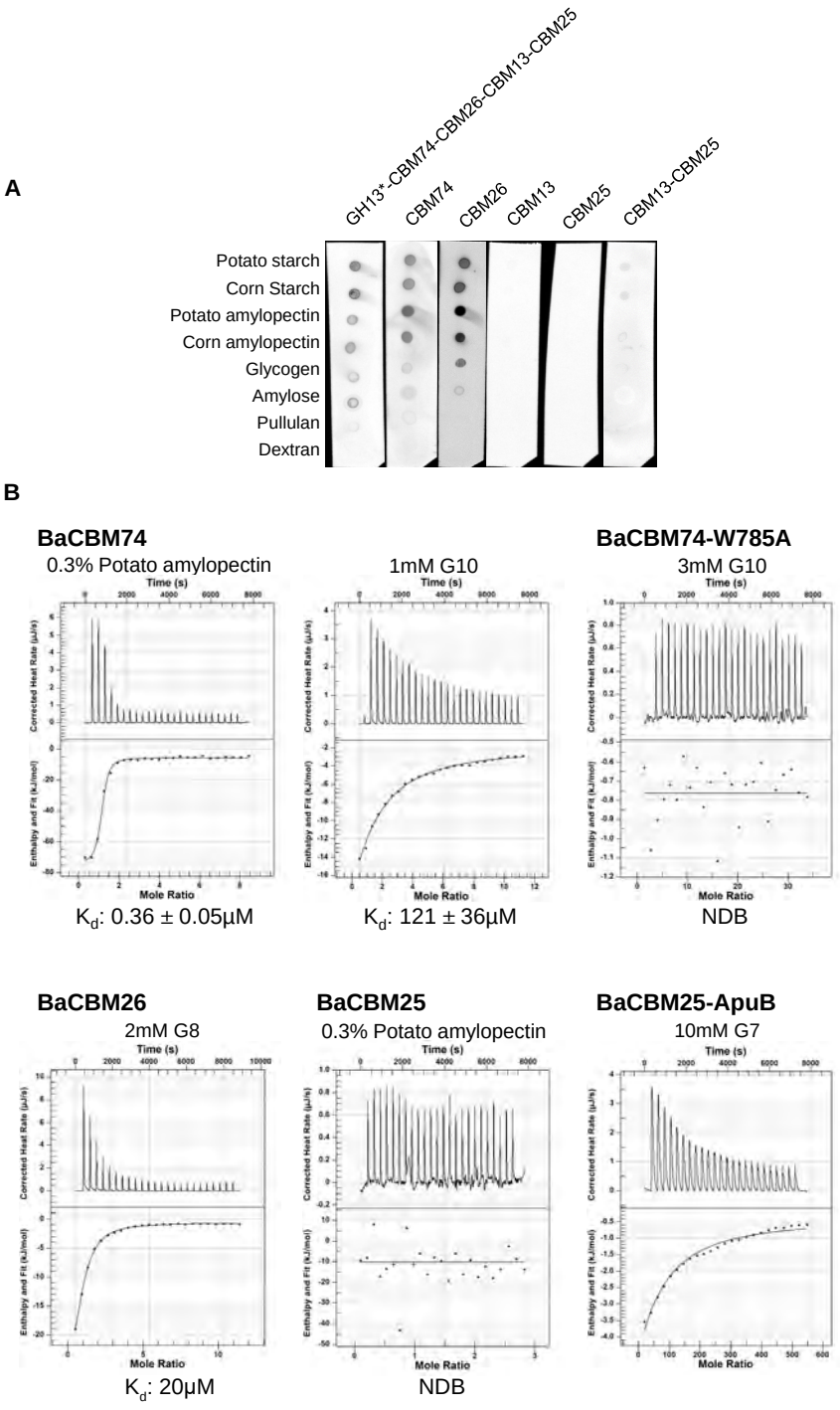


Figure 3.3: BaCBM74 and BaCBM26 drive starch binding by BaAmy7

A. Polysaccharide macroarray with dark dots representing purified recombinant protein binding to the specified polysaccharide. **B.** ITC quantification of 0.1% amylopectin (2 experiments) and G10 binding (3 experiments) by BaCBM74. The average K_d of 2-3 experiments is given below the panel. W785 which corresponds to W373 from Sas6 is hypothesized to be important for starch binding based on the homologous CBM74 from Sas6 (see chapter 2). BaCBM74-W785A is representative of 3 experiments. BaCBM25 is representative of 2 experiments. Note that upward peaks are exothermic in this machine.

amylase (74). We used catalytically inactive BaAmy7 (GH13*-CBM74-CBM26-CBM13-CBM25) to evaluate binding of the full protein including the GH13 domain, which may potentially harbor a surface binding site common with these domains (38). We found that of the 4 individual CBMs, only BaCBM74 and BaCBM26 bind starch polysaccharides (**Fig. 3.3A**). These two CBMs likely drive starch binding by a catalytically inactive full-length BaAmy7. Unexpectedly neither BaCBM13 nor BaCBM25 bind substrate on their own (**Fig. 3.3A**). Though there appears to be faint signal for the pairing of CBM13-CBM25 this has yet to be confirmed by quantitative measures.

We then quantified the binding of CBM74, CBM26, and CBM25 to starch substrates by isothermal titration calorimetry (ITC). Guided by the CBM74 from *Ruminococcus bromii* (2.3.8) we designed the BaCBM74 to include the bacterial Ig-like domains that flank it to provide structural stability and promote proper folding. BaCBM74 binds G10 (121μM) with 20-fold lower affinity compared to *Rb*CBM74 (5.2μM) but has a similar affinity for potato amylopectin (**Fig. 3.3B**)(0.4μM to 0.7μM, see 2.3). Based on the mutational studies in *Rb*CBM74 (2.3.10), we targeted W785 as a hypothesized critical binding residue and found that mutation to alanine eliminated binding to G10 (**Fig. 3.3B**). BaCBM26 also exhibited binding to potato amylopectin by ITC in a single experimental run (**Fig. 3.3B**). Consistent with the polysaccharide macroarray, BaCBM25 exhibited no binding to potato amylopectin (**Fig. 3.3B**). In contrast, the CBM25 from ApuB in the same organism has binding capacity to maltoheptaose, a 7-unit starch oligosaccharide (**Fig. 3.3B**).

3.2.4 The structure of BaCBM25

To understand the molecular basis of the lack of binding by BaCBM25, we sequentially and structurally compared BaCBM25 to other CBM25 family members. The consensus binding site for CBM25s is made up of two surface exposed tryptophans that extend from two loops that form a substrate platform on the surface of the protein (26; 106). When the sequence of BaCBM25 is aligned with the two structurally characterized CBM25s, there is only one tryptophan residue conserved (W1452) (**Fig. 3.4A**).

We next determined the crystal structure of BaCBM25 to 1.55Å resolution (**Table 3.1**) to compare the binding sites of BaCBM25 and other CBM25s. The top structural homologs are the two structurally characterized CBM25s: *Bh*CBM25 from *Bacillus halodurans* amylase in complex with maltotetraose (PDB 2c3x; RMSD: 1.7Å ; 19% seq identity) and *Pp*CBM25, the NMR structure of CBM25-1 of beta/alpha-amylase from *Paenibacillus polymyxa* (PDB 2laa; RMSD: 1.9Å; 20% seq identity). When the 3 CBM25s are aligned it becomes clear that the BaCBM25 binding site is completely disrupted (**Fig. 3.4B**). Both tryptophans are replaced by threonines (T1461 and T1496). Tryptophan W1452 that sometimes aligns with one of the binding site tryptophans (W34 from *Bh*CBM25), is facing the interior of the CBM fold. The structure of *Bh*CBM25 also reveals

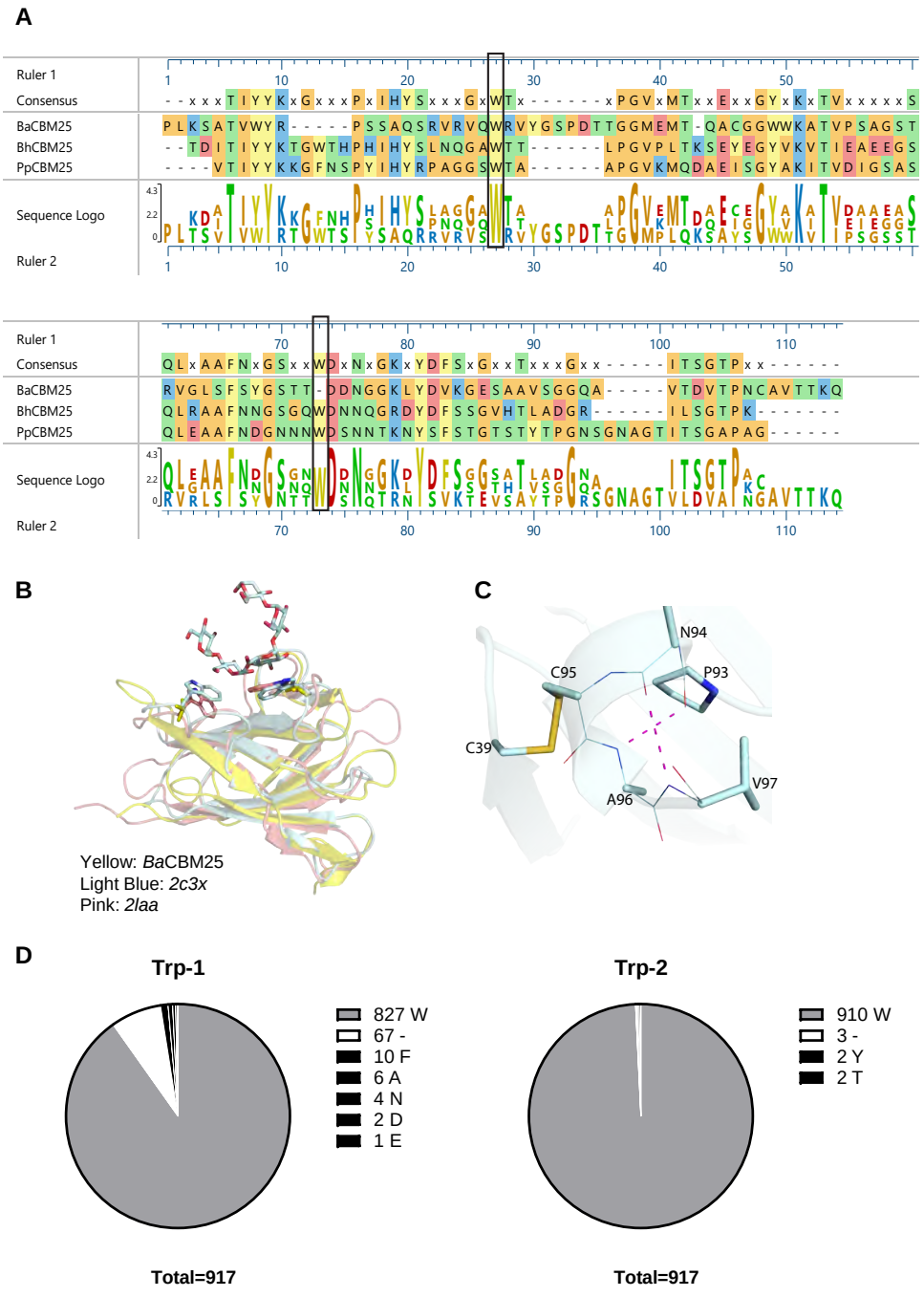


Figure 3.4: Structural basis for lack of BaCBM25 binding

A. A sequence alignment of BaCBM25 with two other structurally characterized CBM25s from *Bacillus halodurans* (PDB 2c3x) and *Paenibacillus polymyxa* (PDB 2laa). The black box indicates the normally conserved binding residues of CBM25s. **B.** A structural alignment of BaCBM25, *Bh*CBM25 (2c3x) and *Pp*CBM25 (2laa). The binding site residues are shown in stick representation. The blue ligand is maltotetraose bound to *Bh*CBM25 for reference. **C.** The C-terminal helix of BaCBM25 showing the hydrogen bonding and disulfide bonds that likely stabilize the C-terminus **D.** Pie chart distribution of variant residues encoded at the site of the Trp residues that make up the binding site among all CBM25 family members classified as part of this family in the CAzy database (cazy.org) (54).

a second binding site on the opposite face of the CBM. This site has a single Trp residue (W20) providing aromatic character. In BaCBM25, there is no corresponding aromatic residue in that position, although there is a Tyr residue from a different loop, Y1492, that occupies the space where W20 would be, but that site is entirely occluded by the loop spanning R1440-R1447. Together the structure of BaCBM25 suggests that it does not have a clear starch binding site.

Once AlphaFold became available we compared our experimentally determined structure to the AlphaFold2 prediction using the Google Colab notebook (156). The central β-sandwich that makes up the CBM25 fold was predicted with high accuracy. The major differences are in the low confidence loops (1428-1433; 1454-1460; 1480-1484; 1525-1532). These regions of low confidence turn out to be incredibly important for the overall arrangement of the protein. For example, the linker between BaCBM13 and BaCBM25, aa1428-1433, is predicted with low confidence but determines the available architecture between the two domains which could be important if they form a shared binding platform. The C-terminal region (1525-1532) was not predicted to form a helix like we see in the crystal structure (**Fig. 3.4C**). However, this helix is stabilized by a disulfide bond between C1470 and C1526 as well as hydrogen bonding between N1525 and V1528, and P1524 and A1527 respectively. This loop could be a Schellman loop that acts as a paperclip cap to stabilize the protein, which would explain why the removal of BaCBM25 results in decreased stability of the overall enzyme (46). Alternatively, it could mediate protein-protein interaction with another nearby amylase for cooperative activity since 3 of the extracellular amylases in *B. adolescentis* L2-32 have identical C-termini (**Fig. 3.1B**).

It is unclear whether the non-binding CBM25 pattern is a sub-classification within the CBM25 family or whether this CBM25 from *B. adolescentis* L2-32 is an outlier within the family. We therefore determined the number of putative CBM25s that lack the canonical binding site. Of the 917 sequences classified as CBM25s within the CAZy database (CAZy.org) 109 of the sequences belong to Bifidobacterial proteins. Only two proteins had the same predicted binding site motif as BaCBM25, both from *B. adolescentis* strain 6. Of the 80 sequences lacking an aromatic residue at the first site, 23 of them are Bifidobacterial proteins (**Fig. 3.4D**). The non-binding CBM25 is prevalent within Bifidobacteria. For example, of the 6 strains of *B. adolescentis* represented in this cohort, half of them have non-binding CBM25s. Within *B. pseudolongum* the non-binding CBM25s are more varied, but within those species at least one protein from each strain has a non-binding CBM25. This would suggest that the evolution of Bifidobacterial proteins containing non-binding CBM25s is a relatively recent offshoot within the *B. adolescentis* and *B. pseudolongum* species. At site 2, the only species lacking an aromatic are Gardnerella species and *Bifidobacterium adolescentis* species 6 and L2-32 (**Fig. 3.4D**). This suggests that the amylases from these two strains of *B. adolescentis* are closely related and unique within their clade. Intriguingly, of the other CBM25s encoded by *B. adolescentis* L2-32, all are missing the binding Trp residues, except the one

from BaApuB. The CBM25 from this protein aligns well with the other CBM25s but has two Trp in the canonical binding site and is binding competent (**Fig. 3.3B**). This domain is likely derived from a different evolutionary source than the ones in BaAmy5, BaAmy6, BaAmy7, and BaAmy10. Though BaCBM25 and BaCBM13 are non-binding domains, they may still play a different but important role in BaAmy7 activity.

3.2.5 Contribution of CBMs to Enzyme Activity

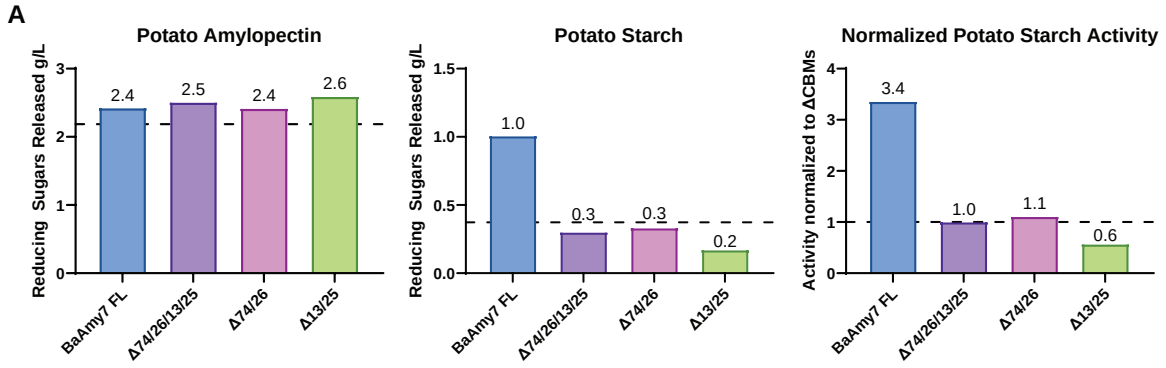


Figure 3.5: BaAmy7 CBMs required for potato starch breakdown
A. Enzyme reactions were set up with 0.3% potato amylopectin or 3% potato starch. Reactions were initialized with the addition of 5 Units of enzyme (see methods for catalytic unit determination). Reducing sugars released from each reaction were determined by DNSA assay (see methods). For panel 3, the potato starch activity was normalized to the GH13-only construct (Δ74/26/13/25). The dotted line represents the activity of porcine pancreatic $/alpha$-amylase (Megazyme).

To determine the contribution of the CBMs to enzymatic activity, we recombinantly expressed constructs of BaAmy7 lacking some or all of the CBMs and characterized their activities on starch using a reducing sugar assay to detect the abundance of products released from the reactions. We tested the potato starch degrading activity of full-length BaAmy7, a construct with just the GH13 (BaAmy7Δ74/26/13/25), one lacking the CBM74 and CBM26 (BaAmy7Δ74/26), and one lacking the C-terminal CBMs (BaAmy7Δ13/25). The following are the results of one experiment and require replication but we will here discuss the trends suggested by these data. We first determined the units of activity (mg protein required to release 1μmol product per minute) using CNP-G3, a colorimetric substrate 2- chloro-p-nitrophenol linked with maltotriose. CNP-G3 was used to measure amylase activity for each construct. We determined the activity of each construct on potato amylopectin and potato starch at 1 hour and 20 hours at 37°C. The activity of each construct on potato amylopectin was similar across all four constructs suggesting that they share similar activity on soluble substrates (**Fig. 3.2A**). We observed a different trend with potato starch. Only the full-length BaAmy7 construct had appreciable activity on potato starch. The fraction that is

degraded by BaAmy7Δ74/26/13/25 likely represents the non-resistant portion of raw starch that would be degraded by human amylases. Indeed, the amount of reducing sugars released is similar to that of pancreatic amylase, used as a control. We therefore normalized the potato starch activity to the construct lacking CBMs, revealing that full-length BaAmy7 has ~3x higher activity on potato starch.

We have shown that BaCBM74 and BaCBM26 are the CBMs driving granular starch binding so we would expect that activity on potato starch would be eliminated when they are absent and that they would be sufficient to confer the ability to break down potato starch. When we deleted both CBMs (BaAmy7Δ74/26), the activity on potato amylopectin was unchanged and activity on potato starch resembled that of the GH13 alone (**Fig. 3.2A**). This supports our hypothesis that BaCBM74 and BaCBM26 are required specifically for granular starch activity.

When we deleted the non-binding CBMs, BaCBM13 and BaCBM25 (BaAmy7Δ13/25), we expected no change to the activity of BaAmy7 on either form of starch. Strikingly, BaAmy7Δ13/25 had lower Units of activity on soluble substrates per mg of enzyme (data not shown), suggesting that overall catalytic efficiency of the enzyme has been impacted. Indeed, we observed pronounced precipitation of the enzyme indicating that the enzyme may be unstable. When we performed the reactions normalizing for Units of activity (as in Fig. 3.2, see Methods 3.4) the BaAmy7Δ13/25 construct had similar activity on potato amylopectin had drastically reduced activity on potato starch, equivalent to or lower than that of the GH13 alone (**Fig. 3.2A**). Though this is the result from one experiment, it suggests that BaCBM13/25 are also necessary for activity on potato starch and that the inclusion of BaCBM74/26 are insufficient alone to confer this activity.

3.3 DISCUSSION

We have here determined the RS-degrading ability of three extracellular amylases from *Bifidobacterium adolescentis*. We found that BaAmy7 is the primary enzyme responsible for granular potato starch breakdown though BaAmy6 and BaApuB are highly active on potato amylopectin. These results are inconsistent with a previous study that found that BaAmy6 had the highest RS2 activity of the 3 resistant starch degrading enzymes (BaAmy6/7/10)(**Fig. 3.1A**) (111). We speculate that this difference in observed activity could be due to two variables: 1) that study used high amylose corn starch whereas we here used raw potato starch as the resistant starch, and 2) the conditions for the granular starch enzyme reactions such as temperature, pH, and reaction time may have influenced the results. To the first point, as discussed in (Chapter 1), there are notable differences in starch structure between high amylose corn starch and potato starch, including crystalline packing and chain length (172). BaAmy6 and BaAmy7 are highly similar across the length of the protein. Both encode a granular starch-binding CBM26 and share nearly 100% identity at their C-termini,

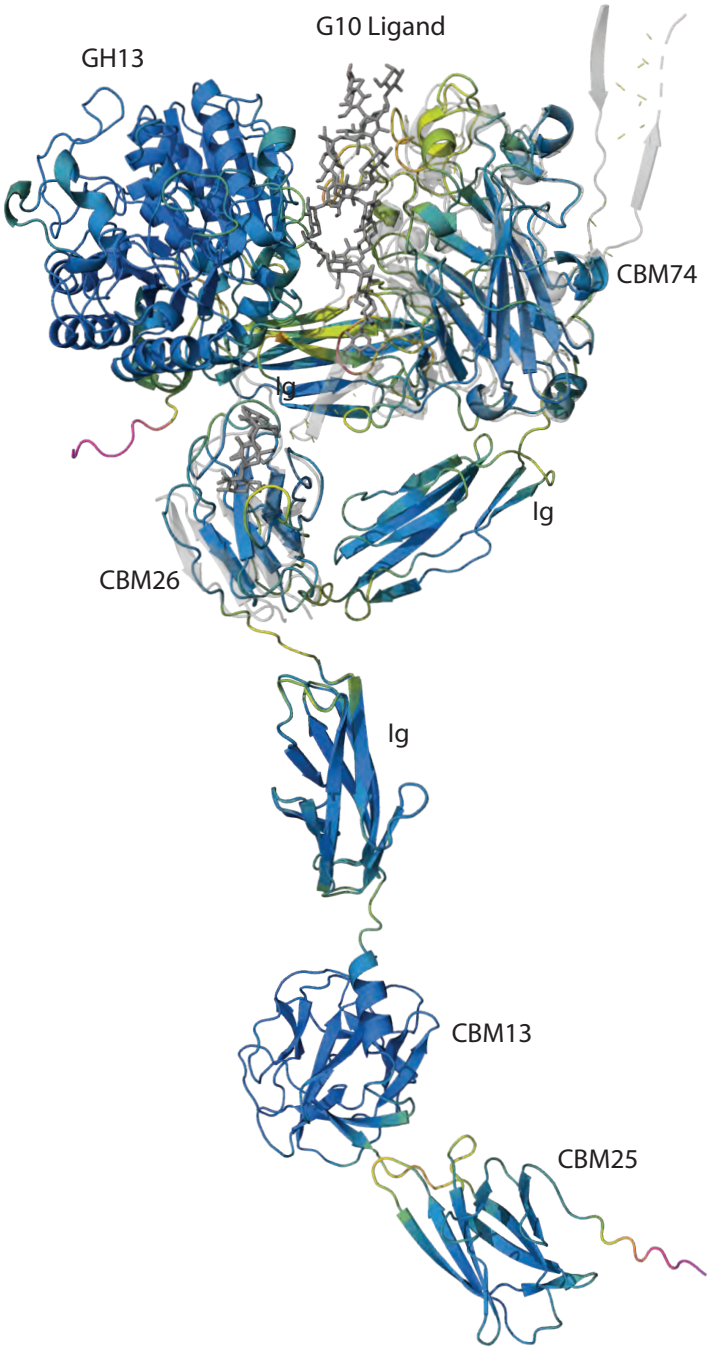


Figure 3.6: AlphaFold prediction of BaAmy7

The N-terminal prediction comes from a BaAmy7 homolog from *B. adolescentis* BIOML-A90 that lacks the C-terminal CBM13 and CBM25s (Uniprot A0A7J5NEI2). For the remaining domains we generated a structural prediction of the Ig-CBM13-CBM25 construct from the protein sequence and overlaid the two structures anchored on the Ig-like domain between CBM26 and CBM13 (225). The structure is colored by confidence score with blue: Very confident, green: confident, yellow: low confidence, magenta: very low confidence. To orient the proposed binding sites of BaCBM74 and BaCBM26, the ligand-bound RbCBM74 (PDB 7uwv) and RBCBM26s (PDB 7uww) from Sas6 (grey) are overlaid on the AlphaFold structure prediction (173).

the region encompassing CBM13 and CBM25s. The central difference between BaAmy6 and BaAmy7 is the presence of a CBM74, which has a high affinity for granular starches, with highest affinity for potato starch which is a particularly recalcitrant granular starch.

We further characterized the contribution of CBM74, CBM26, CBM13, and CBM25 to binding and enzyme activity of BaAmy7. We found that BaCBM74 and BaCBM26 are the CBMs that drive binding of BaAmy7 to granular starch and starch polysaccharides. Neither BaCBM13 nor BaCBM25 bind any starch substrates, a surprising find given that CBM25 is a demonstrated granular starch binding family.

By determining of the crystal structure we were able to show that the structural basis for the lack of binding by BaCBM25 is the replacement of two critical Trp residues that normally make up the CBM25 binding site with Thr residues. A sequence alignment of all CBM25s highlighted that this non-binding variant of CBM25 is a minority within the CBM25 family, with a higher prevalence in Bifidobacteria, though the putative binding of the CBM25 varies by strain. These results spark the question: What is the evolutionary basis of this CBM25 mutation at the binding site in *B. adolescentis* L2-32? BaAmy7 still clearly has RS-degrading activity even without the contribution of the C-terminal CBM25 to binding. The C-termini of BaAmy5/6/7/10 likely arose from a gene duplication event independently of BaApuB as that enzyme also has a CBM25 but it is binding competent (**Fig. 3.3B**).

In this study we have dissected the contribution of four CBMs to overall RS-degradation by BaAmy7. While BaCBM74/26 contribute to substrate binding and BaCBM13/25 contribute in a non-binding capacity, these results are modular. We can endeavor to understand the individual parts that make up the machine but the question remains how the parts assemble for overall function. We can use currently available AI protein prediction tools like AlphaFold to hypothesize a possible model (225; 109). However, care should be taken to avoid over-interpretation of multi-domain protein structure. The individual folds of domains are predicted with high confidence but in the case of multimodular enzymes like BaAmy7, there may be high conformational flexibility that is difficult to predict using currently available tools (90). However, AlphaFold can be used to hypothesize potential overall conformations in the absence of an experimental determined structure (5). We therefore evaluated an AlphaFold prediction of a closely related homolog of BaAmy7 that lacks the C-terminus (AlphaFold identifier AF-A0A7J5NEI2-F1). In parallel, we submitted the C-terminal Ig-CBM13-CBM25 sequence to AlphaFold2 for structure prediction and then overlaid the two predictions anchored on the Ig-like domain. BaAmy7 has a total of eight discrete domains between which there is structural uncertainty in AlphaFold predictions (**Fig. 3.6**). This low confidence score could reflect either low confidence or high flexibility in those regions (90). To compare the predicted structure to the experimentally determined CBM74-CBM26 motif, we overlaid the ligand-bound structures of RbCBM74 and RbCBM26 from Sas6, a starch-binding

protein from *Ruminococcus bromii* (173). In the conformation proposed by AlphaFold the CBM74 binding site would sandwich the ligand against the GH13 catalytic site with the CBM26 wrapped below the site. It is difficult to imagine how this conformation would promote starch capture as the GH13 and CBM74 sites are closed against one another, but it could resemble a potential "closed" conformation. In this prediction the CBM13 and CBM25s are distal from the GH13 which would be consistent with a role other than starch binding. Though these structure predictions allow us to postulate how the enzyme overall might work, it is clear by the yellow and magenta regions of (**Fig. 3.6**) that the linkers between the domains which are critical to the overall conformations the protein can adopt are predicted with low confidence and represent only one possible conformation. Therefore, other structure determination approaches are required to bridge the gap.

One approach to determining how a multi-modular enzyme works in solution is to pair X-ray crystallography with small-angle X-ray scattering (SAXS). One group used this "dissect-and-build" approach to determine that the catalytic domain of a multimodular β-N-acetylglucosaminidase from *Clostridium perfringens* lies on the same side and in the same orientation of its appended CBM32 (67). The next wave of advances in understanding how multi-modular enzymes function is likely going to be driven by single particle analysis using cryo-electron microscopy (cryo-EM). Cryo-EM has the advantage of visualizing protein particles in their native conformations and classification of particles into classes. Cryo-EM for example has been used to determine the structure of a *Thermotoga maritima* β-galactosidase (TmLac) and evaluate variations wherein a CBM had been introduced N-terminal or C-terminal to the catalytic domain (PDB 6S6Z) (153). BaAmy7 encodes eight separate domains but we don't yet understand how they are working together. BaAmy7 is a strong contender for use of cryo-EM to understand how multi-modularity determines function.

3.4 MATERIALS AND METHODS

3.4.1 Protein Expression and Purification

The gene for BaAmy7 (BIFADO_01551, UniProt Accession A7A6R7) was annotated using db-CAN3 (245) and Interpro (23) to determine domains. Where domain boundaries were unclear, we used AlphaFoldDB and AlphaFold2 GoogleColab notebook (156) to predict linkers between domains. The most similar structure in the AlphaFoldDB, A0A7J5N8K5 (GA629_04860), shares 97.9% identity with BaAmy7 over 1266 amino acids, but lacks two C-terminal domains. The AlphaFold colab notebook was used to predict the structure of the two C-terminal domains and assign boundaries. We amplified BIFADO_01551 using *Bifidobacterium adolescentis* L2-32 genomic DNA with primers containing C-terminal or N-terminal 6xHis tags with complementary overhangs to pETite C-His or N-His Kan vectors respectively.

Overlapping PCR was used to generate BaAmy7 domain deletions using primers that match the flanking sequences as previously described (131). Sequences were mixed with linearized pETite backbone and incubated together for 30min on ice and then used to transform "E. cloni 10G" *E. coli* cells (BioResearch Technologies) to propagate the assembled vector. Plasmids were isolated and sequenced to check for any unintended mutations. Chemically competent, chloramphenicol-resistant *E. coli* Rosetta PLySs DE3 was transformed with the sequenced plasmids for protein expression. Rosetta cells containing the plasmid of interest were grown in 1L of Terrific Broth to OD600 0.6-0.8 supplemented with Kanamycin (50ug/mL) and Chloramphenicol (20ug/mL). Protein expression was induced at room temperature with addition of 0.5nM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and expression continued for 16 hours. The cells were pelleted (10,000 xg, 20min) and the remaining media was removed. The resulting pellet was resuspended in TALON Buffer (20mM Tris pH 8, 300mM NaCl) and cells were lysed by sonication. Lysate was centrifuged (30,000 xg, 30min) to pellet insoluble material and cell debris. His tagged protein was purified by Nickel affinity with gravity columns loaded with Ni-NTA resin (ThermoFisher). Protein was eluted with TALON buffer with 500mM imidazole. C-terminal tagged proteins were dialyzed into storage buffer (20mM HEPES pH 7, 100mM NaCl) and concentrated. N-terminal tagged proteins containing a Tev cleavage site were incubated with 6xHis-Tev protease overnight and dialyzed into storage buffer overnight. 6xHis-Tev protease, uncleaved constructs, and non-specific contaminants were removed by an additional round of Nickel affinity purification wherein the protein of interest was collected from the flow-through and concentrated.

3.4.2 Enzyme Assays

We first determined the overall catalytic activity of each enzyme construct for normalization. We used CNP-G3 (2- chloro-p-nitrophenol linked with maltotriose) to determine Units of activity as previously described (221). 25μL of each enzyme was added to 96-well plate wells. A BioTek Synergy H1 plate reader was pre-heated to 37°C and programmed to run an enzyme kinetic assay to read the absorbance at 405nm every 1 min for a total of 10 min. Pre-heated 1mM CNP-G3 in enzyme reaction buffer (50mM Tris-HCl pH7, 10mM $CaCl_2$) was added and the change in absorbance was measured between 2-5 minutes. A standard curve was created using commercial porcine pancreatic α-amylase (Megazyme) with 75,000U/g for which one Unit of α-amylase activity is the amount of enzyme required to release one μmole of p-nitrophenol from blocked p-nitrophenyl-maltoheptaoside per minute, in the presence of excess α-glucosidase at pH 6.9 and 40°C. Several dilutions of the commercial enzyme were generated and measured. The number of Units of our recombinant enzymes were generated using the commercial enzyme as a standard. We added the amount of enzyme equivalent of 5U to the substrate reactions to control for any differences in

activity. We tested the activity of our constructs on potato starch – 1% and potato amylopectin – 0.1%. Samples were collected for reducing sugar analysis at 0hr, 1hr, and 20hr. The concentration of reducing sugars was determined using the 3,5-dinitrosalicylic acid assay as previously described (150) with a $DNSA_{10}$ working reagent that includes 10μM glucose.

3.4.3 Polysaccharide Macroarray

A macroarray screen for carbohydrate binding by CBMs was performed as previously described (223). Potato amylopectin, maize amylopectin, solubilized amylose, glycogen, pullulan, and dextran (Sigma) were prepared as 0.1% solutions. Amylose was solubilized by dissolving in 1M NaOH and the addition of HCl until the solution reached pH 7. 2µL of each polysaccharide solution was blotted on a nitrocellulose membrane and allowed to air dry for 2hr. CBM binding was initiated by the addition of 500μg of CBM construct diluted in 3mL of EveryBlot blocking buffer (Bio-Rad). Blots were incubated with protein for 1hr at room temperature and subsequently blotted with mouse α-polyHis-AP antibody for 1hr, washed 3x5min with TBS-T, and detected using an NBT/BCIP colorimetric substrate. Blots were imaged on a Gel-Doc+ using the colorimetric blot setting.

3.4.4 Crystallography

BaCBM25 was crystallized in 0.14M Calcium chloride dihydrate; 0.07M Sodium acetate; pH 4.6; 14% v/v 2-Propanol; 30% v/v Glycerol determined from a condition in the JCSG-plus screen (Molecular Dimensions). X-ray data were collected at the Life Sciences Collaborative Access Team (LS-CAT) at Argonne National Laboratory's Advanced Photon Source (APS) in Argonne, IL. Data were processed at APS using autoPROC with XDS for spot finding, indexing, and integration followed by Aimless for scaling and merging (228; 112; 64). Intrinsic sulfur SAD phasing was used to determine the structure of BaCBM25 using AutoSol in Phenix (59; 214). The structures were refined using manual model building in WinCoot and refinement in Phenix.refine (60; 3).

3.4.5 Bioinformatic Analysis

BaAmy6 and BaAmy7 were aligned using TCoffee (165; 51). The alignment was annotated using Alignment-to-HTML with manual annotation of domains with boundaries determined above (83; 84). Reported RMSD values and sequence identity for CBM25 structural homologs are those determined by the DALI server (96).

Table 3.1: BaAmy7 Structure Statistics

Construct	**BaCBM25**
PDB Accession	
Wavelength	1.55
Resolution range	44.51 – 1.57 (1.61-1.57)
Space group	C2
Unit cell	64.83 61.65 75.35 90 97.3 90.00
Total reflections	2,838,448 (9,406)
Unique reflections	33374 (351)
Multiplicity	85.0 (26.8)
Completeness (%)	81.3 (11.7)
Mean I/sigma(I)	48.0 (1.9)
R-merge (%)	9.5 (95.4)
CC1/2	1.00 (0.877)
R-work (%)	18.1
R-free (%)	21.4
RMS (bonds)	0.01
RMS (angles)	1.10
Ramachandran favored (%)	96.5
Ramachandran allowed (%)	3.5
Ramachandran outliers (%)	0.0
Rotamer outliers (%)	0.0
Clashscore	5.3
Average B-factor	27.8
macromolecules	27.2
ligands	25.0
solvent	35.1

CHAPTER 4

Discussion

Starch is an abundant part of most human diets. In most cases it is thoroughly cooked, making it accessible to the action of human digestive enzymes, namely salivary and pancreatic amylases (62). However, uncooked granular starch is recalcitrant to the action of all but a few specialized bacterial enzymes qualifying it as resistant starch (RS) (63). Since RS is undigested by human amylases, it becomes food for the collection of microorganisms that inhabit the human gastrointestinal tract, known as the gut microbiota (226). When the primary degraders of RS unleash their specialized starch-degrading machinery, they initiate a crossfeeding cascade that results in the expansion of butyrate-producing bacteria (15). Butyrate is an end-product of bacterial fermentation in the anaerobic environment of the large intestine and serves as the primary energy source for the colonocytes that make up the lining of the intestine (32). In addition, butyrate has a myriad of gut health benefits. RS2, therefore, is being investigated as a potential prebiotic to promote a healthy gut microbiota. A major current limitation to the implementation of RS2 as a prebiotic is the inter-individual variability in the magnitude of the butyrogenic response. In order to design a more targeted implementation of RS2 as a therapeutic, it is necessary to understand how the RS2 specialists initiate breakdown and influence the community.

Bacterial starch breakdown is mediated by amylases that encode glycoside hydrolase family 13 (GH13)s typically with appended carbohydrate binding module (CBM) families that bind starch (106). These CBMs can play a multitude of roles including binding starch ligand and targeting amylases to the insoluble granule surface (216; 87). Aside from their targeting role, CBMs can in some cases structurally disrupt the crystalline packing to allow the catalytic domain better access and anchor the enzyme to the cell surface (87). Of the 16 starch-binding CBM families, CBM74 is of particular interest as it is not broadly encoded by non RS-degrading gut bacteria but each of the known RS2-degraders encodes one protein with a CBM74 (223). CBM74 was discovered as part of a RS-active amylase from bacterium isolated from the starch sludge of a potato starch processing facility (223). The first characterized CBM74 binds granular starch and promotes structural disruption if the granule surface by its concomitant GH13 (223). Therefore, CBM74 has promise as a molecular signature for RS2 utilization.

4.1 Summary of Findings

This thesis endeavors to build on the previous characterization of the starch-binding domain, CBM74, by determining the molecular basis of granular starch binding and its role in enzyme activity (223). The original CBM74 paper reported that CBM74 was a single domain of ~300 amino acids, three times larger than the average CBM and 100 amino acids longer than its closest homolog, CBM9. Despite the large size, CBM74 was bioinformatically predicted to be a single globular domain. The authors showed that the CBM74 from *Microbacterium aurum* AmyA binds amylose, amylopectin, and granular starch. Of granular potato, corn, and wheat starch *Ma*CBM74 had the highest affinity for potato starch. In a phylogenetic comparison of all predicted CBM74 homologs, 62% of the sequences belonged to Bifidobacterial species and 77% belonged to multi-modular amylases with a GH13_28. In all but two of the CBM74-containing proteins in the phylogenetic tree, the CBM74 is encoded adjacent to a CBM25 or CBM26. We also aimed to explore the potential cooperation and basis for the co-occurrence of these domain families. The overarching hypothesis of this thesis was that CBM74s have a specialized binding site for targeting the full protein to the surface of raw starch granules.

In chapter 2, we explored the role of CBM74 in starch binding by characterizing the protein Sas6 from *Ruminococcoccus bromii*. Sas6 is one of the few CBM74-containing proteins without a GH13 but instead has a dockerin domain for protein-protein interactions, enabling it to complex with other binding and GH13-containing proteins in extracellular molecular machines known as amylosomes. The arrangement of Sas6 consists of a CBM26 (*Rb*CBM26) and a CBM74 (*Rb*CBM74 flanked by bacterial Ig-like2 (BIg) domains for structural fidelity (**Fig. 2.1**). We found that Sas6 binds to amylose, amylopectin, and granular starches from corn and potato. Our binding studies with the individually expressed domain found that *Rb*CBM74 requires a maltooligosaccharide of at least 8 residues to serve as a ligand, whereas *Rb*CBM26 is able to bind substrates of at least 4 glucose residues long (**Table 2.3**). We speculate that *Rb*CBM26 binds open and elongated maltooligosaccharides like those found in starch breakdown products and amorphous regions. In contrast, *Rb*CBM74 exhibits significantly (~100-fold) higher affinity for amylopectin than *Rb*CBM26 and preferentially binds longer substrates that more frequently form double helices, as in the patterns in the crystalline layers of granular starch.

We were able to determine the crystal structure of Sas6 with α-cyclodextrin bound at the *Rb*CBM26 binding site and the co-crystal structure of *Rb*CBM74 with maltodecaose (G10) bound (**Fig. 2.2, 2.6**). In addition, we determined the conformational flexibility of Sas6 in solution with small angle X-ray scattering and found that there is minimal wobble between the *Rb*CBM26 and *Rb*CBM74s hinged around the BIg domain between them (**Fig. 2.2**). However, the overall structure remains mostly compact, keeping the *Rb*CBM26 and *Rb*CBM74 binding sites in regional proximity.

In chapter 3, we extrapolated what we had learned about the mode of binding of the CBM74 + CBM26 motif from Sas6 to a multi-modular amylase from *Bifidobacterium adolescentis*, BaAmy7. We recombinantly expressed BaAmy7 and two other extracellular amylases from *B. adolescentis* and found that only BaAmy7 exhibited appreciable activity on raw potato starch, the most recalcitrant of the resistant starches (**Fig. 3.2**). When both CBM74 and CBM26 were removed from BaAmy7 the activity on potato starch matches that of the GH13 alone, indicating that this motif is required for RS activity (**Fig. 3.5**). Based on the reduced activity when the C-terminal CBMs from the CBM13 and CBM25 families are deleted, these domains likely provide structural stability.

With these investigations, we have determined the protein structure of CBM74, how it binds double helical starch motifs in its elongated binding groove, and how it cooperates with adjacent CBM26 for enzymatic activity on RS2.

4.2 Future Directions

As the 2016 FEMS CBM74 paper was for me, I hope that this work serves as a starting point for other curious investigators to explore. There are many more questions that have come to light that could serve as the seeds of new projects.

4.2.1 *R. bromii* Future Directions

I have here characterized the starch-binding protein Sas6 which is a putative member of the *R. bromii* amylosome system. It has a dockerin domain, but we don't yet know *how* it assembles and which scaffoldins (scaffolding proteins for amylosome assembly via cohesin-dockerin interactions) house the cognate cohesin for the Sas6 dockerin. In most cases a dockerin-containing GH13 can bind to several cohesins on several scaffoldins leading to flexibility in how the system can assemble (160). I have mentioned that Sas6 doesn't have a GH13 so it is not catalytically active. Is there a GH13 with which Sas6 is primed to cooperate? We have a custom antibody generated against Sas6 that could be used to co-immunoprecipitate interacting proteins to address this question. Recently, another member of the *R.bromii* starch adherence system, Sas20, was characterized and determined to bind granular starch by attaching to the non-reducing ends abundant at the cell surface (35). Since Sas6 likely attacks the crystalline regions, it has the potential to cooperate with Sas20 to recognize different motifs at the granule surface. Sas20 complexes with Sca5 which only has 2 cohesins, so the cooperation would require an intermediary like Sca1/Amy4 which has both a cohesin and a dockerin (160; 35). This can be addressed by taking an approach similar to the "dissect-and-build" approach in cellulosome field to determine how the pieces of the amylosome may be assembling

in vivo (200). The studies herein contribute to the "dissect" approach while future studies of interacting components can be undertaken using small angle X-ray scattering and cryo-electron microscopy to move to the "building" phase. Once the assembly of amylosome components is better understood, we can synthesize designer amylosomes for maximal RS2 digestion.

4.2.2 *B. adolescentis* Future Directions

As a direct follow-up to the findings in chapter 3, there are several logical next experiments. I expressed and characterized recombinant BaAmy7 lacking pairs of CBMs. We learned that the CBM74-CBM26 motif is important for granular starch targeting, but it is unknown if they play redundant roles or if their cooperation is required for RS2 activity. The C-terminal CBMs contribute a function other than binding. I hypothesize they play a role in stability as they are also required for RS2 activity. It is unclear if both domains are required. To further parse out the contributions of each CBM, it would be prudent to express variants of BaAmy7 lacking only one CBM in each construct. We could also pair only one CBM at a time with the GH13 but it will be important to keep the BIg domains that likely provide the overall structural framework for the enzyme. A less likely possibility but one that should still be considered is that a non-binding domain of BaAmy7 facilitates multi-protein oligomerization. *B. adolescentis* has 3 other extracellular amylases with nearly identical C-termini. The possibility of a multi-enzyme complex would be intriguing as it would effectively generate a dual-catalytic machine much like ApuB. This could be a good strategy for enzyme action in the crystalline layers where double helices are packed closely together. The content of Chapter 3 is a starting point for understanding how BaAmy7 works as an overall machine and cooperates with the other extracellular amylases of *B. adolescentis* for RS2 breakdown.

In Chapter 3.3 I speculated how the different domains of BaAmy7 might arrange in solution based on my previous studies of Sas6 (see 2.5.9). We could learn a lot about the enzymatic action of BaAmy7 if we knew more about its available conformations. One approach is to use X-ray crystallography to determine the lowest energy conformation though these studies are difficult with a target of this size, especially if there is flexibility between the different domains. Given the high confidence AlphaFold predictions of the individual domains, we could use small angle X-ray scattering to generate an envelope into which we could model the individual domains to get some idea of the overall flexibility of the enzyme between the domains. Another enticing approach is to use single particle cryo-electron microscopy to determine the three-dimensional structure and dynamics of BaAmy7. BaAmy7, with a molecular weight of 133 kDa, presents a promising candidate for structural characterization through cryo-EM analysis. I expect that the CBM binding sites are structurally arranged on the same axis as observed in the AlphaFold structure prediction of the overall protein (see 3.3). This would allow the entire polypeptide to bind linearly along the

smooth granule surface for tighter binding. However, it is possible that the CBM binding sites are arranged in opposing orientations. This conformation could facilitate increased initial capture of starch or simultaneous binding to adjacent amylopectin chains.

4.2.3 Genetic engineering of RS2-degrading bacteria

My work has determined the structural basis of granular starch binding by CBM74 and revealed that CBM74 can accommodate starch double helices in its binding site. However, this work has stopped short of providing evidence that the CBM74-containing proteins encoded by *R. bromii* and *B. adolescentis* are instrumental in unlocking granular starch as a substrate for growth. The logical next step to understand the role of CBM74s in RS2 breakdown is to determine if CBM74-containing proteins are *necessary* and *sufficient* for growth on starch. Genetic manipulation tools for *R. bromii* and *B. adolescentis* are still in development. Some roadblocks include the restriction modification systems in these systems that will digest any improperly methylated DNA introduced and the lack of plasmids to be used as vectors (168). If these tools become available it would be informative to delete or disrupt Sas6 in *R. bromii* and BaAmy7 in *B. adolescentis* and test their ability to bind to, degrade, and grow on RS2.

An alternative is to take a gain-of-function approach to heterologously express BaAmy7 in another Gram-positive species like *Lactobacillus plantarum* or *Bacillus subtilis*. There are more tools available for modifying these species so BaAmy7 is a potentially low-hanging fruit for performing these studies. I exclude Sas6 from this proposed approach as we don't know enough about amylosome assembly to reconstitute amylosomes in another species. It may be feasible in the future following the cellulosome field blueprint in which isolated components are expressed in tandem within a model organism to determine how those proteins assemble and work together (205). Once we are able to study *in vitro* assemblies of amylosomes, we can express them in *L. plantarum*. This is a critical next step in moving from biochemical characterization of starch-binding domains to engineering synbiotics to promote human health.

4.3 Broader Implications

The central question of my thesis was: What specialized protein machinery promotes initiation of starch breakdown by RS2-degrading gut bacteria? Each of the known primary RS2 degraders in the human gut, *R. bromii* and *B. adolescentis*, encode one CBM74-containing protein. Therefore, CBM74 shows potential as a marker for RS2-degraders in the human gut. Since the start of this work, several studies have supported this idea. CBM74 is more abundantly represented in resistant starch interventions in metagenomic and metatranscriptomic studies (178; 238). New primary

degraders that have been isolated from the human gut continue to encode CBM74-containing proteins in their genomes (98). A new resistant starch degrading Ruminococcus sp. FMB-CY1 which is closely related to *R. bromii* has been described as degrading raw starch granules faster and more completely than *R. bromii* (98). This organism harbors two CBM74s in its genome, compared to just the one on Sas6 from *R. bromii* (cazy.org) (54). Another potential RS2-degrader, most closely related to the rumen organism, *Clostridium chartatabidum* also encodes a CBM74-containing protein in its genome (15). Future studies should consider CBM74 as a potential marker when analyzing large metagenomic RS2 studies.

A potential application of this work in the gut microbiota field is the development of synbiotics. If we can understand the function of domains like CBM74, then it may be possible to genetically engineer probiotics with RS2-degrading machinery. If these probiotics are administered in parallel with RS2 that is highly selective for strains that can access it, it should promote colonization and expansion of that species.

Alternatively, the molecular characterization of raw starch degrading enzymes like BaAmy7 has an important industrial application. We are in the midst of a climate crisis exacerbated by the use of fossil fuels. There is a push toward more renewable fuel sources, primarily bioethanol. Due to its availability, corn starch makes up the predominant source for bioethanol production in the United States (42). But currently the processes to generate bioethanol from corn starch are inefficient, leading to increased cost and need for raw materials. The cultivation of corn for biofuel production competes with farming land for the human food supply and has its own detrimental impact on the environment. Therefore, improvement of the breakdown of starch for biofuel production could have powerful positive impacts. Currently the process involves cooking the starch at 105°C, followed by digestion with a thermostable α-amylase and glucoamylase (42). These processes require high temperatures and the use of caustic chemicals for pH control (42). Designing a stable amylase that efficiently digests raw starches would reduce the monetary and resource costs associated with biofuel production. Biomining of raw starch degrading bacteria for their enzyme blueprints can streamline this process and help move toward a more efficient biofuel generation process.

4.4 Concluding Thoughts

The microbial world is an abundant source of novel enzymes for a wide variety of applications that have revolutionized our society. A heat-stable DNA polymerase from a thermophilic bacterium made polymerase chain reaction widely applicable. Restriction enzymes used by bacteria to "chew up" foreign DNA at specific sequences are widely employed for molecular cloning. The CRISPR-Cas proteins that make up bacterial defense systems against bacteriophage have made eukaryotic gene editing possible. Microbes are fantastic at evolving new proteins and it's up to structural

microbiologists to determine how new microbial machines work. I hope that my research on resistant starch utilization by gut microbes has left an indelible mark on the starch field and that I can continue to demystify the biological sources of revolutionary new microbial functions.

APPENDIX A

Primary Resistant Starch Degraders Display Differing Approaches to Potato Starch Granule Breakdown

In the preceding chapters we have investigated the granular starch breakdown mechanisms of three different primary resistant starch (RS) degrading bacteria: *Ruminococcus bromii* strain L2-63 and *Bifidobacterium adolescentis* strain L2-32. This appendix addresses the broad questions: "How do primary degraders structurally degrade potato starch granules?" and "Does this degradation facilitate utilization by otherwise non-RS degrading gut symbionts?"

Previous studies have found that an increase in relative abundance of *R. bromii* is most closely correlated with an increase in the butyrate producer, *Eubacterium rectale* (15). The ability of *R. bromii* to cross-feed *E. rectale* with its byproducts has been shown in co-culture as well (241). In contrast, an increase in *B. adolescentis* is correlated with an increase in a different profile of butyrate producers including *Anaerostipes hadrus* and *Eubacterium halii* likely via the production of short chain fatty acids acetate and lactate (15; 17). In co-culture *B. adolescentis* cross-feeds *Faecalibacterium prausnitzii* (183). We propose that an additional mode of cross-feeding is through primary degrader deconstruction of starch granules to the point that it can be used by other bacteria.

The ability of Bifidobacteria to adhere to and degrade granular starch can vary by species and strain. In addition to *B. adolescentis* L2-32, we investigated an isolate of *B. adolescentis* from a human cohort that blooms in response to dietary potato starch administration, designated strain U269 (15). To investigate the starch breakdown dynamics by *R. bromii* L2-63, *B. adolescentis* L2-32, and *B. adolescentis* U269 we grew each strain in liquid culture on potato starch as a sole carbon source. By 24 hours we observed granule aggregation in *R. bromii* and *B. adolescentis* U269. In contrast *B. adolescentis* L2-32 maintains a uniform slurry of insoluble granules.

We analyzed potato starch granules digested by the three RS2-degraders described in this dissertation. With no bacteria the granules are large, smooth, and generally free of cracks or pores (**Fig. A.1A**). At 4 days post-inoculation, using Scanning Electron Microscopy (SEM), we

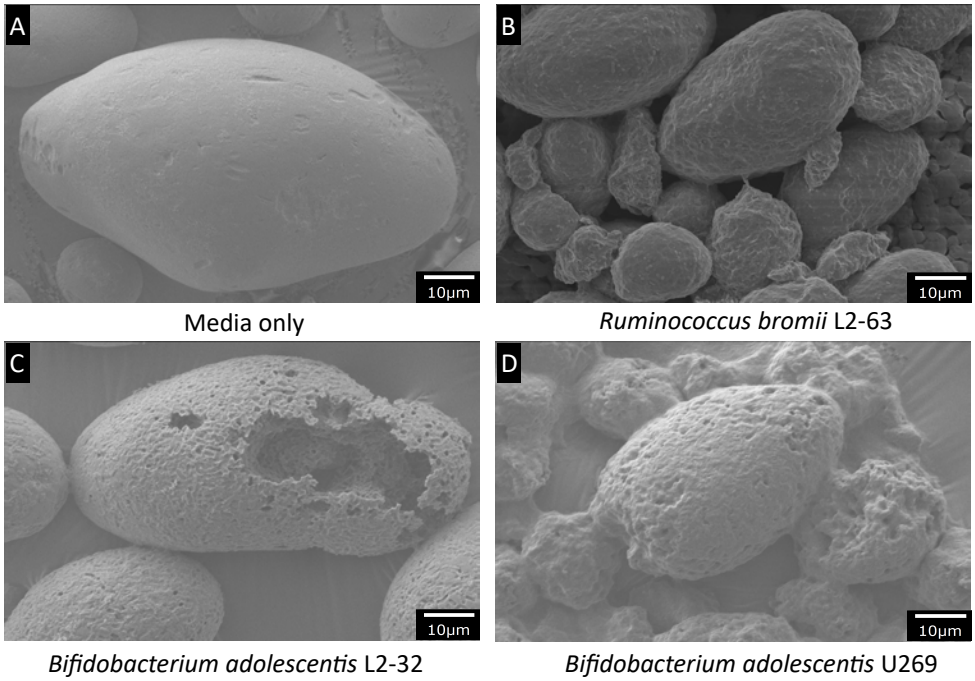


Figure A.1: Scanning electron micrographs of potato starch granules after 72 hours of bacterial utilization by **A.** YCFA medium alone, **B.** *Ruminococcus bromii* L2-63, *C. B. adolescentis* L2-32, or **D.** *B. adolescentis* U269.

see that *R. bromii* generates an adhesive substance that holds the granules together and that the cells form pockets in the surface of the granules (**Fig. A.1B**). Meanwhile, *B. adolescentis* L2-32 is immobilized on the surface of the granule and digests the surrounding starch resulting in large holes into the center of the granule (**Fig. A.1C**). In the *B. adolescentis* U269 culture, the starch granules were nearly entirely digested with only a few clumps of starch found within the imaged sample (**Fig. A.1D**).

In conclusion, all three of the RS-degrading bacterial strains described in this thesis have different starch destruction approaches. Thus even the closely related strains of *B. adolescentis* L2-32 and U269 likely encode variations in their genomes that dictate the aggregation and breakdown of granular starch. These divergent phenotypes may have implications for cross-feeding of the resident gut microbiota by generating differentially accessible granular leftovers. Therefore, it is important to understand the protein machinery that drives these three different phenotypes.

APPENDIX B

Bifidobacterium adolescentis U269 Makes a Sortase-Dependent Pilus With a Starch-Binding Tip

The results presented here are a part of a collaboration with Dr. Tom Schmidt, Dr. Ethan Hillman, Dr. Clegg Waldron, and Haiyan Tang.

The single particle electron microscopy studies were performed in collaboration with Dr. Melanie Ohi, Dr. Louise Chang, and Dr. Katarina Meze with instructional and equipment support from the University of Michigan CryoEM core facility.

B.1 Introduction

Uncooked granular starch is a highly recalcitrant substrate that can only be broken down in the human colon by a few specialized bacterial species that encode dedicated machinery for that purpose (226; 36). These species increase in relative abundance when raw potato or corn starch is administered in the diet as they readily outcompete most of the other gut microbiota residents for this carbon source (15). The two well-characterized RS-utilizers from the human gut are *Ruminococcus bromii* and *Bifidobacterium adolescentis* (132; 2 41). Bifidobacteria are anaerobic bacteria that reside in mammalian gastrointestinal tracts and are one of the first colonizers of the infant gut microbiota (182). While in Chapter 3 we characterized an extracellular amylase from *B. adolescentis* strain L2-32, the ability to adhere to and grow on granular starch as a carbon source is highly strain specific (43; 1 11). A newly isolated strain of *B. adolescentis*, strain U269, grows better than strain L2-32 on granular potato starch in culture and has a different starch degradation approach entirely (see A). *B. adolescentis* U269 breaks down starch by first aggregating the starch granules into large clumps. We therefore investigated the biochemical basis for initial bacterial attachment to the starch granule surface. We speculated that the *B. adolescentis* U269 may have a mode of multivalency that would mediate the starch aggregation we observe in culture.

A 2014 RNA-seq study looked for genes that were upregulated when the closely related strain *B. adolescentis* 22L was grown on starch and maltodextrins over glucose (58). Among the genes

identified were BaApuB and BaAmy6 in addition to several other genes and gene clusters, one of which was a sortase-dependent pilus gene locus, *pil4* (58). Pili are hairlike appendages on the surface of bacteria that play a role in adherence, host colonization, biofilm formation, DNA transfer, immunomodulation, electron transfer, and twitching motility (45; 122). Sortase dependent pili are made by Gram-positive organisms In this type of pilus, the basal pilin is covalently linked to the peptidoglycan of the cell wall via the action of a dedicated sortase enzyme (45). The pilin structure is composed of repeating units of the backbone pilin which polymerize to form the extended pilus structure (122). At the tip of the pilus is a tip protein that provides the main adhesion function (122). Sortase-dependent pili in Bifidobacteria have been shown to be important for gut colonization through the adhesion of the tip pilin to extracellular host glycans and extracellular matrix sugars (218; 71).

B.2 RESULTS

B.2.1 *B. adolescentis* U269 makes a sortase dependent pilus

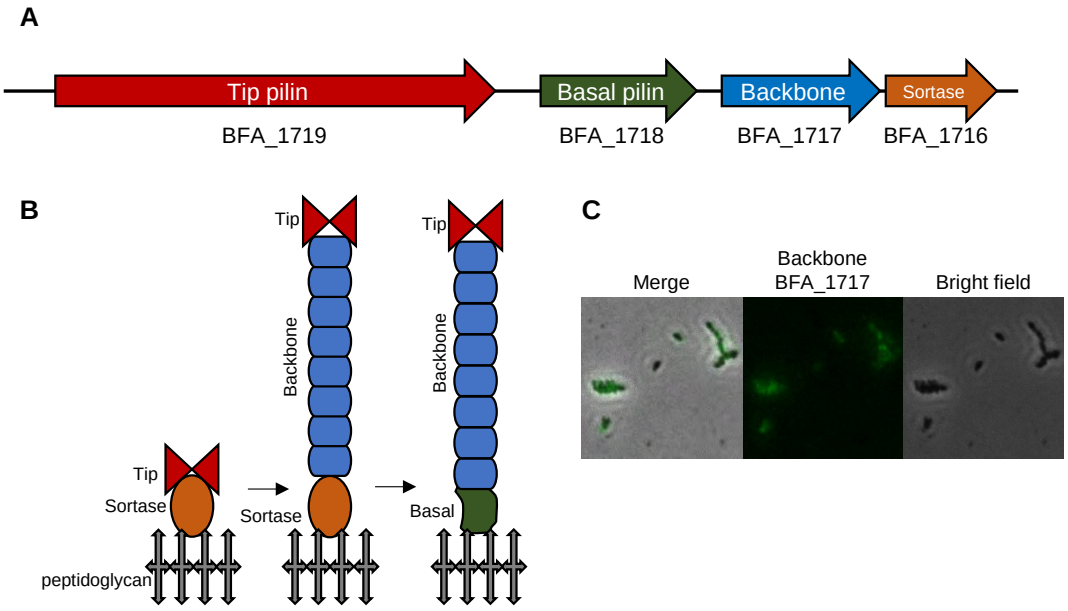


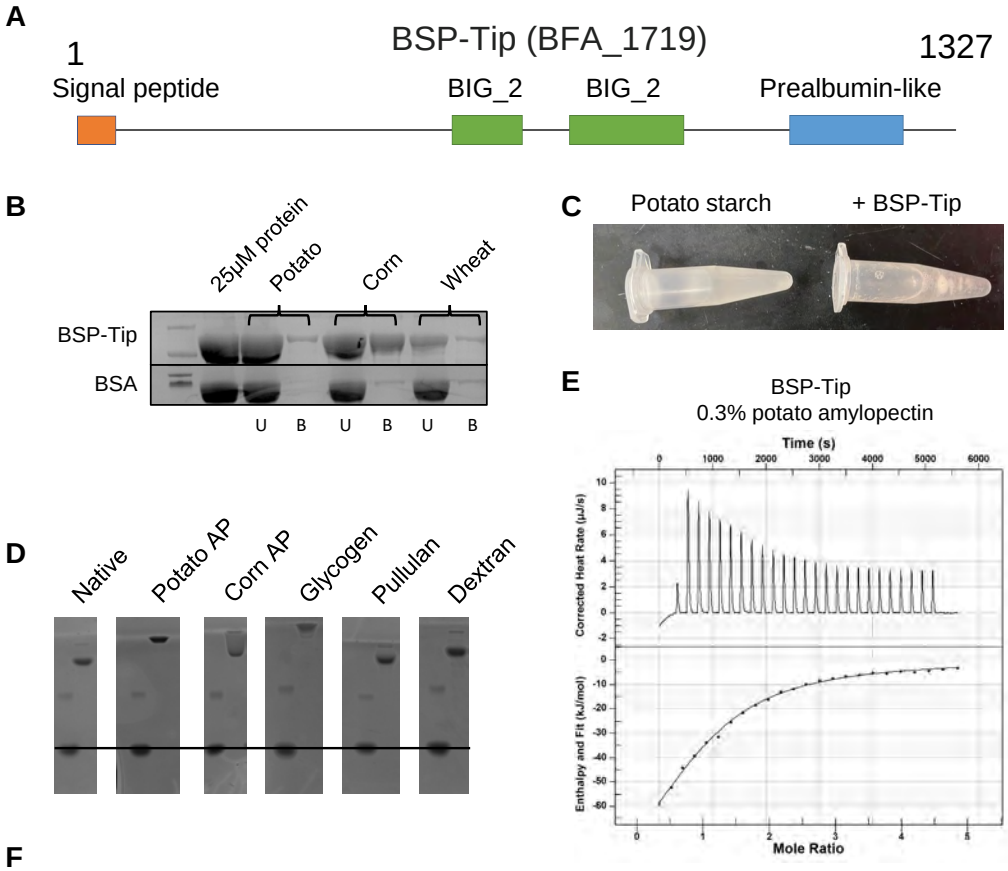
Figure B.1: The gene locus BFA_1719-1716 encodes a sortase dependent pilus
A. The *B. adolescentis* U269 *pil4* locus is an operon of four genes. **B.** A model of sortase dependent pilus assembly with the location of each of the *pil4* genes. First the tip is anchored in the peptidoglycan layer by a *pil4*-specific sortase. Major pilins then get covalently linked to the previous pilin to form the stalk of the pilus. Finally, the pilus gets transferred by the sortase to a basal pilin that anchors the pilus in the cell wall. Modified from (122). **C.** Immunofluorescence of *B. adolescentis* U269 cell grown on 0.1% potato amylopectin and stained with a custom primary antiserum to the backbone pilin (BFA_1717)

The sortase dependent *pil4* locus is composed of four genes: a tip pilin (BFA_1719), a basal pilin (BFA_1718), a backbone pilin (BFA_1717), and a dedicated sortase enzyme (BFA_1716) (**Fig. B.1A**). This locus is analogous to the genes BADO_1455-BADO_1458 in *B. adolescentis* 22L (58). **Fig. B.1B** contextualizes where each of the four *pil4* genes fit into the paradigm of overall pilus assembly. The pilus specific sortase first cleaves the LPXTG motif of the tip pilin, forming a tip-sortase intermediate. The sortase then goes on to catalyze polymerization of the pilus stalk by forming covalent bonds, first between the tip and a backbone pilin, followed by repeated addition of backbone pilins. A housekeeping sortase that is not part of the *pil4* locus presumably adds the basal pilin and covalently links it to a peptidoglycan component, Lipid II (122). We used immunofluorescence to detect the production of the *pil4* pilus *in vivo*. We found that the pilus is made when *B. adolescentis* U269 is grown on potato amylopectin (**Fig. B.1C**).

B.2.2 BSP-Tip binds to starch

To explore the fidelity of the BFA_1716-1719 *pil4* locus as a starch-binding pilus, we recombinantly expressed the tip pilin protein, BFA_1719 and tested it for starch-binding affinity. We first used InterPro and dbCAN2 to annotate any predicted starch-binding domain that might be present in the tip pilin (242; 23). The only annotations were for 2 Bacterial Immunoglobulin 2 (BIG_2) domains and a "Prealbumin-like domain" near the C-terminus (**Fig. B.2A**). The Ig-like fold of BIG_2 domains is a very stable fold that describes the inert spacer domains between functional modules of a multi-modular protein but is also the fold adopted by many starch-binding carbohydrate binding modules. The prealbumin-like annotation is associated with a motif present in SpaA of *Corynebacterium diphtheriae*, the causative agent of diphtheria (23). SpaA is the backbone pilin of the *C. diphtheriae* SpaA pilus, responsible for bacterial adherence to human pharyngeal epithelial cells (113). From sequence analysis there are no domains that stand out as a potential starch-binding domain.

We recombinantly expressed the tip pilin without the signal peptide and C-terminal predicted helix (residues 61-1301) and interrogated its ability to bind granular starch. Potato starch exhibits B-type crystallinity and therefore has low accessibility to binding proteins and enzymes as opposed to corn and wheat starch which exhibit A type crystallinity and are more accessible to breakdown (145). We mixed tip pilin protein with granular starch for one hour and used SDS-PAGE to visualize the unbound protein remaining in the supernatant (U) and the protein bound to starch granules after 3 washes. Bovine serum albumin was used to determine non-specific protein binding typical for each starch. We found that the tip pilin binds potato, corn, and wheat starch (**Fig. B.2B**). The bound portion qualitatively matches the BSA control, the unbound fraction is depleted. This may suggest that BFA_1719 (Tip) binds wheat starch with sufficiently low affinity that it does not stay



F

	K_a (M^{-1})	K_d (μM)	(mM)/% starch	ΔG (kcal.mol^{-1})	ΔH (kcal.mol^{-1})	TΔS (kcal.mol^{-1})
0.3% Potato Amylopectin	6.0e+04 (± 0.5)	16.9 (± 1.5)	1.3 (± 0.0)	-27.2 (± 0.2)	-112.5 (± 4.4)	85.3 (± 4.6)

Figure B.2: *pil4* encodes a starch-binding pilus.
A. Predicted domain schematic of BSP-Tip (BFA_1719) using InterProScan (108) 'BIG_2' refers to Bacterial Immunoglobulin-like2 domains. **B.** BSP-Tip binds granular starch. 8% SDS-PAGE gel showing free protein added, (U) unbound, and (B) bound protein denatured from washed starch granules. Lane 1: molecular weight ladder, lane 2: no starch control, lanes 3-8: unbound and bound fractions from potato starch (lanes 3-4), corn starch (lanes 5-6) or wheat starch (lanes 7-8). **C.** Example of starch aggregation induced by addition of BSA-Tip. **D.** Affinity PAGE gels for 0.3% potato amylopectin, maize amylopectin, glycogen (bovine liver), pullulan, and dextran. BSA (left) as a nonbinding control in each gel to measure band retardation. BSP-Tip (right) migration is compared to that of BSA in the Native gel. **E.** Isothermal titration calorimetry (ITC) results for BSP-Tip titrating 0.3% potato amylopectin into 25μM protein. **F.** Binding parameters of BSP-Tip to potato amylopectin as determined by ITC. Since number of binding sites on the protein cannot be determined using this approach, we instead report the concentration of binding sites in 1% amylopectin (given by the mM/% value) for this construct overall. See Methods section for details.

bound to starch granules during the wash steps. Strikingly, we observed that when we added the BFA_1719 (Tip) protein to all of the starch, but potato starch especially, it aggregated the starch into large clumps much like we would see with this strain in culture (**Fig. B.2C**). We determined the *pil4* locus encodes a verified starch-binding pilus, hereafter called the Bifidobacterial starch pilus, or BSP.

We next used affinity PAGE in which a polysaccharide of interest is incorporated into the matrix of a non-denaturing gel to qualitatively test for binding of BSP-Tip protein to starch polysaccharides. We tested binding to polysaccharides composed primarily of α1,4-linkages with α1,6 branch points occurring an average of every 23.1 residues (potato amylopectin), 19.7 residues (maize amylopectin), or 13 residues (glycogen) (151; 20). Amylose is almost exclusively composed of α1,4-linkages and forms tight single helices that make it largely insoluble in water. Pullulan is composed of maltotriose units connected by α1,6 branches and can only be bound and broken down by proteins that can accommodate the branch point in their binding or active sites or only require a three glucose footprint for recognition. Dextran is a non-starch α-glucan composed mainly of α1,6 linkages with branch points formed by α1,3-linkages and is not typically accommodated by starch-binding proteins.

We found that BSP-Tip binds potato amylopectin, corn amylopectin, glycogen, amylose, and pullulan but not dextran (**Fig. B.2D**). We further quantified the BSP-Tip affinity for potato amylopectin by isothermal titration calorimetry (ITC) (**Fig. B.2E**). The K_d is $\approx 17\mu$M which is 50x lower affinity than the affinity of Sas6 for the same substrate (For Sas6 comparison see table 2.3)(**Fig. B.2F**).

B.2.3 Demarcation of starch-binding regions of BSP-Tip

We used trROSETTA to predict the domain boundaries from which we created an N-terminal BSP-Tip constuct (BSP-N1) and a C-terminal construct (BSP-Tip-C1) split at amino acids 695-696 (55). The BSP-Tip-N1 and BSP-Tip-C1 constructs were then interrogated for binding to starch components using a polysaccharide macroarray. We found that BSP-Tip-FL binds to raw potato and corn starch, amylopectin from both potato and corn, glycogen, amylose, and pullulan (**Fig. B.3A**). As expected, it did not bind to dextran. These results recapitulate the affinity PAGE results (**Fig. B.2D**). For the BSP-Tip-N1 construct, we observed faint signal for the granular starches, amylopectin, and glycogen (**Fig. B.3A**). These results are qualitative and require further quantification by ITC. In contrast, the profile of the BSP-Tip-C1 construct resembled the binding pattern of the FL protein (**Fig. B.3A**). These results suggest that there is minimal starch-binding ability conferred by the N-terminus of BSP-Tip but that starch-binding is primarily determined by the C-terminal portion.

When the AlphaFold database was launched with the goal of predicting nearly every structure

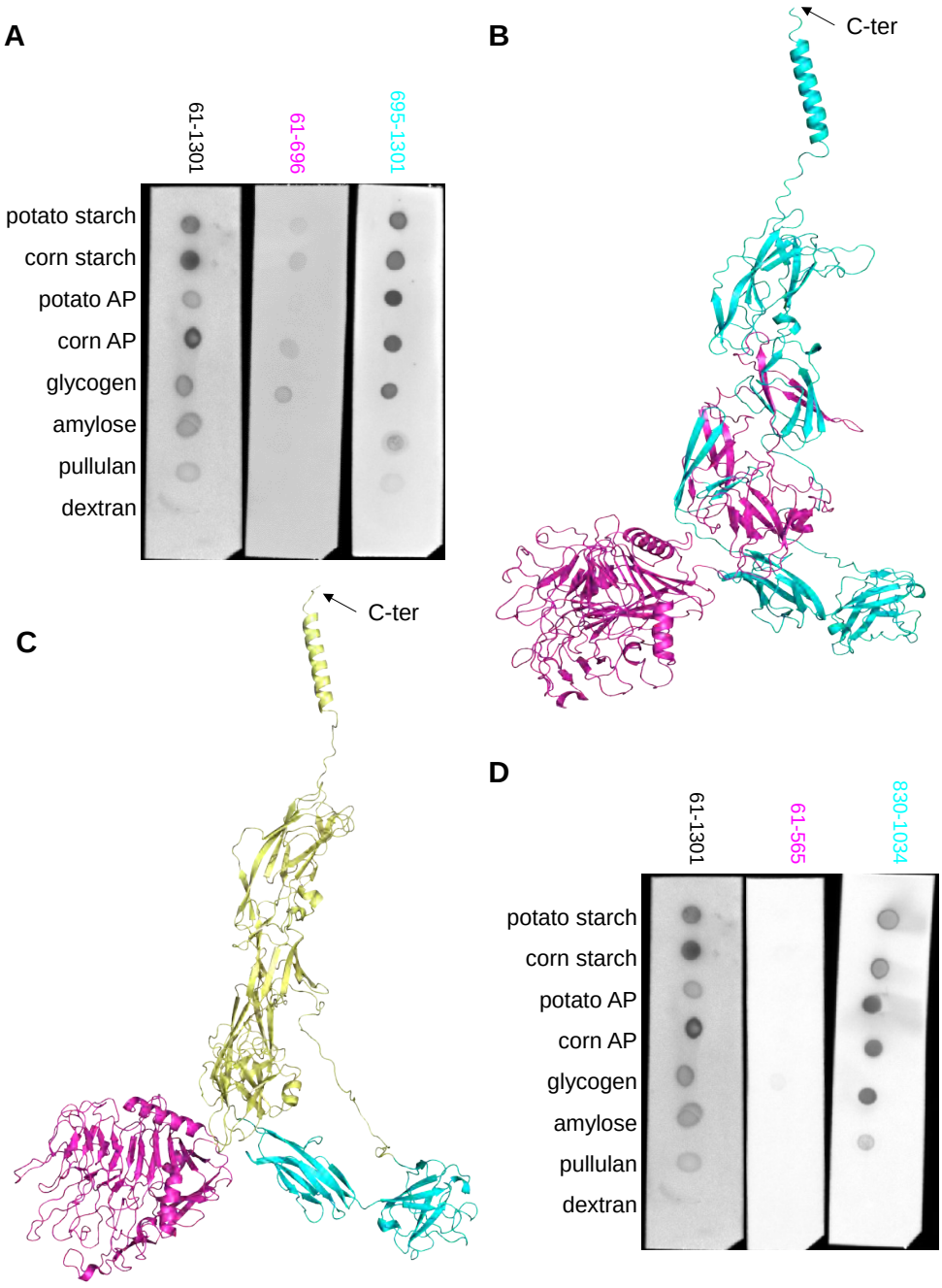


Figure B.3: The C-terminal globular domain binds starch

A. Polysaccharide macroarray in which 2μL of each polysaccharide is spotted on a nitrocellulose membrane, incubated with His_6 tagged protein and detected with mouse α-His_6-AP antibody and NBT/BCIP. The residue ranges of each BSP-Tip construct are listed above each blot. From left to right the residue range is given for the FL BSP-Tip, BSP-Tip-N1, and BSP-Tip-C1 respectively. **B.** An updated AlphaFold prediction of a homologous protein from AlphaFoldDB (UniProt A0A8B3ANM0) (225). The region corresponding to BSP-Tip-N1 is colored in magenta and the region corresponding to BSP-Tip-C1 colored in cyan. The C-terminal helix is oriented upward at the top of the protein. **C.** New constructs of just globular regions, BSP-Tip-G1 and BSP-Tip-G2, are colored in magenta and cyan respectively on the AlphaFoldDB homolog. The remainder of the protein is colored in yellow. **D.** Polysaccharide macroarray of BSP-Tip, BSP-Tip-G1 (magenta) and BSP-Tip-G2 (cyan).

for which there is a determined protein sequence, we were able to search the database for a closely related homolog of BSP-Tip. The closest match is an uncharacterized protein from a different strain of *B. adolescentis* with the UniProt ID A0A8B3ANM0 (**Fig. B.3B**). Though they share 86% sequence identity, this protein differs in that it doesn't include the N-terminal signal peptide and instead starts at the corresponding residue M96. The prediction is graded with high confidence for each of the individual domains with lower confidence for the linkers between the domains. While it is important to remain skeptical of the prediction, we can learn about how the protein may *likely* be folding in one conformation. In the prediction, there is a shaft region composed of Ig-like domains and two globular regions coming off of the shaft (**Fig. B.3B**). We mapped our N-terminal and C-terminal constructs and found that the protein has an unusual arrangement in which the N-terminal construct extends into the shaft region and starts the fold of two of the Ig-like domains that are completed by the C-terminus. It is important to recall that this is a protein structure prediction, but we used that prediction to design new constructs of just the globular domains (**Fig. B.3C**). We cloned and expressed the N-terminal globular region (BSP-Tip-G1) and the C-terminal globular region (BSP-Tip-G2) and tested their binding to starch. We found that there was only a very faint signal for corn amylopectin (**Fig. B.3D**) in the BSP-Tip-G1 blot. In contrast, BSP-Tip-G2 binds all of the same polysaccharides as the FL construct except that it shows no signal for pullulan (**Fig. B.3D**). This further narrows down BSP-Tip-G2 as the primary domain driving starch-binding. The region from 836-916 is annotated as a bacterial Ig-like domain and the region spanning 936-1034 is unannotated (171). There is no predicted starch-binding domain in this region so the molecular basis for starch-binding remains unknown.

B.2.4 Single Particle electron microscopy (EM) of BSP-Tip

To better understand how BSP-Tip might be binding starch, we moved from a structure prediction approach to determining the available structures that BSP-Tip can take on in solution. We therefore used single particle electron microscopy to evaluate the available conformations of BSP-Tip. We first performed negative stain EM to determine the most abundant 2D classes (**Fig. B.4A**). We then used cryoSPARC to create *ab initio* 3D reconstructions of the top 2D classes (176) (**Fig. B.4B**). In the most abundant class, the map shows a shaft-like region flanked on one end by two round globular regions. The two other classes likely represent a retraction of some of the shaft Ig-like domains. When the AlphaFold prediction structure is placed in the map, it becomes clear which domains may be part of the shaft and which may constitute the globular regions. These roughly correspond to the globular regions we tested for binding. Protruding from the tip of the shaft region is a C-terminal helix that is predicted to be involved in pilin oligomerization leading us to hypothesize that the tip of the shaft links BSP-Tip to the backbone pilins that make up the pilus

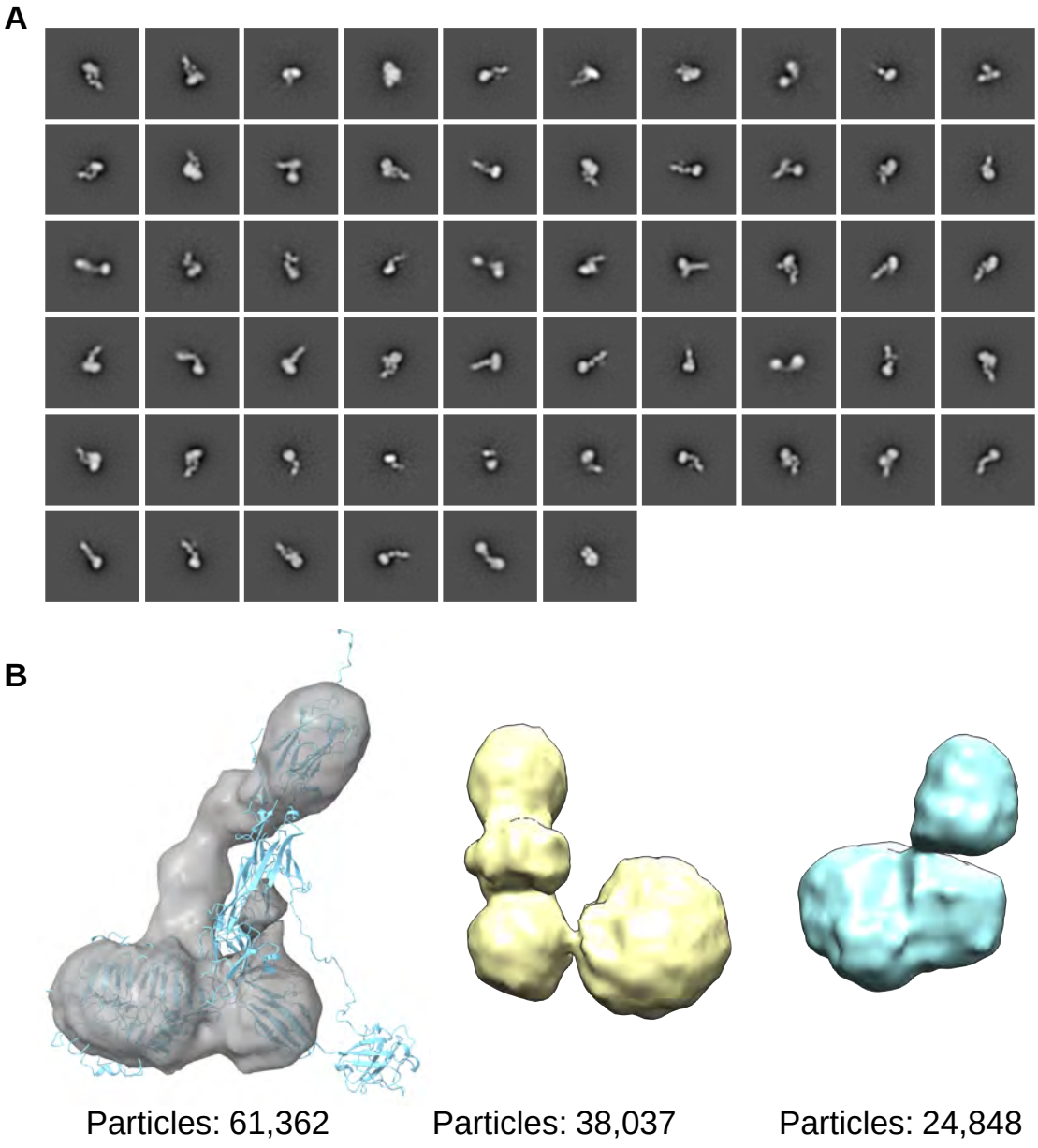


Figure B.4: BSP-Tip adopts several conformations in solution
A. 2D classes of negative stained BSP-Tip. Particles were picked and 2D classes were created using crYOLO within cryoSPARC. See Methods for details. **B.** *ab inito* 3D reconstructions of the top 2D classes with the number of particles that bin into that class listed below. The AlphaFold prediction in light blue is placed in the top 3D reconstruction for comparison.

stalk and the globular regions protrude into solution for starch granule capture.

B.3 DISCUSSION

We have previously described methods of starch-binding (Ch. 2) and starch breakdown at the cell surface (Ch. 3). In this appendix we characterized a mode of long-range starch granule capture by *B. adolescentis* U269. We found that *B. adolescentis* U269 makes a dedicated starch-binding pilus, BSP that is composed of backbone pilins that give the pilus stalk its reach, and a tip pilin, BSP-Tip that specifically binds starch. As the BSP-Tip had no obvious known starch-binding domains, we parsed the protein into short constructs to determine which region is primarily responsible for binding. We found that a globular region from 836-1036 has the ability to bind all starch substrates except pullulan.

We then took a structural approach to determine the potential conformations of BSP-Tip by single particle electron microscopy. Using only negative stain, we generated 2D class averages and an *ab initio* 3D reconstruction. We found that BSP-Tip has structural heterogeneity in solution but that generally there is a shaft region that we hypothesize is where the pilin assembles into the overall BSP and two adjacent globular regions, one of which is the primary starch-binding region. These globular domains likely prime the BSP for starch capture. Together the overall structure of BSP-Tip and its enigmatic starch-binding regions prime BSP-Tip for long-range starch granule capture.

B.4 METHODS

B.4.1 Immunofluorescence

We grew *Bifidobacterium adolescentis* U269 cells to stationary phase overnight on Yeast Casitone Fatty Acid media (241) with 0.1% potato amylopectin as a carbon source. We spun down cells at 5,000xg and washed in 1mL of 1X PBS (composition). A 1μL inoculating loop was used to spread 1μL on microscope slides (Fisher) inside of 2 circles drawn by a hydrophobic marker. Cells were fixed for 10min with 10% formalin and then washed 3x in PBS. We then blocked the cells in blocking buffer (1% BSA, 10% goat serum in PBS) for 30min. Custom antiserum against either BSP-Tip or BSP-Backbone was diluted 1:1000 in blocking buffer and applied to 1 of the 2 samples on each slide while blocking buffer was applied to the other as a negative control. Slides were incubated in a humid chamber at room temperature for 1 hour and then washed 3x5min with PBS. Goat anti-rabbit AlexaFluor 488 secondary was diluted 1:500 in blocking buffer and applied to

slides for 30 min at room temperature, then washed 3x5min in PBS. Slides were coverslipped with ProlongGold and allowed to dry in the dark overnight.

B.4.2 Protein Expression and Purification

Generally, protein cloning, expression, and purification was performed as previously described (see 2.5.1). Briefly, constructs were cloned from *B. adolescentis* U269 using primers with matching overhangs to an N-His pETite vector. The plasmid was propogated in *E. cloni* 10G cells and sequenced for sequence fidelity. *E. coli* ROSETTA pLysS cells were transformed with the plasmid for IPTG inducible expression. Cells were induced at mid-log phase and grown for 16 hours and room temperature. Cells were sonicated and the protein of interest was purified from the cell lysate by Ni^{2+} affinity purification (see Ch. 2.5. At this point we proceeded with the polysaccharide macroarray or further purified for electron microscopy.

B.4.3 Polysaccharide macroarray

Custom antisera against BSP-Tip and BSP-Backbone were generated by the immunization of rabbits with recombinantly expressed BSP-Tip and BSP-Backbone proteins (Lampire). A macroarray screen for carbohydrate binding by CBMs was performed as previously described (223). Potato amylopectin, maize amylopectin, solubilized amylose, glycogen, pullulan, and dextran (Sigma) were prepared as 0.1% solutions. Amylose was solubilized by dissolving in 1M NaOH and the addition of HCl until the solution reached pH7. 2µL of each polysaccharide solution was blotted on a nitrocellulose membrane and allowed to air dry for 2hr. CBM binding was initiated by the addition of 500μg of CBM construct diluted in 3mL of EveryBlot blocking buffer (Bio-Rad). Blots were incubated with protein for 1hr at room temperature and subsequently blotted with mouse α-polyHis-AP antibody for 1hr, washed 3x5min with TBS-T, and detected using an NBT/BCIP colorimetric substrate. Membranes were imaged on a Gel-Doc+ (Bio-Rad) with the non-colorimetric blot setting with a manual exposure of 0.2s.

B.4.4 Protein Purification for Negative stain EM

BSP-Tip was affinity purified with Ni^{2+} resin to capture His_6-tagged BSP-Tip (Binding buffer: 20mM Tris-HCl pH8, 300mM NaCl; Elution buffer: 20mM Tris-HCl pH8, 300mM NaCl, 500mM imidazole). The tag was then cleaved with a His_6-tagged Tev protease and then affinity purified with Ni^{2+} resin again, this time collecting the flow-through untagged protein. The purified protein sample was then concentrated to 1mL and subjected to size exclusion chromatography with a Sephacryl S-200 column in buffer (20mM HEPES, 100mM NaCl). The main peak was monodisperse. The

SEC-purified protein was then dialyzed into 20mM Tris buffer pH7 with no salt and further purified by anion exchange. The protein was eluted by an increasing gradient of 500mM NaCl.

B.4.5 Negative stain

For negative staining we first glow discharged Formvar/carbon-coated copper grid with 200 mesh size. 3μL of 0.008mg/mL protein in 20mM HEPES, 100mM NaCl was placed on the grid for 1 min. The grids were washed twice by floating on drops of water. The grids were floated briefly in 0.75% uranyl formate and moved to a new drop and floated for 1 min before blotting dry. The negative stained grids were screened on a 100kV Morgagni transmission electron microscope (TEM) in the University of Michigan Cryo-EM core facility. Grids that were well-stained with well-dispersed particles were then imaged on a 120kV T12 TEM.

B.4.6 2D Classification and 3D reconstruction

We used cryoSPARC (176) for particle picking, alignment, and 2D classification. In total, 750 micrographs were used. CTF correction was performed with CTF4FIND (186), using Amplitude contrast of 0.4. We used crYOLO (229) to perform automated particle picking. In total 182,745 particles were extracted leading to 176,903 particles after exposure curation. The automated particle diameter average was 109 pixels. The extracted particles generated 3 2D class averages. We then generated an *ab initio* reconstruction of the 2D classes. The resulting maps were masked to remove noise.

BIBLIOGRAPHY

[1] Abbott, D. W., and Boraston, A. B. *Chapter eleven - Quantitative Approaches to The Analysis of Carbohydrate-Binding Module Function*, vol. 510. Academic Press, 2012, pp. 211–231.

[2] Adlerberth, I., and Wold, A. Establishment of the gut microbiota in western infants. *Acta paediatrica 98*, 2 (2009), 229–238.

[3] Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr 68*, Pt 4 (2012), 352–67.

[4] Agirre, J., Iglesias-Fernández, J., Rovira, C., Davies, G. J., Wilson, K. S., and Cowtan, K. D. Privateer: software for the conformational validation of carbohydrate structures. *Nature Structural & Molecular Biology 22*, 11 (2015), 833–834.

[5] Akdel, M., Pires, D. E., Pardo, E. P., Jänes, J., Zalevsky, A. O., Mészáros, B., Bryant, P., Good, L. L., Laskowski, R. A., Pozzati, G., et al. A structural biology community assessment of alphafold2 applications. *Nature Structural & Molecular Biology 29*, 11 (2022), 1056–1067.

[6] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. Basic local alignment search tool. *J Mol Biol 215*, 3 (1990), 403–10.

[7] Amaretti, A., Raimondi, S., Volpi, N., and Rossi, M. In vitro assessment of prebiotic activity. *Bifidobacteria: Methods and Protocols* (2021), 209–223.

[8] Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M., et al. Enterotypes of the human gut microbiome. *nature 473*, 7346 (2011), 174–180.

[9] Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. Consurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res 44*, W1 (2016), W344–50.

[10] Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. Consurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res 38*, Web Server issue (2010), W529–33.

[11] Asp, N.-G., and Björck, I. Resistant starch. *Trends in Food Science & Technology 3* (1992), 111–114.

[12] Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science 307*, 5717 (2005), 1915–20.

[13] Báez Bolivar, E. G., Bui, D. T., Kitova, E. N., Han, L., Zheng, R. B., Luber, E. J., Sayed, S. Y., Mahal, L. K., and Klassen, J. S. Submicron emitters enable reliable quantification of weak protein–glycan interactions by esi-ms. *Analytical Chemistry 93*, 9 (2021), 4231–4239.

[14] Barba-Cedillo, V., and Montanier, C. Y. Effect of multimodularity and spatial organization of glycoside hydrolases on catalysis. *Essays in Biochemistry 67*, 3 (04 2023), 629–638.

[15] Baxter, N. T., Schmidt, A. W., Venkataraman, A., Kim, K. S., Waldron, C., and Schmidt, T. M. Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers. *MBio 10*, 1 (2019).

[16] Bayer, E. A., Morag, E., and Lamed, R. The cellulosome–a treasure-trove for biotechnology. *Trends Biotechnol 12*, 9 (1994), 379–86.

[17] Belenguer, A., Duncan, S. H., Calder, A. G., Holtrop, G., Louis, P., Lobley, G. E., and Flint, H. J. Two routes of metabolic cross-feeding between bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. *Appl Environ Microbiol 72*, 5 (2006), 3593–9.

[18] Bergman, E. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiological reviews 70*, 2 (1990), 567–590.

[19] Berry, C. Resistant starch: formation and measurement of starch that survives exhaustive digestion with amylolytic enzymes during the determination of dietary fibre. *Journal of cereal science 4*, 4 (1986), 301–314.

[20] Bertoft, E. Understanding starch structure: Recent progress. *Agronomy v. 7*, no. 3 (2017), 2017 v.7 no.3.

[21] Birt, D. F., Boylston, T., Hendrich, S., Jane, J.-L., Hollis, J., Li, L., McClelland, J., Moore, S., Phillips, G. J., Rowling, M., et al. Resistant starch: promise for improving human health. *Advances in nutrition 4*, 6 (2013), 587–601.

[22] Bjursell, M. K., Martens, E. C., and Gordon, J. I. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, bacteroides thetaiotaomicron, to the suckling period. *Journal of biological chemistry 281*, 47 (2006), 36269–36279.

[23] Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Necci, M., Orengo, C. A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J. A., Sillitoe, I., Thanki, N., Thomas, P. D., Tosatto, S. C. E.,

Wu, C. H., Bateman, A., and Finn, R. D. The interpro protein families and domains database: 20 years on. *Nucleic Acids Res 49*, D1 (2021), D344–D354.

[24] Bolam, D. N., and van den Berg, B. Tonb-dependent transport by the gut microbiota: novel aspects of an old problem. *Current opinion in structural biology 51* (2018), 35–43.

[25] Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical Journal 382*, Pt 3 (Sep 2004), 769–781.

[26] Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A., and Law, V. A structural and functional analysis of alpha-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *J Biol Chem 281*, 1 (2006), 587–98.

[27] Boraston, A. B., Tomme, P., Amandoron, E. A., and Kilburn, D. G. A novel mechanism of xylan binding by a lectin-like module from streptomyces lividans xylanase 10a. *The Biochemical Journal 350 Pt 3*, Pt 3 (Sep 2000), 933–941.

[28] Božić, N., Lončar, N., Slavić, M. Š., and Vujčić, Z. Raw starch degrading α-amylases: an unsolved riddle. *Amylase 1*, 1 (2017), 12–25.

[29] Brewer, M. K., and Gentry, M. S. Brain glycogen structure and its associated proteins: Past, present and future. *Adv Neurobiol 23* (2019), 17–81.

[30] Brownlee, I. A., Havler, M. E., Dettmar, P. W., Allen, A., and Pearson, J. P. Colonic mucus: secretion and turnover in relation to dietary fibre intake. *Proceedings of the Nutrition Society 62*, 1 (2003), 245–249.

[31] Cameron, E. A., Maynard, M. A., Smith, C. J., Smith, T. J., Koropatkin, N. M., and Martens, E. C. Multidomain carbohydrate-binding proteins involved in bacteroides thetaiotaomicron starch metabolism. *Journal of Biological Chemistry 287*, 41 (2012), 34614–34625.

[32] Canani, R. B., Di Costanzo, M., Leone, L., Pedata, M., Meli, R., and Calignano, A. Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. *World journal of gastroenterology: WJG 17*, 12 (2011), 1519.

[33] Candussio, A., Schmid, G., and Bock, A. Biochemical and genetic analysis of a maltopentaose-producing amylase from an alkaliphilic gram-positive bacterium. *Eur J Biochem 191*, 1 (1990), 177–85.

[34] Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., and Ben-Tal, N. Consurf: Using evolutionary data to raise testable hypotheses about protein function. 199–206.

[35] Cerqueira, F. M., Photenhauer, A. L., Doden, H. L., Brown, A. N., Abdel-Hamid, A. M., Morais, S., Bayer, E. A., Wawrzak, Z., Cann, I., Ridlon, J. M., Hopkins, J., and Koropatkin, N. M. Sas20 is a highly flexible starch-binding protein in the ruminococcus bromii cell-surface amylosome. *J Biol Chem* (2022), 101896.

[36] Cerqueira, F. M., Photenhauer, A. L., Pollet, R. M., Brown, H. A., and Koropatkin, N. M. Starch digestion by gut bacteria: Crowdsourcing for carbs. *Trends Microbiol 28*, 2 (2020), 95–108.

[37] Cho, K. H., and Salyers, A. A. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by bacteroides thetaiotaomicron. *J Bacteriol 183*, 24 (2001), 7224–30.

[38] Cockburn, D., Nielsen, M. M., Christiansen, C., Andersen, J. M., Rannes, J. B., Blennow, A., and Svensson, B. Surface binding sites in amylase have distinct roles in recognition of starch structure motifs and degradation. *Int J Biol Macromol 75* (2015), 338–45.

[39] Cockburn, D. W., Suh, C., Medina, K. P., Duvall, R. M., Wawrzak, Z., Henrissat, B., and Koropatkin, N. M. 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. *Molecular microbiology 107*, 2 (2018), 249–264.

[40] Collins, M. D., and Gibson, G. R. Probiotics, prebiotics, and synbiotics: approaches for modulating the microbial ecology of the gut. *The American journal of clinical nutrition 69*, 5 (1999), 1052s–1057s.

[41] Conway, J. M., Pierce, W. S., Le, J. H., Harper, G. W., Wright, J. H., Tucker, A. L., Zurawski, J. V., Lee, L. L., Blumer-Schuette, S. E., and Kelly, R. M. Multidomain, surface layer-associated glycoside hydrolases contribute to plant polysaccharide degradation by caldicellulosiruptor species. *J Biol Chem 291*, 13 (2016), 6732–47.

[42] Cripwell, R. A., Favaro, L., Viljoen-Bloom, M., and van Zyl, W. H. Consolidated bioprocessing of raw starch to ethanol by saccharomyces cerevisiae: achievements and challenges. *Biotechnology Advances 42* (2020), 107579.

[43] Crittenden, R., Laitila, A., Forssell, P., Matto, J., Saarela, M., Mattila-Sandholm, T., and Myllarinen, P. Adhesion of bifidobacteria to granular starch and its implications in probiotic technologies. *Appl Environ Microbiol 67*, 8 (2001), 3469–75.

[44] Cummings, J. H., and Macfarlane, G. T. The control and consequences of bacterial fermentation in the human colon. *J Appl Bacteriol 70*, 6 (1991), 443–59.

[45] Danne, C., and Dramsi, S. Pili of gram-positive bacteria: roles in host colonization. *Research in microbiology 163*, 9-10 (2012), 645–658.

[46] Datta, S., Shamala, N., Banerjee, A., Pramanik, A., Bhattacharjya, S., and Balaram, P. Characterization of helix terminating schellman motifs in peptides. crystal structure and nuclear overhauser effect analysis of a synthetic heptapeptide helix. *Journal of the American Chemical Society 119*, 39 (1997), 9246–9251.

[47] De Vrese, M., and Schrezenmeir, J. Probiotics, prebiotics, and synbiotics. *Food biotechnology* (2008), 1–66.

[48] D'Elia, J. N., and Salyers, A. A. Effect of regulatory protein levels on utilization of starch by bacteroides thetaiotaomicron. *Journal of bacteriology 178*, 24 (1996), 7180–7186.

[49] Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., Pudlo, N. A., Kitamoto, S., Terrapon, N., Muller, A., et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell 167*, 5 (2016), 1339–1353.

[50] Di Tommaso, N., Gasbarrini, A., and Ponziani, F. R. Intestinal barrier in human health and disease. *International journal of environmental research and public health 18*, 23 (2021), 12836.

[51] Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., Taly, J. F., and Notredame, C. T-coffee: a web server for the multiple sequence alignment of protein and rna sequences using structural information and homology extension. *Nucleic Acids Res 39*, Web Server issue (2011), W13–7.

[52] Dobranowski, P. A., and Stintzi, A. Resistant starch, microbiome, and precision modulation. *Gut Microbes 13*, 1 (2021), 1926842.

[53] Dreher, M. L., Dreher, C. J., Berry, J. W., and Fleming, S. E. Starch digestibility of foods: a nutritional perspective. *Critical Reviews in Food Science & Nutrition 20*, 1 (1984), 47–71.

[54] Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res 50*, D1 (2022), D571–D577.

[55] Du, Z., Su, H., Wang, W., Ye, L., Wei, H., Peng, Z., Anishchenko, I., Baker, D., and Yang, J. The trrosetta server for fast and accurate protein structure prediction. *Nature protocols 16*, 12 (2021), 5634–5651.

[56] Duncan, S. H., Louis, P., and Flint, H. J. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Applied and environmental microbiology 70*, 10 (2004), 5810–5817.

[57] Dupont, A., Heinbockel, L., Brandenburg, K., and Hornef, M. W. Antimicrobial peptides and the enteric mucus layer act in concert to protect the intestinal mucosa. *Gut microbes 5*, 6 (2014), 761–765.

[58] Duranti, S., Turroni, F., Lugli, G. A., Milani, C., Viappiani, A., Mangifesta, M., Gioiosa, L., Palanza, P., van Sinderen, D., and Ventura, M. Genomic characterization and transcriptional studies of the starch-utilizing strain bifidobacterium adolescentis 22l. *Appl Environ Microbiol 80*, 19 (2014), 6080–90.

[59] El Omari, K., Iourin, O., Kadlec, J., Fearn, R., Hall, D. R., Harlos, K., Grimes, J. M., and Stuart, D. I. Pushing the limits of sulfur sad phasing: de novo structure solution of the n-terminal domain of the ectodomain of hcv e1. *Acta Crystallogr D Biol Crystallogr 70*, Pt 8 (2014), 2197–203.

[60] Emsley, P., and Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr 60*, Pt 12 Pt 1 (2004), 2126–32.

[61] Englyst, H., Wiggins, H. S., and Cummings, J. H. Determination of the non-starch polysaccharides in plant foods by gas-liquid chromatography of constituent sugars as alditol acetates. *Analyst 107*, 1272 (1982), 307–318.

[62] Englyst, H. N., and Cummings, J. H. Non-starch polysaccharides (dietary fiber) and resistant starch. *New Developments in Dietary Fiber: Physiological, Physicochemical, and Analytical Aspects* (1990), 205–225.

[63] Englyst, H. N., Kingman, S., and Cummings, J. Classification and measurement of nutritionally important starch fractions. *European journal of clinical nutrition 46* (1992), S33–50.

[64] Evans, P. R., and Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr 69*, Pt 7 (2013), 1204–14.

[65] Faghfoori, Z., Faghfoori, M. H., Saber, A., Izadi, A., and Yari Khosroushahi, A. Anticancer effects of bifidobacteria on colon cancer cell lines. *Cancer cell international 21*, 1 (2021), 1–12.

[66] Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution 39*, 4 (1985), 783–791.

[67] Ficko-Blean, E., Gregg, K. J., Adams, J. J., Hehemann, J.-H., Czjzek, M., Smith, S. P., and Boraston, A. B. Portrait of an enzyme, a complete structural analysis of a multimodular β-n-acetylglucosaminidase from clostridium perfringens. *Journal of Biological Chemistry 284*, 15 (2009), 9876–9884.

[68] Flowers, S. A., Baxter, N. T., Ward, K. M., Kraal, A. Z., McInnis, M. G., Schmidt, T. M., and Ellingrod, V. L. Effects of atypical antipsychotic treatment and resistant starch supplementation on gut microbiome composition in a cohort of patients with bipolar disorder or schizophrenia. *Pharmacotherapy 39*, 2 (2019), 161–170.

[69] Foley, M. H., Cockburn, D. W., and Koropatkin, N. M. The sus operon: a model system for starch uptake by the human gut bacteroidetes. *Cellular and Molecular Life Sciences 73* (2016), 2603–2617.

[70] Foresti, M. L., Williams, M. d. P., Martínez-García, R., and Vázquez, A. Analysis of a preferential action of α-amylase from b. licheniformis towards amorphous regions of waxy maize starch. *Carbohydrate Polymers 102* (2014), 80–87.

[71] Foroni, E., Serafini, F., Amidani, D., Turroni, F., He, F., Bottacini, F., O'Connell Motherway, M., Viappiani, A., Zhang, Z., Rivetti, C., et al. Genetic analysis and morphological identification of pilus-like structures in members of the genus bifidobacterium. In *Microbial Cell Factories* (2011), vol. 10, BioMed Central, pp. 1–13.

[72] Franke, D., Jeffries, C. M., and Svergun, D. I. Machine learning methods for x-ray scattering data analysis from biomacromolecular solutions. *Biophysical journal 114*, 11 (2018), 2485–2492.

[73] Fraser, J. S., Yu, Z., Maxwell, K. L., and Davidson, A. R. Ig-like domains on bacteriophages: A tale of promiscuity and deceit. *Journal of Molecular Biology 359*, 2 (2006), 496–507.

[74] Fujimoto, Z., Takase, K., Doui, N., Momma, M., Matsumoto, T., and Mizuno, H. Crystal structure of a catalytic-site mutant α-amylase from bacillus subtilis complexed with maltopentaose. *Journal of molecular biology 277*, 2 (1998), 393–407.

[75] Fuller, R. Probiotics in man and animals. *The Journal of applied bacteriology 66*, 5 (1989), 365–378.

[76] Furrie, E., Macfarlane, S., Kennedy, A., Cummings, J., Walsh, S., O'neil, D., and Macfarlane, G. Synbiotic therapy (bifidobacterium longum/synergy 1) initiates resolution of inflammation in patients with active ulcerative colitis: a randomised controlled pilot trial. *Gut 54*, 2 (2005), 242–249.

[77] Gagliardi, A., Totino, V., Cacciotti, F., Iebba, V., Neroni, B., Bonfiglio, G., Trancassini, M., Passariello, C., Pantanella, F., and Schippa, S. Rebuilding the gut microbiota ecosystem. *International journal of environmental research and public health 15*, 8 (2018), 1679.

[78] Gallant, D., Bouchet, B., Buleon, A., and Perez, S. Physical characteristics of starch granules and susceptibility to enzymatic degradation. *European journal of clinical nutrition 46*, 2 (1992), 3–16.

[79] Gernat, C., Radosta, S., Anger, H., and Damaschun, G. Crystalline parts of three different conformations detected in native and enzymatically degraded starches. *Starch-Stärke 45*, 9 (1993), 309–314.

[80] Gessler, K., Usón, I., Takaha, T., Krauss, N., Smith, S. M., Okada, S., Sheldrick, G. M., and Saenger, W. V-amylose at atomic resolution: X-ray structure of a cycloamylose with 26 glucose residues (cyclomaltohexaicosaose). 4246–4251.

[81] Gibson, G. R. Dietary modulation of the human gut microflora using prebiotics. *British Journal of Nutrition 80*, S2 (1998), S209–S212.

[82] Gibson, G. R., and Roberfroid, M. B. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *J Nutr 125*, 6 (1995), 1401–12.

[83] Gille, C., Birgit, W., and Gille, A. Sequence alignment visualization in html5 without java. *Bioinformatics 30*, 1 (2014), 121–2.

[84] Gille, C., Fahling, M., Weyand, B., Wieland, T., and Gille, A. Alignment-annotator web server: rendering and annotating sequence alignments. *Nucleic Acids Res 42*, Web Server issue (2014), W3–6.

[85] Glenwright, A. J., Pothula, K. R., Bhamidimarri, S. P., Chorev, D. S., Baslé, A., Firbank, S. J., Zheng, H., Robinson, C. V., Winterhalter, M., Kleinekathöfer, U., et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature 541*, 7637 (2017), 407–411.

[86] Goodman, C., Keating, G., Georgousopoulou, E., Hespe, C., and Levett, K. Probiotics for the prevention of antibiotic-associated diarrhoea: a systematic review and meta-analysis. *BMJ open 11*, 8 (2021), e043054.

[87] Guillén, D., Sánchez, S., and Rodríguez-Sanoja, R. Carbohydrate-binding domains: multiplicity of biological roles. *Applied Microbiology and Biotechnology 85*, 5 (Feb 2010), 1241–1249.

[88] Guillen, D., Santiago, M., Linares, L., Perez, R., Morlon, J., Ruiz, B., Sanchez, S., and Rodriguez-Sanoja, R. Alpha-amylase starch binding domains: cooperative effects of binding to starch granules of multiple tandemly arranged domains. *Appl Environ Microbiol 73*, 12 (2007), 3833–7.

[89] Gulbakan, B., Barylyuk, K., Schneider, P., Pillong, M., Schneider, G., and Zenobi, R. Native electrospray ionization mass spectrometry reveals multiple facets of aptamer-ligand interactions: From mechanism to binding constants. *J Am Chem Soc 140*, 24 (2018), 7486–7497.

[90] Guo, H.-B., Perminov, A., Bekele, S., Kedziora, G., Farajollahi, S., Varaljay, V., Hinkle, K., Molinero, V., Meister, K., Hung, C., et al. Alphafold2 models indicate that protein sequence determines both structure and dynamics. *Scientific Reports 12*, 1 (2022), 10696.

[91] Hanashiro, I., Abe, J.-i., and Hizukuri, S. A periodic distribution of the chain length of amylopectin as revealed by high-performance anion-exchange chromatography. *Carbohydrate Research 283* (1996), 151–159.

[92] Hempel, S., Newberry, S. J., Maher, A. R., Wang, Z., Miles, J. N., Shanman, R., Johnsen, B., and Shekelle, P. G. Probiotics for the prevention and treatment of antibiotic-associated diarrhea: a systematic review and meta-analysis. *Jama 307*, 18 (2012), 1959–1969.

[93] Henrick, B. M., Rodriguez, L., Lakshmikanth, T., Pou, C., Henckel, E., Arzoomand, A., Olin, A., Wang, J., Mikes, J., Tan, Z., et al. Bifidobacteria-mediated immune system imprinting early in life. *Cell 184*, 15 (2021), 3884–3898.

[94] Hillmann, G. *Measurement by End-point Determination on Paper*. Academic Press, 1974, pp. 903–909.

[95] Hizukuri, S. Relationship between the distribution of the chain length of amylopectin and the crystalline structure of starch granules. *Carbohydrate Research 141*, 2 (1985), 295–306.

[96] Holm, L. Dali and the persistence of protein shape. 128–140.

[97] Holm, L. Using dali for protein structure comparison. *Methods Mol Biol 2112* (2020), 29–42.

[98] Hong, Y.-S., Jung, D.-H., Chung, W.-H., Nam, Y.-D., Kim, Y.-J., Seo, D.-H., and Park, C.-S. Human gut commensal bacterium ruminococcus species fmb-cy1 completely degrades the granules of resistant starch. *Food Science and Biotechnology 31*, 2 (2022), 231–241.

[99] Hopkins, J. B., Gillilan, R. E., and Skou, S. Bioxtas raw: improvements to a free open-source program for small-angle x-ray scattering data reduction and analysis. *J Appl Crystallogr 50*, Pt 5 (2017), 1545–1553.

[100] Hu, Y., Le Leu, R. K., Christophersen, C. T., Somashekar, R., Conlon, M. A., Meng, X. Q., Winter, J. M., Woodman, R. J., McKinnon, R., and Young, G. P. Manipulation of the gut microbiota using resistant starch is associated with protection against colitis-associated colorectal cancer in rats. *Carcinogenesis 37*, 4 (2016), 366–375.

[101] Huang, L., Zhang, H., Wu, P., Entwistle, S., Li, X., Yohe, T., Yi, H., Yang, Z., and Yin, Y. dbcan-seq: a database of carbohydrate-active enzyme (cazyme) sequence and annotation. *Nucleic Acids Research 46*, D1 (2018), D516–D521.

[102] Huang, X.-F., Nazarian, F., Vincken, J.-P., Visser, R. G. F., and Trindade, L. M. A tandem cbm25 domain of α-amylase from microbacterium aurum as potential tool for targeting proteins to starch granules during starch biosynthesis. *BMC biotechnology 17*, 1 (Dec 2017), 86.

[103] Hugon, P., Lagier, J.-C., Colson, P., Bittar, F., and Raoult, D. Repertoire of human gut microbes. *Microbial pathogenesis 106* (2017), 103–112.

[104] Imberty, A., Chanzy, H., Pérez, S., Buléon, A., and Tran, V. The double-helical nature of the crystalline part of a-starch. *Journal of Molecular Biology 201*, 2 (1988), 365–378.

[105] Imberty, A., and Pérez, S. A revisit to the three-dimensional structure of b-type starch. 1205–1221.

[106] Janeček, Š., Mareček, F., MacGregor, E. A., and Svensson, B. Starch-binding domains as cbm families-history, occurrence, structure, function and evolution. *Biotechnol Adv 37*, 8 (2019), 107451.

[107] Janeček, Š., and Svensson, B. How many α-amylase gh families are there in the cazy database? 1–10.

[108] Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., and Hunter, S. Interproscan 5: genome-scale protein function classification. *Bioinformatics 30*, 9 (2014), 1236–40.

[109] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. Highly accurate protein structure prediction with alphafold. *Nature 596*, 7873 (2021), 583–589.

[110] Jung, D. H., Kim, G. Y., Kim, I. Y., Seo, D. H., Nam, Y. D., Kang, H., Song, Y., and Park, C. S. Bifidobacterium adolescentis p2p3, a human gut bacterium having strong non-gelatinized resistant starch-degrading activity. *J Microbiol Biotechnol 29*, 12 (2019), 1904–1915.

[111] Jung, D. H., Seo, D. H., Kim, Y. J., Chung, W. H., Nam, Y. D., and Park, C. S. The presence of resistant starch-degrading amylases in bifidobacterium adolescentis of the human gut. *Int J Biol Macromol 161* (2020), 389–397.

[112] Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr 66*, Pt 2 (2010), 125–32.

[113] Kang, H. J., Paterson, N. G., Gaspar, A. H., Ton-That, H., and Baker, E. N. The corynebacterium diphtheriae shaft pilin spaa is built of tandem ig-like modules with stabilizing isopeptide and disulfide bonds. *Proceedings of the National Academy of Sciences 106*, 40 (2009), 16967–16971.

[114] Kay, R. Dietary fiber. *Journal of lipid research 23*, 2 (1982), 221–242.

[115] Khalikova, E., Susi, P., and Korpela, T. Microbial dextran-hydrolyzing enzymes: fundamentals and applications. *Microbiol Mol Biol Rev 69*, 2 (2005), 306–25.

[116] Kikhney, A. G., Borges, C. R., Molodenskiy, D. S., Jeffries, C. M., and Svergun, D. I. Sasbdb: Towards an automatically curated and validated repository for biological scattering data. *Protein Science 29*, 1 (2020), 66–75.

[117] Kim, M. J., Lee, D. K., Park, J. E., Park, I. H., Seo, J. G., and Ha, N. J. Antiviral activity of bifidobacterium adolescentis spm1605 against coxsackievirus b3. *Biotechnology & Biotechnological Equipment 28*, 4 (2014), 681–688.

[118] Kim, S. Y., Kim, H., Kim, Y. J., Jung, D. H., Seo, D. H., Jung, J. H., and Park, C. S. Enzymatic analysis of truncation mutants of a type ii pullulanase from bifidobacterium adolescentis p2p3, a resistant starch-degrading gut bacterium. *Int J Biol Macromol 193*, Pt B (2021), 1340–1349.

[119] Kirby, N., Cowieson, N., Hawley, A. M., Mudie, S. T., McGillivray, D. J., Kusel, M., Samardzic-Boban, V., and Ryan, T. M. Improved radiation dose efficiency in solution saxs using a sheath flow sample environment. *Acta Crystallographica Section D: Structural Biology 72*, 12 (2016), 1254–1266.

[120] Koropatkin, N. M., Cameron, E. A., and Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol 10*, 5 (2012), 323–35.

[121] Koropatkin, N. M., and Smith, T. J. Susg: a unique cell-membrane-associated alpha-amylase from a prominent human gut symbiont targets complex starch molecules. *Structure 18*, 2 (2010), 200–15.

[122] Krishnan, V. Pilins in gram-positive bacteria: a structural perspective. *IUBMB life 67*, 7 (2015), 533–543.

[123] Krissinel, E., and Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol 372*, 3 (2007), 774–97.

[124] Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. Mega x: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol 35*, 6 (2018), 1547–1549.

[125] Lammerts van Bueren, A., Ficko-Blean, E., Pluvinage, B., Hehemann, J.-H., Higgins, M. A., Deng, L., Ogunniyi, A. D., Stroeher, U. H., El Warry, N., Burke, R. D., Czjzek, M., Paton, J. C., Vocadlo, D. J., and Boraston, A. B. The conformation and function of a multimodular glycogen-degrading pneumococcal virulence factor. *Structure 19*, 5 (2011), 640–651.

[126] Lattimer, J. M., and Haub, M. D. Effects of dietary fiber and its components on metabolic health. *Nutrients 2*, 12 (2010), 1266–1289.

[127] Le Leu, R. K., Hu, Y., Brown, I. L., Woodman, R. J., and Young, G. P. Synbiotic intervention of bifidobacterium lactis and resistant starch protects against colorectal cancer development in rats. *Carcinogenesis 31*, 2 (2010), 246–251.

[128] Le Leu, R. K., Young, G. P., Hu, Y., Winter, J., and Conlon, M. A. Dietary red meat aggravates dextran sulfate sodium-induced colitis in mice whereas resistant starch attenuates inflammation. *Digestive diseases and sciences 58* (2013), 3475–3482.

[129] Le Poul, E., Loison, C., Struyf, S., Springael, J.-Y., Lannoy, V., Decobecq, M.-E., Brezillon, S., Dupriez, V., Vassart, G., Van Damme, J., et al. Functional characterization of human receptors for short chain fatty acids and their role in polymorphonuclear cell activation. *Journal of Biological Chemistry 278*, 28 (2003), 25481–25489.

[130] Lee, D. K., Kang, J. Y., Shin, H. S., Park, I. H., and Ha, N. J. Antiviral activity of bifidobacterium adolescentis spm0212 against hepatitis b virus. *J Archives of pharmacal research 36*, 12 (2013), 1525–1532.

[131] Lee, J., Shin, M.-K., Ryu, D.-K., Kim, S., and Ryu, W.-S. Insertion and deletion mutagenesis by overlap extension pcr. *In Vitro Mutagenesis Protocols: Third Edition* (2010), 137–146.

[132] Leitch, E. C., Walker, A. W., Duncan, S. H., Holtrop, G., and Flint, H. J. Selective colonization of insoluble substrates by human faecal bacteria. *Environ Microbiol 9*, 3 (2007), 667–79.

[133] Letunic, I., and Bork, P. Interactive tree of life (itol): an online tool for phylogenetic tree display and annotation. *Bioinformatics 23*, 1 (2007), 127–8.

[134] Li, D., Breiman, A., Le Pendu, J., and Uyttendaele, M. Anti-viral effect of bifidobacterium adolescentis against noroviruses. *J Frontiers in microbiology 7* (2016), 864.

[135] Lin, C., Lin, Y., Zhang, H., Wang, G., Zhao, J., Zhang, H., and Chen, W. Intestinal 'infant-type'bifidobacteria mediate immune system development in the first 1000 days of life. *Nutrients 14*, 7 (2022), 1498.

[136] Little, E., Bork, P., and Doolittle, R. F. Tracing the spread of fibronectin type iii domains in bacterial glycohydrolases. *Journal of Molecular Evolution 39*, 6 (1994), 631–643.

[137] Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. The carbohydrate-active enzymes database (cazy) in 2013. *Nucleic Acids Res 42*, Database issue (2014), D490–5.

[138] Loos, M., Gerber, C., Corona, F., Hollender, J., and Singer, H. Accelerated isotope fine structure calculation using pruned transition trees. *Analytical Chemistry 87*, 11 (2015), 5738–5744.

[139] Macfarlane, G. T., and Englyst, H. N. Starch utilization by the human large intestinal microflora. *J Appl Bacteriol 60*, 3 (1986), 195–201.

[140] Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. Search and sequence analysis tools services from embl-ebi in 2022. *Nucleic Acids Res 50*, W1 (2022), W276–W279.

[141] Mahmood, T., Turner, M. A., and Stoddard, F. L. Comparison of methods for colorimetric amylose determination in cereal grains. *Starch - Stärke 59*, 8 (2007), 357–365.

[142] Makki, K., Deehan, E. C., Walter, J., and Backhed, F. The impact of dietary fiber on gut microbiota in host health and disease. *Cell Host Microbe 23*, 6 (2018), 705–715.

[143] Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., and Franke, D. Atsas 3.0: expanded functionality and new tools for small-angle scattering data analysis. *Journal of Applied Crystallography 54*, 1 (2021), 343–355.

[144] Mareček, F., and Janeček, S. A novel subfamily gh13_46 of the α-amylase family gh13 represented by the cyclomaltodextrinase from flavobacterium sp. no. 92. 8735.

[145] Martens, B. M., Gerrits, W. J., Bruininx, E. M., and Schols, H. A. Amylopectin structure and crystallinity explains variation in digestion kinetics of starches across botanic sources in an in vitro pig model. *Journal of animal science and biotechnology 9*, 1 (2018), 1–13.

[146] Martens, E. C., Koropatkin, N. M., Smith, T. J., and Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the bacteroidetes sus-like paradigm. *Journal of Biological Chemistry 284*, 37 (2009), 24673–24677.

[147] Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K., Benesch, J. L., and Robinson, C. V. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal Chem 87*, 8 (2015), 4370–6.

[148] Matsui, M., Kakuta, M., and Misaki, A. Comparison of the unit-chain distributions of glycogens from different biological sources, revealed by anion exchange chromatography. *Bioscience, Biotechnology, and Biochemistry 57*, 4 (1993), 623–627.

[149] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. Phaser crystallographic software. *Journal of Applied Crystallography 40*, 4 (2007), 658–674.

[150] McKee, L. S. *Measuring Enzyme Kinetics of Glycoside Hydrolases Using the 3,5-Dinitrosalicylic Acid Assay*. Springer New York, New York, NY, 2017, pp. 27–36.

[151] Melendez-Hevia, E., Waddell, T., and Shelton, E. Optimization of molecular design in the evolution of metabolism: the glycogen molecule. *Biochemical Journal 295*, 2 (1993), 477–483.

[152] Micah, H., Claire, F.-L., Rob, K., et al. The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature 449*, 7164 (2007), 804–810.

[153] Miguez Amil, S., Jiménez-Ortega, E., Ramírez-Escudero, M., Talens-Perales, D., Marín-Navarro, J., Polaina, J., Sanz-Aparicio, J., and Fernandez-Leiro, R. The cryo-em structure of thermotoga maritima β-galactosidase: quaternary structure guides protein engineering. *ACS Chemical Biology 15*, 1 (2019), 179–188.

[154] Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., Belzer, C., Delgado Palacio, S., Arboleya Montes, S., Mancabelli, L., et al. The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. *Microbiology and molecular biology reviews 81*, 4 (2017), e00036–17.

[155] Milles, L. F., Unterauer, E. M., Nicolaus, T., and Gaub, H. E. Calcium stabilizes the strongest protein fold. *Nature communications 9*, 1 (2018), 4764.

[156] Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. Colabfold: making protein folding accessible to all. *Nature methods 19*, 6 (2022), 679–682.

[157] Møller, M. S., and Svensson, B. Structural biology of starch-degrading enzymes and their regulation. *Current opinion in structural biology 40* (2016), 33–42.

[158] Moore, W., CATO, E. P., and HOLDEMAN, L. V. Ruminococcus bromii sp. n. and emendation of the description of ruminococcus sijpestein. *International Journal of Systematic and Evolutionary Microbiology 22*, 2 (1972), 78–80.

[159] Mukerjea, R., and Robyt, J. F. Isolation, structure, and characterization of the putative soluble amyloses from potato, wheat, and rice starches. *Carbohydrate research 345*, 3 (2010), 449–451.

[160] Mukhopadhya, I., Morais, S., Laverde-Gomez, J., Sheridan, P. O., Walker, A. W., Kelly, W., Klieve, A. V., Ouwerkerk, D., Duncan, S. H., Louis, P., Koropatkin, N., Cockburn, D., Kibler, R., Cooper, P. J., Sandoval, C., Crost, E., Juge, N., Bayer, E. A., and Flint, H. J. Sporulation capability and amylosome conservation among diverse human colonic and rumen isolates of the keystone starch-degrader ruminococcus bromii. *Environ Microbiol 20*, 1 (2018), 324–336.

[161] Murphy, R. D., Chen, T., Lin, J., He, R., Wu, L., Pearson, C. R., Sharma, S., Vander Kooi, C. D., Sinai, A. P., Zhang, Z.-Y., Vander Kooi, C. W., and Gentry, M. S. The toxoplasma glucan phosphatase tglaforin utilizes a distinct functional mechanism that can be exploited by therapeutic inhibitors. *Journal of Biological Chemistry 298*, 7 (2022), 102089.

[162] Naim, H., Sterchi, E., and Lentze, M. Structure, biosynthesis, and glycosylation of human small intestinal maltase-glucoamylase. *Journal of Biological Chemistry 263*, 36 (1988), 19709–19717.

[163] Nash, M. A., Smith, S. P., Fontes, C. M., and Bayer, E. A. Single versus dual-binding conformations in cellulosomal cohesin-dockerin complexes. *Curr Opin Struct Biol 40* (2016), 89–96.

[164] Notenboom, V., Boraston, A. B., Kilburn, D. G., and Rose, D. R. Crystal structures of the family 9 carbohydrate-binding module from thermotoga maritima xylanase 10a in native and ligand-bound forms. *Biochemistry 40*, 21 (2001), 6248–56.

[165] Notredame, C., Higgins, D. G., and Heringa, J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol 302*, 1 (2000), 205–17.

[166] O'Callaghan, A., and van Sinderen, D. Bifidobacteria and their role as members of the human gut microbiota. *Front Microbiol 7* (2016), 925.

[167] O'Connell Motherway, M., Fitzgerald, G. F., Neirynck, S., Ryan, S., Steidler, L., and van Sinderen, D. Characterization of apub, an extracellular type ii amylopullulanase from bifidobacterium breve ucc2003. *Appl Environ Microbiol 74*, 20 (2008), 6271–9.

[168] O'Connell Motherway, M., O'Driscoll, J., Fitzgerald, G. F., and Van Sinderen, D. Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in bifidobacterium breve ucc2003. *Microbial biotechnology 2*, 3 (2009), 321–332.

[169] Olaya Galan, N., Ulloa Rubiano, J., Velez Reyes, F., Fernandez Duarte, K., Salas Cardenas, S., and Gutierrez Fernandez, M. In vitro antiviral activity of lactobacillus casei and bifidobacterium adolescentis against rotavirus infection monitored by nsp 4 protein production. *Journal of applied microbiology 120*, 4 (2016), 1041–1051.

[170] O'Sullivan, A. C., and Perez, S. The relationship between internal chain length of amylopectin and crystallinity in starch. 381–390.

[171] Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., et al. Interpro in 2022. *Nucleic Acids Research 51*, D1 (2023), D418–D427.

[172] Pérez, S., and Bertoft, E. The molecular structures of starch components and their contribution to the architecture of starch granules: A comprehensive review. *Starch-Stärke 62*, 8 (2010), 389–420.

[173] Photenhauer, A. L., Cerqueira, F. M., Villafuerte-Vega, R., Armbruster, K. M., Mareček, F., Chen, T., Wawrzak, Z., Hopkins, J. B., Kooi, C. W. V., Janeček, Š., et al. The ruminococcus bromii amylosome protein sas6 binds single and double helical α-glucan structures in starch. *bioRxiv* (2022), 2022–11.

[174] Piiadov, V., Ares de Araujo, E., Oliveira Neto, M., Craievich, A. F., and Polikarpov, I. Saxsmow 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental saxs data measured on a relative scale. *Protein Sci 28*, 2 (2019), 454–463.

[175] Pollet, R. M., Martin, L. M., and Koropatkin, N. M. Tonb-dependent transporters in the bacteroidetes: Unique domain structures and potential functions. *Molecular Microbiology 115*, 3 (2021), 490–501.

[176] Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. cryosparc: algorithms for rapid unsupervised cryo-em structure determination. *Nature methods 14*, 3 (2017), 290–296.

[177] Rambo, R. P., and Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature 496*, 7446 (2013), 477–481.

[178] Ravi, A., Troncoso-Rey, P., Ahn-Jarvis, J., Corbin, K. R., Harris, S., Harris, H., Aydin, A., Kay, G. L., Le Viet, T., Gilroy, R., Pallen, M. J., Page, A. J., O'Grady, J., and Warren, F. J. Hybrid metagenome assemblies link carbohydrate structure with function in the human gut microbiome. *Commun Biol 5*, 1 (2022), 932.

[179] Rees, D. A., and Welsh, E. J. Secondary and tertiary structure of polysaccharides in solutions and gels. 214–224.

[180] Reeves, A. R., D'Elia, J. N., Frias, J., and Salyers, A. A. A bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *Journal of bacteriology 178*, 3 (1996), 823–830.

[181] Reeves, A. R., Wang, G., and Salyers, A. A. Characterization of four outer membrane proteins that play a role in utilization of starch by bacteroides thetaiotaomicron. *Journal of bacteriology 179*, 3 (1997), 643–649.

[182] Reyman, M., van Houten, M. A., van Baarle, D., Bosch, A. A., Man, W. H., Chu, M. L. J., Arp, K., Watson, R. L., Sanders, E. A., Fuentes, S., et al. Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. *Nature communications 10*, 1 (2019), 4997.

[183] Rios-Covian, D., Gueimonde, M., Duncan, S. H., Flint, H. J., and de los Reyes-Gavilan, C. G. Enhanced butyrate formation by cross-feeding between faecalibacterium prausnitzii and bifidobacterium adolescentis. *FEMS Microbiol Lett 362*, 21 (2015).

[184] Rodriguez Sanoja, R., Morlon-Guyot, J., Jore, J., Pintado, J., Juge, N., and Guyot, J. Comparative characterization of complete and truncated forms of lactobacillus amylovorus α-amylase and role of the c-terminal direct repeats in raw-starch binding. *Applied and Environmental Microbiology 66*, 8 (2000), 3350–3356.

[185] Rodriguez-Sanoja, R., Oviedo, N., Escalante, L., Ruiz, B., and Sanchez, S. A single residue mutation abolishes attachment of the cbm26 starch-binding domain from lactobacillus amylovorus alpha-amylase. *J Ind Microbiol Biotechnol 36*, 3 (2009), 341–6.

[186] Rohou, A., and Grigorieff, N. Ctffind4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology 192*, 2 (2015), 216–221.

[187] Romano, A., Mackie, A., Farina, F., Aponte, M., Sarghini, F., and Masi, P. Characterisation, in vitro digestibility and expected glycemic index of commercial starches as uncooked ingredients. *J Food Sci Technol 53*, 12 (2016), 4126–4134.

[188] Rosenblum, J. L., Irwin, C. L., and Alpers, D. H. Starch and glucose oligosaccharides protect salivary-type amylase activity at acid ph. *American Journal of Physiology-Gastrointestinal and Liver Physiology 254*, 5 (1988), G775–G780.

[189] Russell, D., Ross, R., Fitzgerald, G., and Stanton, C. Metabolic activities and probiotic potential of bifidobacteria. *International journal of food microbiology 149*, 1 (2011), 88–105.

[190] Salminen, S., Isolauri, E., and Onnela, T. Gut flora in normal and disordered states. *Chemotherapy 41(suppl 1)*, Suppl. 1 (1995), 5–15.

[191] Sanders, L. M. *Carbohydrate: Digestion, Absorption and Metabolism*. Academic Press, Oxford, 2016, pp. 643–650.

[192] Santos, J., Saus, E., Smalley, S., Cataldo, L., Alberti, G., Parada, J., Gratacós, M., and Estivill, X. Copy Number Polymorphism of the Salivary Amylase Gene: Implications in Human Nutrition Research. *Journal of Nutrigenetics and Nutrigenomics 5*, 3 (09 2012), 117–131.

[193] Sarian, F. D., van der Kaaij, R. M., Kralj, S., Wijbenga, D.-J., Binnema, D. J., van der Maarel, M. J., and Dijkhuizen, L. Enzymatic degradation of granular potato starch by microbacterium aurum strain b8. a. *Applied microbiology and biotechnology 93* (2012), 645–654.

[194] Sayers, E. W., Cavanaugh, M., Clark, K., Pruitt, K. D., Schoch, C. L., Sherry, S. T., and Karsch-Mizrachi, I. Genbank. *Nucleic Acids Res 49*, D1 (2021), D92–D96.

[195] Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. Nih image to imagej: 25 years of image analysis. *Nature Methods 9*, 7 (2012), 671–675.

[196] Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., and Sali, A. Foxs, foxsdock and multifoxs: Single-state and multi-state structural modeling of proteins and their complexes based on saxs profiles. *Nucleic Acids Res 44*, W1 (2016), W424–9.

[197] Sidar, A., Albuquerque, E. D., Voshol, G. P., Ram, A. F., Vijgenboom, E., and Punt, P. J. Carbohydrate binding modules: diversity of domain architecture in amylases and cellulases from filamentous microorganisms. *Frontiers in Bioengineering and Biotechnology 8* (2020), 871.

[198] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol 7* (2011), 539.

[199] Singh, R. S., Saini, G. K., and Kennedy, J. F. Pullulan: Microbial sources, production and applications. *Carbohydr Polym 73*, 4 (2008), 515–31.

[200] Smith, S. P., and Bayer, E. A. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Current Opinion in Structural Biology 23*, 5 (2013), 686–694.

[201] Sonestedt, E. Salivary amylase gene variations influence the physiologic response to starchy foods: 2 sides of the story. *The American Journal of Clinical Nutrition 108*, 4 (2018), 656–657.

[202] Soper, M. T., DeToma, A. S., Hyung, S.-J., Lim, M. H., and Ruotolo, B. T. Amyloid-β–neuropeptide interactions assessed by ion mobility-mass spectrometry. *Physical Chemistry Chemical Physics 15*, 23 (2013), 8952–8961.

[203] Spiwok, V. Ch/π interactions in carbohydrate recognition. *Molecules (Basel, Switzerland) 22*, 7 (Jun 2017), 1038.

[204] Stam, M. R., Danchin, E. G., Rancurel, C., Coutinho, P. M., and Henrissat, B. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. *Protein Engineering, Design and Selection 19*, 12 (2006), 555–562.

[205] Stern, J., Moraïs, S., Ben-David, Y., Salama, R., Shamshoum, M., Lamed, R., Shoham, Y., Bayer, E. A., and Mizrahi, I. Assembly of synthetic functional cellulosomal structures onto the cell surface of lactobacillus plantarum, a potent member of the gut microbiome. *Applied and Environmental Microbiology 84*, 8 (2018), e00282–18.

[206] Strynadka, N. C. J., and James, M. N. G. Towards an understanding of the effects of calcium on protein structure and function. *Current Opinion in Structural Biology 1*, 6 (1991), 905–914.

[207] Sumitani, J., Tottori, T., Kawaguchi, T., and Arai, M. New type of starch-binding domain: the direct repeat motif in the c-terminal region of bacillus sp. no. 195 alpha-amylase contributes to starch binding and raw starch degrading. *The Biochemical Journal 350 Pt 2*, Pt 2 (Sep 2000), 477–484.

[208] Svergun, D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of applied crystallography 25*, 4 (1992), 495–503.

[209] Tahri, K., Crociani, J., Ballongue, J., and Schneider, F. Effects of three strains of bifidobacteria on cholesterol. *Letters in Applied Microbiology 21*, 3 (1995), 149–151.

[210] Tancula, E., Feldhaus, M., Bedzyk, L., and Salyers, A. Location and characterization of genes involved in binding of starch to the surface of bacteroides thetaiotaomicron. *Journal of Bacteriology 174*, 17 (1992), 5609–5616.

[211] Tannous, S., Stellbrinck, T., Hoter, A., and Naim, H. Y. Interaction between the α-glucosidases, sucrase-isomaltase and maltase-glucoamylase, in human intestinal brush border membranes and its potential impact on disaccharide digestion. *Frontiers in Molecular Biosciences 10* (2023).

[212] Teichmann, J., and Cockburn, D. W. In vitro fermentation reveals changes in butyrate production dependent on resistant starch source and microbiome composition. *Frontiers in Microbiology 12* (2021), 640253.

[213] Terrapon, N., Lombard, V., Drula, E., Lapébie, P., Al-Masaudi, S., Gilbert, H. J., and Henrissat, B. Puldb: the expanded database of polysaccharide utilization loci. *Nucleic acids research 46*, D1 (2018), D677–D683.

[214] Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L.-W. Decision-making in structure solution using bayesian estimates of map quality: the phenix autosol wizard. *Acta Crystallographica Section D 65*, 6 (2009), 582–601.

[215] Thomson, P., Medina, D. A., and Garrido, D. Human milk oligosaccharides and infant gut bifidobacteria: Molecular strategies for their utilization. *Food Microbiology 75* (2018), 37–46.

[216] Tian, Y., Wang, Y., Zhong, Y., Moller, M. S., Westh, P., Svensson, B., and Blennow, A. Interfacial catalysis during amylolytic degradation of starch granules: Current understanding and kinetic approaches. *Molecules 28*, 9 (2023).

[217] Turnbull, W. B., and Daranas, A. H. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc 125*, 48 (2003), 14859–66.

[218] Turroni, F., Serafini, F., Foroni, E., Duranti, S., O'Connell Motherway, M., Taverniti, V., Mangifesta, M., Milani, C., Viappiani, A., Roversi, T., and B., S. Role of sortase-dependent pili of bifidobacterium bifidum prl2010 in modulating bacterium-host interactions. *Proceedings of the National Academy of Sciences 110*, 27 (2013), 11151–11156.

[219] Tuson, H. H., Foley, M. H., Koropatkin, N. M., and Biteen, J. S. The starch utilization system assembles around stationary starch-binding proteins. *Biophysical Journal 115*, 2 (2018), 242–250.

[220] UniProt, C. Uniprot: the universal protein knowledgebase in 2021. *Nucleic Acids Res 49*, D1 (2021), D480–D489.

[221] Valk, V., Eeuwema, W., Sarian, F. D., van der Kaaij, R. M., and Dijkhuizen, L. 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*, 19 (2015), 6610–20.

[222] Valk, V., Kaaij, R. M. v. d., and Dijkhuizen, L. The evolutionary origin and possible functional roles of fniii domains in two microbacterium aurum b8.a granular starch degrading enzymes, and in other carbohydrate acting enzymes. *Amylase 1*, 1 (2017), 1–11.

[223] Valk, V., Lammerts van Bueren, A., van der Kaaij, R. M., and Dijkhuizen, L. Carbohydrate-binding module 74 is a novel starch-binding domain associated with large and multidomain alpha-amylase enzymes. *FEBS J 283*, 12 (2016), 2354–68.

[224] van de Waterbeemd, M., Fort, K. L., Boll, D., Reinhardt-Szyba, M., Routh, A., Makarov, A., and Heck, A. J. R. High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. *Nature Methods 14*, 3 (2017), 283–286.

[225] Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research 50*, D1 (2022), D439–D444.

[226] Venkataraman, A., Sieber, J. R., Schmidt, A. W., Waldron, C., Theis, K. R., and Schmidt, T. M. Variable responses of human microbiomes to dietary supplementation with resistant starch. *Microbiome 4*, 1 (2016), 33.

[227] Vihinen, M., and Mantsiila, P. Microbial amylolytic enzyme. *Critical reviews in biochemistry and molecular biology 24*, 4 (1989), 329–418.

[228] Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. Data processing and analysis with the autoproc toolbox. *Acta Crystallographica Section D 67*, 4 (2011), 293–302.

[229] Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., et al. Sphire-cryolo is a fast and accurate fully automated particle picker for cryo-em. *Communications biology 2*, 1 (2019), 218.

[230] Wang, S., and Copeland, L. Effect of acid hydrolysis on starch structure and functionality: A review. *Critical Reviews in Food Science and Nutrition 55*, 8 (2015), 1081–1097. PMID: 24915341.

[231] Wang, W., Hu, H., Zijlstra, R. T., Zheng, J., and Ganzle, M. G. Metagenomic reconstructions of gut microbial metabolism in weanling pigs. *Microbiome 7*, 1 (2019), 48.

[232] Wang, W., Kitova, E. N., and Klassen, J. S. Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem 75*, 19 (2003), 4945–55.

[233] Whelan, S., and Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol 18*, 5 (2001), 691–9.

[234] Wong, J. M., De Souza, R., Kendall, C. W., Emam, A., and Jenkins, D. J. Colonic health: fermentation and short chain fatty acids. *Journal of clinical gastroenterology 40*, 3 (2006), 235–243.

[235] Xiao, J., Kondo, S., Takahashi, N., Miyaji, K., Oshida, K., Hiramatsu, A., Iwatsuki, K., Kokubo, S., and Hosono, A. Effects of milk products fermented by bifidobacterium longum on blood lipids in rats and healthy adult male volunteers. *Journal of dairy science 86*, 7 (2003), 2452–2461.

[236] Xiao, S., Jiang, S., Qian, D., and Duan, J. Modulation of microbially derived short-chain fatty acids on intestinal homeostasis, metabolism, and neuropsychiatric disorder. *Applied microbiology and biotechnology 104* (2020), 589–601.

[237] Xiao, Y., Zhao, J., Zhang, H., Zhai, Q., and Chen, W. Colonized niche, evolution and function signatures of bifidobacterium pseudolongum within bifidobacterial genus. *Foods 10*, 10 (2021).

[238] Xu, J., Xu, R., Jia, M., Su, Y., and Zhu, W. Metatranscriptomic analysis of colonic microbiota's functional response to different dietary fibers in growing pigs. *Animal Microbiome 3*, 1 (2021), 45.

[239] Zaman, S. A., and Sarbini, S. R. The potential of resistant starch as a prebiotic. *Critical Reviews in Biotechnology 36*, 3 (2016), 578–584.

[240] Ze, X., Ben David, Y., Laverde-Gomez, J. A., Dassa, B., Sheridan, P. O., Duncan, S. H., Louis, P., Henrissat, B., Juge, N., Koropatkin, N. M., Bayer, E. A., and Flint, H. J. Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic firmicutes bacterium ruminococcus bromii. *MBio 6*, 5 (2015), e01058–15.

[241] Ze, X., Duncan, S. H., Louis, P., and Flint, H. J. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J 6*, 8 (2012), 1535–43.

[242] Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., and Yin, Y. dbcan2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research 46*, W1 (2018), W95–W101.

[243] Zheng, H., Chordia, M. D., Cooper, D. R., Chruszcz, M., Muller, P., Sheldrick, G. M., and Minor, W. Validation of metal-binding sites in macromolecular structures with the checkmymetal web server. *Nat Protoc 9*, 1 (2014), 156–70.

[244] Zheng, H., Cooper, D. R., Porebski, P. J., Shabalin, I. G., Handing, K. B., and Minor, W. Checkmymetal: a macromolecular metal-binding validation tool. *Acta Crystallogr D Struct Biol 73*, Pt 3 (2017), 223–233.

[245] Zheng, J., Ge, Q., Yan, Y., Zhang, X., Huang, L., and Yin, Y. dbcan3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Research* (2023).

[246] Zheng, T., Eswaran, S., Photenhauer, A. L., Merchant, J. L., Chey, W. D., and D'Amato, M. Reduced efficacy of low fodmaps diet in patients with ibs-d carrying sucrase-isomaltase (si) hypomorphic variants. *Gut 69*, 2 (2020), 397–398.

[247] Zhou, B., Yuan, Y., Zhang, S., Guo, C., Li, X., Li, G., Xiong, W., and Zeng, Z. Intestinal flora and disease mutually shape the regional immune system in the intestinal tract. *Frontiers in Immunology 11* (2020), 575.

[248] Zivkovic, A. M., German, J. B., Lebrilla, C. B., and Mills, D. A. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. *Proceedings of the National Academy of Sciences 108*, supplement_1 (2011), 4653–4658.