Chapter 9
Cellulosomes: Highly Efficient Cellulolytic Complexes

Victor D. Alves, Carlos M. G. A. Fontes, and Pedro Bule

Abstract Cellulosomes are elaborate multienzyme complexes capable of efficiently deconstructing lignocellulosic substrates, produced by cellulolytic anaerobic microorganisms, colonizing a large variety of ecological niches. These macromolecular structures have a modular architecture and are composed of two main elements: the cohesin-bearing scaffoldins, which are non-catalytic structural proteins, and the various dockerin-bearing enzymes that tenaciously bind to the scaffoldins. Cellulosome assembly is mediated by strong and highly specific interactions between the cohesin modules, present in the scaffoldins, and the dockerin modules, present in the catalytic units. Cellulosomal architecture and composition varies between species and can even change within the same organism. These differences seem to be largely influenced by external factors, including the nature of the available carbon-source. Even though cellulosome producing organisms are relatively few, the development of new genomic and proteomic technologies has allowed the identification of cellulosomal components in many archea, bacteria and even some primitive eukaryotes. This reflects the importance of this cellulolytic strategy and suggests that cohesin-dockerin interactions could be involved in other non-cellulolytic processes. Due to their building-block nature and highly cellulolytic capabilities, cellulosomes hold many potential biotechnological applications, such as the conversion of lignocellulosic biomass in the production of biofuels or the development of affinity based technologies.

Keywords Cellulosome · Modular · CAZymes · Cohesin-Dockerin · Scaffoldin · Biofuels

V. D. Alves · C. M. G. A. Fontes · P. Bule (✉)
CIISA, Faculdade de Medicina Veterinária, ULisboa, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisbon, Portugal
e-mail: pedrobule@fmv.ulisboa.pt

V. D. Alves
e-mail: vdalves@fmv.ulisboa.pt

C. M. G. A. Fontes
e-mail: cafontes@fmv.ulisboa.pt

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_9

The Cellulosome—Introduction

Plant cell wall polysaccharides are one of the most abundant sources of carbon and energy on Earth. They mainly consist of cellulose fibers, which are usually associated with hemicellulose, lignin (a non-saccharide polymer), pectins and proteins. The interactions between the various plant cell wall components form an intricate structure that is highly resistant to degradation, making all the stored energy largely inaccessible (Albersheim et al. 2011; Carpita et al. 2015). Because of this complexity, plant cell wall polysaccharides require a large number of highly specialized enzymes to catalyse their degradation. These enzymes are included in a group of biocatalysts capable of degrading, modifying or creating glycosidic bonds, termed Carbohydrate Active Enzymes or CAZymes (Henrissat and Davies 2000; Lombard et al. 2010). In light of its potential as a clean alternative source of renewable energy, the conversion of lignocellulosic biomass into fermentable sugars has fuelled a significant amount of dedicated research, leading to the identification and characterization of an ever expanding number of CAZymes (Jørgensen et al. 2007; Cantarel et al. 2009; Lombard et al. 2014). Most of these enzymes are produced by cellulolytic microorganisms, including bacteria, fungi and protozoa, which are able to hydrolyse the lignocellulose in plant cell walls, recycling the photosynthetically fixed carbon and therefore playing a crucial role in the carbon cycle (Himmel and Bayer 2009; Himmel et al. 2010). In general, aerobic cellulolytic species secret copious amounts of enzymes which act individually during structural polysaccharide hydrolysis. The released products are then captured by the cellulolytic cell and used as an energy source. However, the energetic constraints posed by anaerobic ecosystems limit the enzyme producing capabilities of microorganisms. This has imposed an evolutionary pressure on anaerobic bacteria and fungi leading to the development of an alternative and more efficient cellulolytic system, termed the cellulosome (Bayer et al. 1994; Fontes and Gilbert 2010).

The cellulosome was first described in the early 1980's, in the gram-positive thermophilic bacteria *Hungateiclostridium thermocellum* (*Clostridium thermocellum*) (Bayer et al. 1983; Lamed et al. 1983). This macromolecular structure consists of a remarkably efficient and well-organized cell surface multi-enzyme complex that allows enzyme recycling and direct assimilation of polysaccharide hydrolysis products. Cellulosomes have several different components, but are essentially made of two types of building blocks: dockerin (Doc) bearing enzymes and cohesin (Coh) bearing structural proteins, named scaffoldins (Bayer et al. 2004; Fontes and Gilbert 2010; Artzi et al. 2017). The high affinity and high specificity protein-protein interactions established between cohesin and dockerin modules define the molecular platform for cellulosome assembly. These two components bind tightly with each other according to a specificity pattern, dictating final cellulosomal architecture and composition (Fig. 9.1) (Carvalho et al. 2003; Doi and Kosugi 2004; Hamberg et al. 2014).

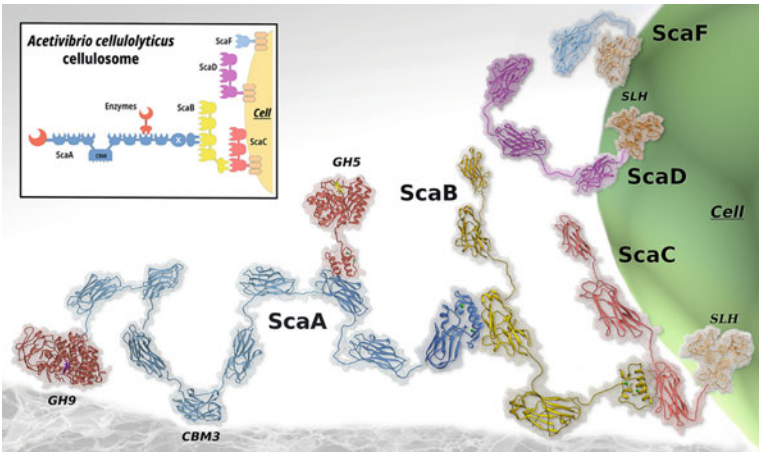

Fig. 9.1 Acetivibrio cellulolyticum cellulosome. A structure-based concept art illustration. The cellulosome architecture is organized around the non-catalytic primary scaffoldin (ScaA), a unique adaptor scaffoldin (ScaB) and several anchoring scaffoldins (ScaC, ScaD and ScaF). ScaA bears 9 cohesin (Coh) modules that congregate the catalytic enzymes *via* their appended dockerin (Doc) modules, it also can bind the carbohydrate substrate through a carbohydrate binding module (CBM) and it is crowned by a N-terminal glycoside hydrolase (GH) and by a C-terminal X-Doc that enables the anchoring of the cellulosome to the bacterial cell-wall. The schematic cellulosome in the small image insert is color-coded according to the main 3D image representation. The quaternary structure as well as the unstructured linkers between modules, are an "artistic" rendering speculation of a cellulosome. The structure artwork was generated in UCSF Chimera v 1.15 *Adapted* from (Bule et al. 2018b)

In this chapter, we dissect the cellulosome in order to explore the functional and structural properties of each of its key components. We also focus on cellulosomal diversity by discussing some of the known cellulosome-producing organisms, including the architecture and composition of their complexes. Finally, we look at the regulation of cellulosomal gene transcription and discuss some of the potential biotechnological applications for these macromolecular structures.

Cellulosomal Components

Cellulosomes result from the assembly of several distinct proteins, usually displaying a modular organization, with each module having a specific function within the macromolecular cellulolytic complex. Those functions include cellulosome assembly, structural support, catalytic activity, substrate binding and cellular tethering. The assembly elements are what distinguish the cellulosomal system from the free enzyme systems. They include the cohesin modules, present in purely structural non-catalytic proteins, termed scaffoldins, and the dockerin modules, present

in the enzymes and some scaffoldins. The strong and specific protein-protein interactions occurring between cohesin and dockerin modules allow the several cellulosomal components to congregate in a highly-efficient cellulolytic nanomachine. Nonetheless, several other elements are also present in the cellulosome, contributing to its catalytic efficiency, flexibility and ability to recognize its substrates. Here, we break down the cellulosome into its fundamental pieces in order to provide a better understanding of their importance within the macromolecular complex.

The Scaffoldin

The backbone that creates the cellulosome is built around scaffoldins, which are structural proteins mainly composed of linker spaced multiple cohesin modules (Fig. 9.2). According to their function, scaffoldins are classified into four major types: primary, anchoring, adaptor and cell-free scaffoldins. It is around primary scaffoldins that cellulosomes are assembled, by simultaneously incorporating up to hundreds of CAZymes through high affinity Coh-Doc interactions (Artzi et al. 2017; Smith et al. 2017). Cohesins and dockerins are classified into types (I, II, III, R) based on primary structure homology (Bayer et al. 2004). Apart from a few exceptions, type I interactions are usually committed to cellulosome assembly, while type II interactions are associated with cell wall tethering. Therefore, most known primary scaffoldins possess a C-terminal type II dockerin, which interacts with the complementary type II cohesin modules in cell-bound anchoring scaffoldins. Many scaffoldins have an Ig-like stabilizing module, termed the X-module, adjacent to the N-terminal end of the type II dockerin (Adams et al. 2006). Another common feature in primary scaffoldins is the presence of a non-catalytic domain with carbohydrate-binding activity. These modules, named Carbohydrate-Binding Modules (CBMs), target the cellulosome to the crystalline regions of the plant cell wall. Interestingly, scaffoldin CBMs are almost exclusively family 3 CBMs, as it will be further discussed below (Tormo et al. 1996; Shimon et al. 2000; Hong et al. 2014).

Complex cellulosomes have dedicated scaffoldins responsible for mediating the attachment of the cellulosome to the bacterial cell surface, called anchoring scaffoldins (Fig. 9.2). The importance of this cell-bound arrangement *vis-a-vis* the alternative cell-free cellulosome is reflected by the clustering of cellulosomal genes, with primary and anchoring scaffoldin genes usually clustered together (Zhivin et al. 2017). Anchoring scaffoldins are non-covalently bound to the organism surface through an SLH domain, or covalently bound through a sortase recognition motif (Lemaire et al. 1995; Rincon et al. 2005). Another type of scaffoldin, the adaptor scaffoldin, has also been described on more elaborate and highly structured cellulosomes (Fig. 9.2). Polyvalent adaptor scaffoldins that have multiple cohesin modules, can serve as an amplifier for the number of CAZymes integrated into the cellulosome (Xu et al. 2003; Dassa et al. 2012). Alternatively, monovalent adaptor scaffoldins use a single cohesin module to intermediate binding between enzymes bearing divergent

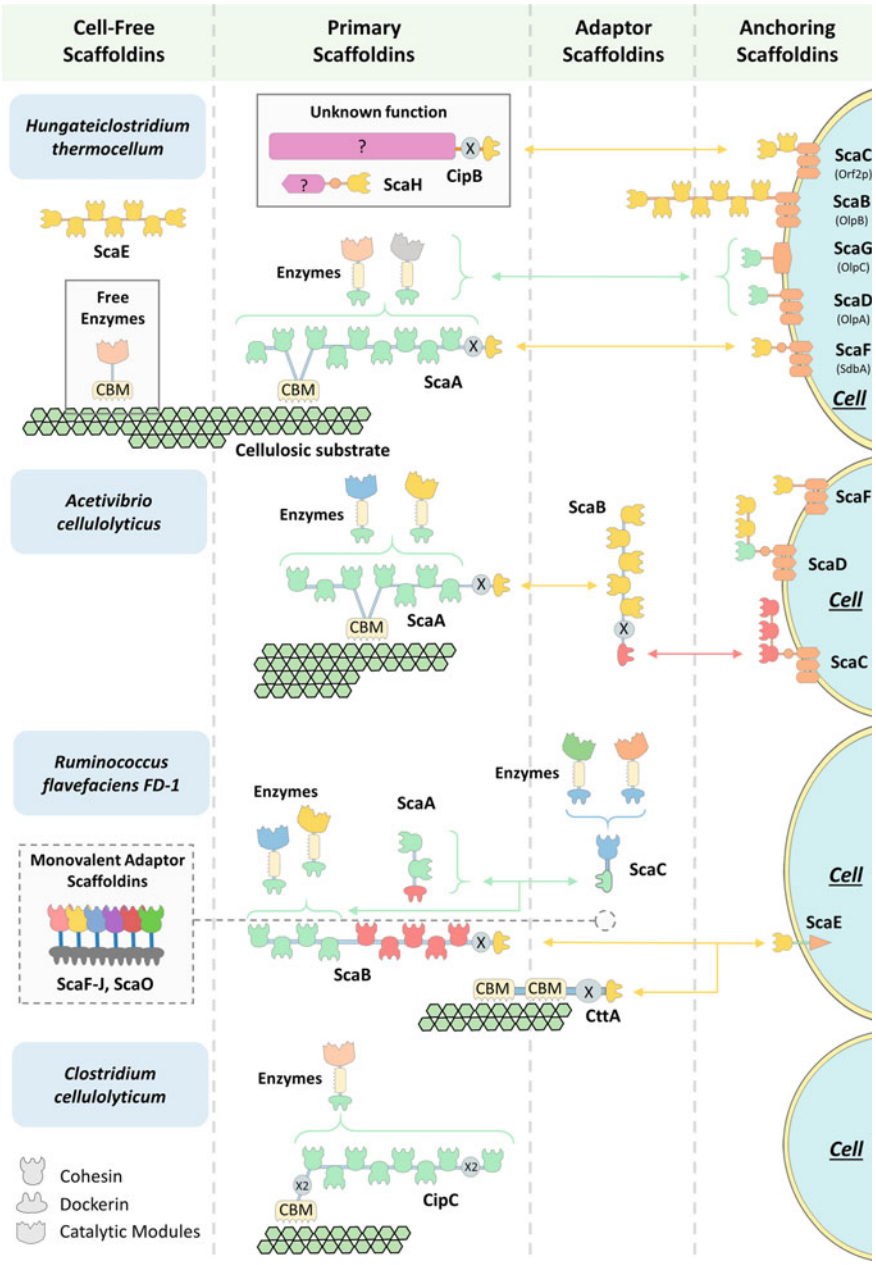

Fig. 9.2 Architecture of cellulosome systems. Schematic representation of four distinct cellulosomal systems belonging to the bacteria *H. thermocellum*, *A. cellulolyticus*, *R. flavefaciens* and *C. cellulolyticus*. Within each system, the same color is used for dockerins sharing similar specificity and for the cohesins recognized by those dockerins. X-modules are represented by gray circles marked with an X. Polyssacharides are represented as green hexagons

dockerins and the primary scaffoldin (Fig. 9.2). They can also function as a repertoire change inducer, by switching the specificity of the CAZymes that populate the primary scaffoldin in response to substrate availability (Rincon et al. 2004; Moraïs et al. 2016a). Cell-free scaffoldins are exclusively composed of cohesin modules, separated by linker sections (Fig. 9.2). They originate untethered cellulosomes that are released to the extracellular space, like the ones secreted by *Clostridium clarifavum* or *Hungateiclostridium thermocellum*, which can coexist with cell-bound cellulosomes (Artzi et al. 2015; Xu et al. 2016).

Some noteworthy exceptions from this general scaffoldin blueprint are the absence of a C-terminal dockerin in the primary scaffoldins of simple cellulosomes (Dassa et al. 2017) and the reversed cohesin-dockerin arrangement of *Bacteroides (Pseudobacteroides) cellulosolvens* cellulosome system. In this species, the assembly of enzymes onto the primary scaffoldin relies on type II cohesins, while the terminal type I dockerin mediates cellulosome attachment to an anchoring scaffoldin (Zhivin et al. 2017).

Cohesins and Dockerins

Cohesin and dockerin modules are undoubtedly the most important components of the cellulosome, such that the basis for the identification of cellulosomal proteins is dependent upon the presence of one or both of these modular elements. These complementary modules are generally classified based on sequence similarity. In the archetypal cellulosome of *Hungateiclostridium thermocellum*, type I interactions are reserved for CAZYme recruitment onto the primary scaffoldin, while type II interactions allow tethering the cellulosome to the bacterial cell surface. Although other species present similar arragements, this dichotomy was broken upon the characterization of the highly complex cellulosome of *Ruminococcus flavefaciens* (Berg Miller et al. 2009), which possesses divergent cohesins and dockerins, subsequently classified as type III modules. More recently, a comprehensive bioinformatics analysis of *Bacteroides (Pseudobacteroides) cellulosolvens*'s genome, led to the discovery of yet another type of interaction, named type R, in what is currently the most complex cellulosome system known to date (Zhivin et al. 2017). As more data are generated from genomic and metagenomics studies, it seems that the type classification for cohesins and dockerins might be more diverse than initially suspected. Metagenomic data mining offers new opportunities when dealing with uncultured bacteria and such analysis have already uncovered some divergent unclassified cohesins that do not fall in any of the known types (Bensoussan et al. 2017). Structural homology type classification versus a functional classification, although convenient, can be misleading, and a number of noteworthy discrepancies imply a cautious approach, namely the already mentioned type I and II function reversal of *Bacteroides (Pseudobacteroides) cellulosolvens* (Xu et al. 2004a; Bule et al. 2018b).

Cohesin-dockerin interactions are one of the strongest non-covalent protein-protein interactions known (Stahl et al. 2012) and the basis for cellulosomal assembly.

They tend to be species-specific, although cross-specificity (predominantly for type II interactions) has been previously described (Pagès et al. 1997; Haimovitz et al. 2008; Fontes and Gilbert 2010; Bule et al. 2018b). Both cohesins and dockerins display a striking overall conservation of structural topology. Cohesins are approximately 150 amino acids long, arranged in a flattened and elongated β-sandwich, composed of nine β-strands divided between two β-sheets (A and B), arranged in a jellyroll topology, and bearing a hydrophobic core. β-sheet A is composed of β-strands 4, 7, 2, 1 and 9, while β-sheet B is composed of β-strands 5, 6, 3 and 8 (Fig. 9.3a, b). A distinctive and noticeable planar plateau is formed by β-sheet B which establishes the dockerin binding interface of cohesins (Carvalho et al. 2003; Adams et al. 2006). Structurally, dockerins are short protein modules (~ 70 residues long) with an α-helix dominated configuration, arranged around two highly conserved duplicate sequences (N-terminal or helix-1 and C-terminal or helix-3) presenting nearly perfect 2-fold symmetry, each composed of about 22 amino acid that, alternatively, comprise the cohesin binding sites of the dockerin. These two tandem repeats are connected by a smaller third α-helix (helix-2), and hold a stable tertiary configuration based upon two conserved F-hand Ca^{2+}-binding loop-helix structural motifs (Fig. 9.3). An almost fully conserved Gly residue on the N-term helix-1 constitutes a null reference (#0) for a convenient relative dockerin residue position numbering. As a consequence, residues at positions #1, 3, 5, 9, and 12 (and the equivalent positions on the C-term helix-2) are typically aspartate or asparagine residues and are involved in calcium-coordination, whereas positions #10 and 11 are recognized as main residues involved in cohesin recognition, along with positions #17, 18 and 22. A prevailing residue conservation on both helix-1 and -3 cohesin recognition sites and the discovery that some dockerin can bind to their cohesin partners by either an N- or C-terminal helix dominated interaction, disclosed the existence of a dual-binding mode for type-I interactions (Fig. 9.3) (Carvalho et al. 2003). This means that some dockerins can bind to the same partner in two distinct orientations, which are related by a 180º rotation relatively to the cohesin. On the other hand, the first structurally characterized type II cohesin-dockerin complex revealed a single-binding mode of interaction. In this complex, the lack of residue conservation between the two dockerin repeats only allows binding in a single orientation, with no dominating helix, since residues on both helices are significantly involved in the interaction. Interestingly, in type II complexes, the adjacent X-module also participates in cohesin recognition (Adams et al. 2006; Brás et al. 2012; Salama-Alber et al. 2013).

Initially, it was believed that the dual-binding mode would be ubiquitous for cellulosome CAZyme recruitment, contributing to avert steric clashes among cellulosome components while facilitating substrate access, by providing two alternative binding conformations. In contrast, the single-binding mode would be the norm for cellulosome cell-anchoring. However, it is now clear that in some organisms, like *Acetivibrio cellulolyticus*, there is a widespread dual-binding mode interaction in both cellulosome assembly and anchoring, even when it involves adaptor scaffoldins (Cameron et al. 2015; Brás et al. 2016; Bule et al. 2018a). Furthermore, ruminococal cellulosomes, like that of *Ruminococcus flavefacien*s, can be assembled and anchored exclusively through single binding mode cohesin-dockerin interactions (Bule et al.

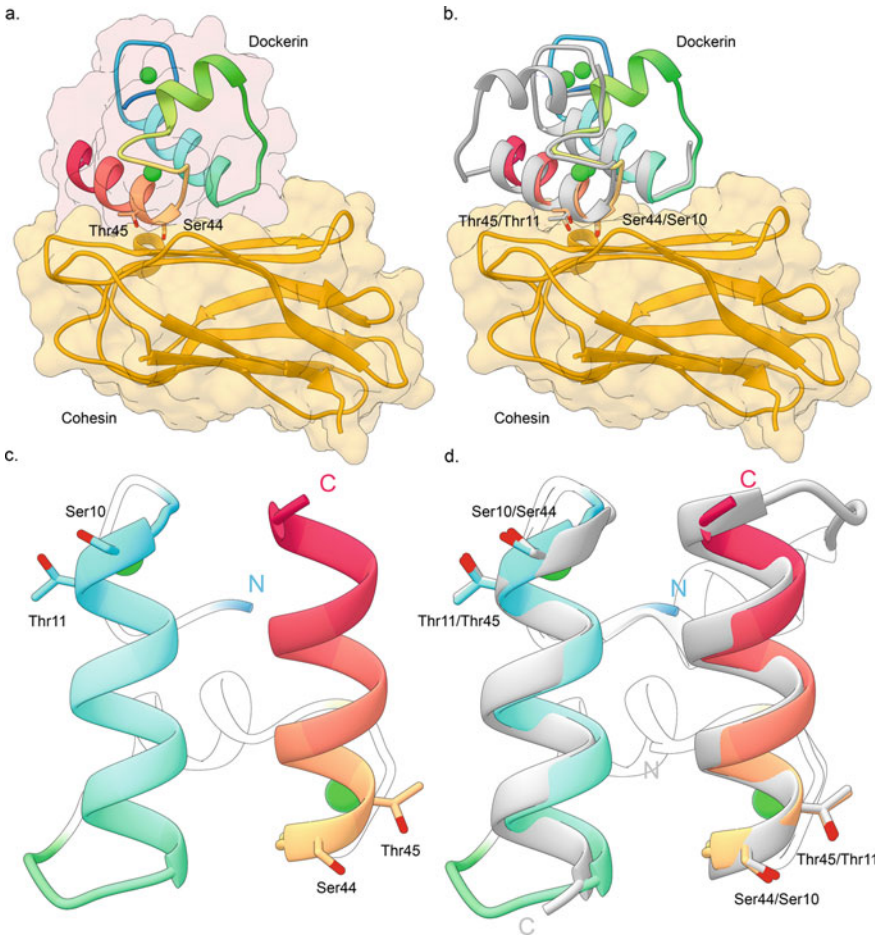

Fig. 9.3 The dual binding mode. **a**. Ribbon representation of a type I cohesin-dockerin complex from *Hungateiclostridium thermocellum* (PDB accession code 1OHZ), containing the dockerin from a xylanase (in rainbow coloring: N-terminus is blue, C-terminus is red) and the second cohesin from primary scaffoldin ScaA (in yellow). In this panel it is possible to observe that the dockerin is interacting with the cohesin mainly through its C-terminal helix. **b**. Overlay of 1OHZ with the same Coh-Doc complex (PDB accession code 2CCL) in which the dockerin (in grey) has been mutated to force interaction with the cohesin through its N-terminal helix. The overlay was obtained by matching the cohesins of both complexes. It is possible to observe that the dockerin of 2CCL is rotated 180° relative to the dockerin in 1OHZ. **c**. Cohesin binding interface of the type I dockerin present in the 1OHZ Coh-Doc complex (N-terminus in blue, C-terminus in red). **d**. Comparison of the two cohesin binding helices of the type I Doc, by overlaying the dockerin with a version of itself rotated by 180° (in grey). This comparison shows that there is conservation of the two key cohesin recognition residues (Serine/Threonine pair) at the same relative positions (#10, #11) in both helices. Van der Waals' surfaces are represented in translucent colors and calcium ions are shown in green. The structures in this figure were generated using UCSF Chimera v 1.13.1

2016, 2017). Interestingly, the lack of conservation between the two symmetric dockerin halves has, in some cases, resulted in a dual-specificity-binding strategy, in which each of the helices at the binding interface recognizes a different cohesin. This means that the dockerin can bind distinct partners by using opposite orientations, revealing the enormous versatility of such interactions (Voronov-Goldman et al. 2015; Brás et al. 2016). Ultimately, such diversity, although based on a very conserved main chain Cα trace, relies on the particular composition of key interface recognition amino acid residues, defining its conserved or divergent character, and subtly modifying inter- and intra-species barriers, binding specificity and binding mode for the cohesin-dockerin interaction (Slutzki et al. 2015; Bule et al. 2016, 2017; Brás et al. 2016).

Enzymes

Cellulosomal enzymes are, by definition, enzymes containing a dockerin domain. These enzymes are typically modular and, in addition to the dockerin that mediates protein-protein interactions, possess one or more catalytic modules and sometimes one or more CBMs, involved in protein-carbohydrate interactions (Fontes and Gilbert 2010). As a plant cell wall degrading macromolecular complex, the cellulosome can integrate numerous different carbohydrate active enzymes. The first identified cellulosomal enzymes were cellulases, which is not surprising considering that the first cellulosome was discovered due to its highly efficient cellulolytic properties (Lamed et al. 1983). In addition to cellulases, other cellulosomal enzymes were subsequently identified, most notably xylanases, pectinases, carbohydrate esterases, mannanases and xyloglucanases.

Much like their substrates, complex polysaccharide degrading enzymes are extremely diverse and tend to be highly specific. They include glycoside hydrolases, carbohydrate esterases and polysaccharide lyases and have been grouped into families according to primary sequence homology, kept in an ever-growing online database called CAZy.org (Henrissat and Davies 1997, 2000; Cantarel et al. 2009; Lombard et al. 2010, 2014). The degradation of complex polysaccharides often requires the synergistic action of several different CAZymes. Synergy may be due to different modes of action towards the same substrate, like when an endoglucanase hydrolyses the substrate, thereby producing additional chain ends that can be cleaved by an exoglucanase. It can also result from the hydrolysis of two tightly associated substrates, in which the action of one enzyme could make the concealed substrate accessible for the action of the second enzyme (e.g. cellulases and xylanases). It can also work by limiting product inhibition effects, when a second enzyme acts on the product of a primary enzyme, thus restoring the activity of the first enzyme (e.g. β-glucosidases and cellulases) (Morag et al. 1991). Even the hydrolysis of a relatively simple polysaccharide such as cellulose involves the cooperative efforts of at least three groups of enzymes: endo-(1,4)-β-D-glucanases, exo-(1,4)-β-D-glucanases and β-glucosidases. The endoglucanase randomly attacks the internal O-glycosidic

bonds, resulting in glucan chains of different lengths; the exoglucanase acts on the ends of the cellulose chains and releases β-cellobiose as the major end product and finally, the β-glucosidases act specifically on the β-cellobiose disaccharides releasing glucose (Kuhad et al. 2011).

The genome sequencing of several cellulosome producing bacteria has identified a large variety of dockerin bearing enzymes, which include not only cellulases but also other CAZymes such as hemicellulases, ligninases (Fan and Yuan 2010; Cragg et al. 2015), pectinases (Tamaru and Doi 2001; Chakraborty et al. 2015), mannanases (Kurokawa et al. 2001; Sabathé et al. 2002; Perret et al. 2004) and chitinases, suggesting a complex mechanism of enzymatic synergy. Nonetheless, almost all known cellulosome-producing bacteria characteristically express large amounts of a single family 48 glycoside hydrolase (GH48) exoglucanase, which is a key component for the highly efficient cellulolytic activity displayed by these structures (Morag et al. 1991; Ravachol et al. 2015). Contrastingly, an extensive repertoire of family 9 glycoside hydrolases is generally secreted by these bacteria, which work in close synergy with GH48 since, on its own, GH48 displays very little activity on cellulose. Many GH9 enzymes from different cellulosomic systems have been characterized, displaying different activities, distinct abilities to bind to cellulosic substrates and diverse synergies with the major GH48 exoglucanase (Ravachol et al. 2014; Moraïs et al. 2016b). This eventually suggests that enzyme diversity, especially of GH9 enzymes, reflects the structural diversity of the substrates. In fact, the nature of the available substrate seems to be an important factor in cellulosome composition and efficiency, as revealed by some proteomic studies (Gold and Martin 2007; Cho et al. 2010; Tsai et al. 2010; Wang et al. 2019). The expression levels of enzymatic components can change depending on the carbon source, which means that cellulosomes with different compositions can be assembled by the same microorganism when grown in the presence of different substrates. Other glycoside hydrolase families, such as GH5, GH10, GH11 and GH43, are also commonly found in cellulosome systems, meaning that these complexes are able to utilize a large variety of plant cell wall polysaccharides (Artzi et al. 2017).

Interestingly, dockerin bearing proteins are not restricted to CAZymes. Many other activities have been reported for potential cellulosomic proteins, which are not directly linked to polyssacharide degradation. These include serpins (Kang et al. 2006), proteases (Levy-Assaraf et al. 2013) and expansins (Chen et al. 2016; Artzi et al. 2017). The role of these biocatalysts in cellulosome function is still unclear although it has been suggested that they may have protective and regulatory functions or even act in synergy with CAZymes to degrade highly complex substrates (Chen et al. 2016; Moraïs et al. 2016b; Bensoussan et al. 2017).

Carbohydrate Binding Modules

In general, multi-modular CAZymes are appended to non-catalytic Carbohydrate Binding Modules (CBMs) that bind carbohydrates, thus playing a critical role in

targeting their associated enzymes towards the substrate, resulting in improved catalytic activity (Boraston et al. 2004; Gilbert et al. 2013). Since the first described CBMs had crystalline cellulose as their primary ligand, these non-catalytic polysaccharide recognizing modules were initially defined as Cellulose-Binding Domains, CBDs (Gilkes et al. 1988). After these initial studies, it has been shown that many carbohydrates can be targeted by different 'CBDs', including crystalline cellulose, non-crystalline cellulose, chitin, β-1,3-glucans, β-1,3-1,4-mixed linkage glucans, xylan, mannan, galactan and starch (Boraston et al. 2004). Some can even display 'lectin-like' specificity and bind to a variety of cell-surface glycans (Fujimoto et al. 2013). In order to reflect their diverse ligand specificity, these modules were renamed Carbohydrate Binding Modules, or CBMs (Boraston et al. 1999).

Much like the catalytic modules of CAZymes, CBMs are classified into families in the CAZy database, according to primary structure homology. Currently there are 86 different CBM families registered on cazy.org, a number that has been consistently growing (Boraston et al. 2004; Lombard et al. 2014). According to their function, CBMs have also been classified into one of three types: type A CBMs which interact with crystalline polysaccharides (primarily cellulose); type B modules which bind to internal regions of single glycan chains, and type C CBMs that recognize small saccharides in the context of a complex carbohydrate (Boraston et al. 2004; Fontes and Gilbert 2010; Armenta et al. 2017).

It is now clearly recognized that CBMs potentiate CAZyme function (Bolam et al. 1998). However, the mechanism by which they do so is still not fully understood. It was suggested that CBMs may function through a proximity effect or have a more active role leading to a disruptive effect on the polysaccharides (Boraston et al. 2004; Guillén et al. 2009). Many studies have shown that maintaining enzymes in the proximity of the insoluble substrate leads to a more efficient degradation, as the removal of CBMs from enzymes results in a dramatic reduction in enzymatic activity (Bolam et al. 1998; Boraston et al. 2003). Curiously, enzyme activity on soluble substrates is not frequently affected when CBMs are removed (Kleine and Liebl 2006), meaning that CBM function is highly dependent on the nature of the substrate.

CBMs can also be found in many cellulosomal systems, usually occupying a central position within the primary scaffoldin (Boraston et al. 2004; Fontes and Gilbert 2010). Generally, primary scaffoldins contain a family 3 CBM, such as the one present in ScaA of *H. thermocellum* (Poole 1992; Morag et al. 1995). This is a type A CBM that binds strongly to the crystalline surface of cellulose, accounting for the primary targeting of the cellulosome onto its substrate (Bayer et al. 2004; Artzi et al. 2017). In contrast to what has been described for many species such as *H. thermocellum*, *A. cellulolyticus* or *P. cellulosolvens*, the scaffoldins produced by ruminal bacterium *R. flavefaciens* do not possess a CBM module targeting cellulosic substrates. Rather, this bacterium expresses an independent 'CttA' scaffoldin with two putative CBMs and a dockerin that specifically recognizes an anchoring scaffoldin, which may serve to attach the bacterial cell directly to the substrate (Fig. 9.2) (Rincon et al. 2010; Salama-Alber et al. 2013). Interestingly, a cell-free scaffoldin produced by *A. cellulolyticus*, was found to contain two family 2 CBMs as opposed

to a family 3 CBM (Dassa et al. 2012). Comparable CBM2-containing cell-free scaffoldins were also observed in the genome of *Clostridium clariflavum* (Artzi et al. 2014)(Artzi et al. 2014). These are the only known examples of scaffoldins containing CBMs from a family other than CBM3. Family 2 CBMs are more commonly found associated with free, non-cellulosomal enzymes and the fact that they have only been found in cell-free scaffoldins suggests that they could have a distinct function in cellulosomes to that of family 3 CBMs (Simpson et al. 2000; Artzi et al. 2017).

CBMs seem to play a vital role in cellulosome function and not only as a part of the scaffoldin proteins. Although only two CBM families have been reported in scaffoldins, many cellulosomal enzymes have CBMs from various families, exhibiting different carbohydrate specificities. Presently, the study of cellulosomes has been an important source of novel carbohydrate binding molecules and has significantly helped our understanding of the mechanisms behind carbohydrate recognition (Venditto et al. 2016; Ribeiro et al. 2019). Much like free CAZymes, the enzyme-borne cellulosomal CBMs likely work to position the catalytic modules in the proximity of the substrate, resulting in optimal hydrolysis (Hammel et al. 2005).

Anchoring Modules

As mentioned above, most cellulosome producing species assemble different cellulosomes presenting distinct conformations and, depending on the nature of their main scaffoldin, cellulosomes can either be found in a cell-bound or cell-free form (Fig. 9.2). Nonetheless, in most species there appears to be a preference for the former. From an energy-efficiency point of view, one could infer the obvious advantages that a tethered complex presents to the anaerobic microorganism since (1) there is close proximity between the products of hydrolysis and the cell, resulting in optimal product uptake and (2) the ability to retain the complex at the cell surface allows the microorganism to benefit from its activity for long periods of time, as opposed to having to express multiple copies as they diffuse away. Furthermore, using ex vivo assembled mini-cellulosomes with identical catalytic components, it has been observed that the cell-bound versions display up to 4.5 times higher hydrolytic activity than the cell-free versions. This suggests that tethering the cellulosome could also be important for its potential biotechnological applications, as it optimizes function (You et al. 2012).

The surface layer, or S-layer, is a common feature of many gram-positive bacteria. They are monomolecular crystalline arrays of proteinaceous subunits that surround the exterior cell wall, providing various functions, including protection and cellular interaction (Sleytr et al. 1993; Doi and Kosugi 2004). Since the S-layer was first described, much attention was given to finding how its proteins target and bind the cell wall. Curiously, the first clue came from the sequence of a gene cluster encoding some of *H. thermocellum* scaffoldins, including ScaA (then called CipA). The cluster sequence revealed an ORF encoding a cohesin module and a C-terminal domain displaying significant similarity to the N-terminal region of S-layer proteins

from various species. The authors suggested that this protein could be somehow involved in the cell-anchoring of *H. thermocellum*'s cellulosome (Fujino et al. 1993) and later confirmed its ability to bind the cell envelope (Lemaire et al. 1995). Since then, the now called SLH (S-layer homology) domain has been identified in many bacterial surface proteins. It is composed of either one or three repeated GPTS (glycine, proline, threonine, serine)-rich segments that bind peptidoglycan associated polysaccharides in a non-covalent interaction (Mesnage 2000), tethering extracellular proteins to the bacterial wall, including many cellulosome anchoring scaffoldins (Xu et al. 2004b; Zhivin et al. 2017). On the other hand, most mesophilic cellulosomes, such as the ones produced by *R. flavefaciens* and *Ruminococcus champanelensis*, do not possess SLH domains in their anchoring scaffoldins. Instead, they are covalently bound to the cell wall via a sortase-mediated attachment mechanism (Rincon et al. 2005; Salama-Alber et al. 2013). This mechanism involves three crucial features, present at the C-terminus of the protein, and collectively called the cell wall sorting signal: an LPXTG (leucine, proline, X, threonine and glycine, where X is any amino acid) motif, a hydrophobic domain which interacts with the cytoplasmic membrane and a tail of charged residues (Schneewind et al. 1992; Mazmanian 1999). During secretion, the hydrophobic domain and charged tail prevents protein secretion by stopping membrane translocation, allowing recognition of the LPXTG motif by the sortase enzyme. The sortase then cleaves the LPXTG motif between the threonine and glycine residues and covalently attaches the threonine to the amino group of the pentaglycine cell wall cross-bridge resulting in a cell wall attached protein (Paterson and Mitchell 2004).

The X-Module

A common feature of primary scaffoldins is the presence of an enigmatic module at the N-terminus of their dockerins, which has been termed the X module (Fig. 9.2). This X module-dockerin pairing (XDoc) has been described in many cellulosomes, such as the ones assembled by *H. thermocellum*, *A. cellulolyticus, P. cellulosolvens* and *R. flavefaciens.* The exact function of this domain is still relatively unknown, although it has been repeatedly suggested that its presence could provide structural stability and enhanced solubility to the adjacent cellulosomal components (Mosbah et al. 2000; Schubot et al. 2004; Xu et al. 2004a; Kataeva et al. 2004; Adams et al. 2005). The crystal structure of a *H. thermocellum* type II cohesin-dockerin complex together with its neighbouring X-module shows direct hydrogen-bond contacts between the Ig-like X module and the cohesin. Together with the fact that the X module stabilizes the type II dockerin in solution, this suggests that there is a functional enhancement of adjacent components in the presence of the X module (Adams et al. 2006). The authors also report an increased affinity of the type II Coh for the XDoc modular pair *versus* that for the isolated type II Doc, supporting this assumption.

Unlike the X-module in the type II Coh-XDoc interaction of *H. thermocellum*, the X-module in the type III Coh-XDoc complex of *R. flavefaciens* does not appear to contribute directly to the CohE-Doc binding surface (Fig. 9.2). Rather, its elongated stalk-like conformation, supported by 3 dockerin inserts that act as structural buttresses, appears to serve as an extended spacer, projecting the cellulosome away from the bacterial cell wall (Salama-Alber et al. 2013). This would position the catalytic domains away from the cell surface facilitating the action of the cellulosomal enzymes on the plant cell polysaccharides. In fact, the *R. flavefaciens* X module does not bear homology with any other X modules from cellulolytic bacteria. Instead, it shares significant similarity with the G5 module of *Streptococcus pneumoniae's StrH* protein, whose function is to position the catalytic modules away from the cell, allowing optimal processing of host cell surface-presented glycoconjugates.

Cellulosome Diversity

Since the cellulosome was first described, it has become increasingly apparent that its prevalence in nature is not as pronounced as initially predicted. As far as our knowledge extends, this system is rather restricted to a few specialized microorganisms inhabiting specific ecological niches such as the soil, sewage sludge and some mammalian gastrointestinal compartments. Nonetheless, there is no denying that cellulosomes play an essential role in the conversion of recalcitrant lignocellulosic substrates in anaerobic environments, and with the expanding availability of genomic, metagenomic and proteomic tools, it is very likely that new cellulosome producing species will continue to emerge.

Depending on whether cellulosomes possess a single scaffoldin or multiple types of interacting scaffoldins, cellulosome systems are divided into two major types: the simple and the complex. As mentioned, scaffoldins are the molecular platforms that organize cellulosomes and have the potential to be exceptionally varied in size and constitution, leading to cellulosome systems that are structurally diverse. Cellulosomes can be rather simple and small, like those of *Clostridium acetobutylicum*, which can only assemble 5 enzymes in its single scaffoldin (Sabathé et al. 2002). They can also be large and complex such as those of the *P. cellulosolvens* system, with its 31 different scaffoldins and a theoretical assembly that can hold up to 110 enzymes in one single cellulosome unit (Zhivin et al. 2017). As opposed to complex cellulosomes, which are found in species living in a wide range of environmental conditions, simple cellulosomes have only been identified in mesophilic organisms. These bacteria produce cellulosomes with a single scaffoldin backbone that can attach up to 9 cohesins, depending on the species, and are usually not tethered to the bacterial wall. These include *Clostridium josui* (Ichikawa et al. 2014), *Clostridium cellulovorans* (Cho et al. 2010), *Clostridium papyrosolvens* (Pohlschröder et al. 1994), *Clostridium cellulolyticum* (Fig. 9.2) (Xu et al. 2015) and *Ruminococcus bromii* (Ze et al. 2015). This last species is a key resistant-starch degrader found in the human gut (Ze et al.

2012) and, as such, its dockerin-bearing enzymes were found to hydrolyze starch, rather than cellulose. Hence, *R. bromii* complexes were referred to as amylosomes.

The first described cellulosome and currently best-characterized system, is that of the thermophilic bacterium *H. thermocellum* (Lamed et al. 1983). Interestingly, *H. thermocellum*'s highly efficient cellulolytic performance is supported by three systems: a free enzyme system; a cell-free cellulosome system and a cell-bound cellulosome system (Xu et al. 2016), which has been extensively studied and still serves as the archetypal model to this day (Fig. 9.2). *H. thermocellum*'s genome encodes 72 dockerin-containing proteins, suggesting an elaborate cellulosome system (Pinheiro et al. 2009). This bacterium can produce eight different scaffoldins including primary scaffoldin ScaA, the five cell-anchoring scaffoldins ScaB, C, D, F and G, the two cell-free scaffoldins ScaE and H and the elusive CipB. Primary scaffoldin ScaA has eight type I cohesins, separated by linkers of various lengths, which can bind to the dockerins present on cellulosomal enzymes (Gerwig et al. 1993; Kruus et al. 1995). It also possesses an internal family 3 CBM and a C-terminal type II X-dockerin that can bind to the cohesins on anchoring scaffoldins ScaB, C or F, tethering the cellulosome to the bacterial surface, or to the cohesins on ScaE, originating cell-free cellulosomes (Xu et al. 2016). Considering that ScaB and ScaE have seven type II cohesins each, it is possible for *H. thermocellum* to assemble both cell-free and cell-bound cellulosomes with up to 63 enzymes on a single unit. Enzymes can also be directly attached to the bacterial surface by interacting with the cohesins on ScaG and ScaD which, like the other anchoring scaffoldins on this bacterium, are bound to the cell wall by their C-terminal SLH domains (Leibovitz and Béguin 1996. The roles of CipB and ScaH are still not fully understood.

The concept of adaptor scaffoldin was first introduced by the cellulosome of *Acetivibrio cellulolyticus* (Fig. 9.2) which, with 16 different scaffoldins and over 140 dockerin-bearing proteins encoded in its genome, has a considerably more intricate system than that of *H. thermocellum* (Xu et al. 2003; Dassa et al. 2012). These adaptor scaffoldins promote the incorporation of more enzymes into one single complex allowing *A. cellulolyticus* to integrate up to 84 enzymes in its cellulosome, which, including ScaA's own catalytic module, totalizes 96 catalytic components. The thermophilic *Clostridium clariflavum* has a similar system to that of *A. cellulolyticus* in the sense that it also relies on adaptor scaffoldins to create larger assemblies, like its theoretical, record breaking, 160-enzyme cellulosome (Artzi et al. 2014). Interestingly, in the cellulosomal system of the ruminal bacterium *R. flavefaciens* (Fig. 9.2), monovalent adaptor scaffoldins possessing a single cohesin have also been identified. These adaptors, rather than increasing the number of enzymes, increase the repertoire of different catalytic units that can integrate the complex (Rincón et al. 2004; Bule et al. 2016). The genome of *R. flavefaciens* encodes an enormous number of cellulosomal components, including 222 dockerin-containing proteins, the largest number described so far (Rincon et al. 2010). These dockerins have been classified into six different families according to their primary structure homology, a classification that translates into function, as members of the same family recognize similar cohesin partners (Israeli-Ruimy et al. 2017; Bule et al. 2017, 2018c). The monovalent adaptor scaffoldins are capable of recruiting specific families into the

cellulosome, integrating different catalytic activities in the complex's profile (Bule et al. 2016). Intriguingly, numerous dockerin bearing proteins were also identified in the *Ruminococcus albus* genome, another cellulolytic ruminal bacterium, but only a single cohesin was found (Dassa et al. 2014). Given that *R. albus* shares the same ecological niche as *R. flavefaciens*, it has been proposed that these bacteria have a symbiotic relationship, in which the enzymes produced by *R. albus* can be integrated into *R. flavefaciens* cellulosome, with both benefiting from its catalytic action. This hypothesis is supported by the fact that the dockerins from the two species share a high degree of homology, which in some cases can be as high as 60% identity. The recently described cellulosome system of *Ruminococcus champanelensis*, a cellulolytic bacteria found in the human gut, was also found to possess a very elaborate architecture, similar to that of *R. flavefaciens* (Cann et al. 2016; Moraïs et al. 2016a).

The existence of fungal multi-enzyme complexes has also been proposed from as early as 1992 (Wilson and Wood 1992). Although it has been suggested that these complexes are assembled through cohesin-dockerin interactions (Fanutti et al. 1995), there is still very little information regarding fungal cellulosomes when compared to their bacterial counterparts. Fungal dockerins have been identified in numerous cellulolytic enzymes but differ significantly from bacterial dockerins (Haitjema et al. 2017). They can be found either at the N or C-terminus of enzymes and exist in tandem repeats rather than single copies at the C-terminal region (Steenbakkers et al. 2001). This tandem repetition seems to have a functional significance as it has been observed that a double-dockerin can bind more tightly to the cellulosome than a single domain one (Nagy et al. 2007). The only reported 3-dimensional structures of a fungal dockerin were obtained by NMR and do not display the typical double F-hand calcium binding fold seen in all bacterial dockerins, but rather a single β-sheet composed of three or four small antiparallel strands and a short α-helix (Raghothama et al. 2001; Nagy et al. 2007). Furthermore, there seems to exist pronounced interspecies recognition, with dockerins from one fungal organism binding the cellulosome of others (Nagy et al. 2007). Until recently, there was no convincing evidence supporting the existence of fungal cohesin domains, but a comparative genomics study allowed the definitive identification of scaffoldin proteins in the anaerobic fungi *Anaeromyces robustus*, *Neocallimastix californiae* and *Piromyces finnis* (Haitjema et al. 2017). These scaffoldin proteins were shown to bind non-catalytic dockerin domains present in enzymes and contained multiple repeating motifs, which were suggested to be the cohesin domains. The minimum sequence that defines a single fungal cohesin remains to be determined but scaffoldin fragments with only four of these repeats were capable of dockerin recognition. Clearly, there is still much left to know about fungal cellulosomes but they could potentially have distinct advantages over the bacterial ones such as a larger variety of catalytic activities and the fact that they produce glucose and not cellobiose as final product (Gilmore et al. 2015). It is becoming increasingly evident that fungi have a key role in the conversion of recalcitrant biomass, much due to their cellulosomal enzymes (Hagen et al. 2020). This makes fungal cellulosomes attractive blueprints for the engineering of complexes with multiple biotechnological applications (Gilmore et al. 2020).

Recent advances in genome-sequencing methods and new metagenomic approaches are promoting the rapid discovery of new cellulosome systems and expanding our understanding of well-known cellulosomal species (Dassa et al. 2012, 2014; Artzi et al. 2014; Zhivin et al. 2017). Cohesins and dockerins are being described in the genome of uncultured bacteria within species not previously documented (Bensoussan et al. 2017). New types of cohesins that do not fit within the previously described three types were also found, suggesting that these modules are more diverse than originally predicted. In addition, cohesins and dockerins have been identified in the genomes of non-cellulolytic species such as *Clostridium perfringens*, an opportunistic pathogen that inhabits the gastrointestinal tract (Bayer et al. 1999; Adams et al. 2008; Peer et al. 2009). Some of these dockerin-bearing proteins were found to be toxins or to form multi-protein complexes that promote virulence. This could mean that the cohesin-dockerin interaction can be present in other systems besides the cellulosome, possibly serving other purposes other than polysaccharide degradation.

Quaternary Structure

Even though there is now abundant information regarding the 3-dimensional structures of each of the several cellulosomal components, very little is known about the conformation of an entire complex. After the first successful attempt at determining the structure of a cohesin-dockerin complex (Carvalho et al. 2003), some were inspired to crystallize larger portions of cellulosomes, albeit with relatively little success (Currie et al. 2012). The intrinsic plasticity of linker segments connecting the various cohesin modules, allow scaffoldins to adopt a variety of dramatically different conformations, making cellulosome quaternary architecture extremely hard to decipher. This conformational flexibility between cohesin modules allow for the optimal positioning of the enzymatic subunits onto the substrate (Hammel et al. 2005). In fact, data comparing the activity of designer cellulosomes containing different linker lengths between cohesins suggest that longer linkers improve hydrolytic efficiency (Vazana et al. 2013). Interestingly, the linker section connecting the catalytic cores of cellulosomal enzymes with their dockerins seem to be predominantly rigid (Hammel et al. 2004, 2005).

The largest crystal structure of a cellulosome ever to be determined by X-ray crystallography belongs to the canonical cellulosome of *H. thermocellum* and includes five modules from three different proteins (Currie et al. 2012). It contains the type I Doc of Cel9D bound to a C-terminal trimodular fragment of ScaA scaffoldin, which includes a type I cohesin connected by a linker to the X module-type II dockerin dyad, which in turn is interacting with ScaF's type II cohesin (Fig. 9.4). The structure reveals an elongated topology with a flexible 13-residue linker connecting the ninth type I cohesin (CohI9) module and the X module. Elevated temperature factors suggest that the linker is highly dynamic. In addition, the four copies of the penta-modular complex found in the asymmetric unit of the crystal structure had

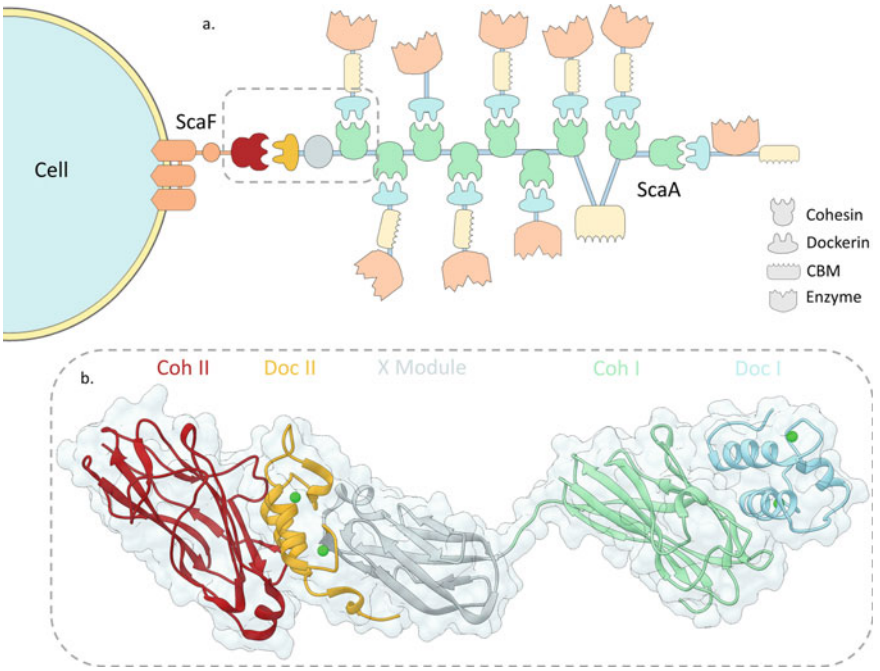

Fig. 9.4 Tridimensional structure of the largest crystallized cellulosome fragment. **a**. Cellulosome of *Hungateiclostridium thermocellum* resulting from the interaction of anchoring scaffoldin ScaF and primary scaffoldin ScaA. The modules within the dashed box correspond to the modules seen in the structure below. **b**. Ribbon representation of a ternary cellulosome complex structure from *Hungateiclostridium thermocellum,* containing a total of five modules (PDB accession code 4FL4). The type II cohesin from anchoring scaffoldin ScaF is represented in red and is interacting with the type II dockerin from primary scaffoldin ScaA, represented in yellow. The ScaA fragment also shows the dockerin adjacent X-module (in grey) separated from the 9th type II cohesin (in green) by a linker region. Lastly, a type I dockerin from the enzyme Cel9D can be seen in light blue interacting with ScaA's 9th cohesin. The van der Waals' surface of the complex is shown in translucent blue. Calcium ions are represented in green. The structure in this figure was generated using UCSF Chimera v 1.13.1

slightly different orientations in the DocI-CohI9 region. This suggests the existence of several possible conformations of the linker sequences that contribute to modulate the overall conformation of the cellulosome.

Regulation of Cellulosomal Components

The genes encoding most cellulosomal subunits are organized in clusters on the bacterial chromosome. On simpler systems, the scaffoldin gene is followed downstream by the dockerin-bearing enzyme encoding genes, while complex systems have

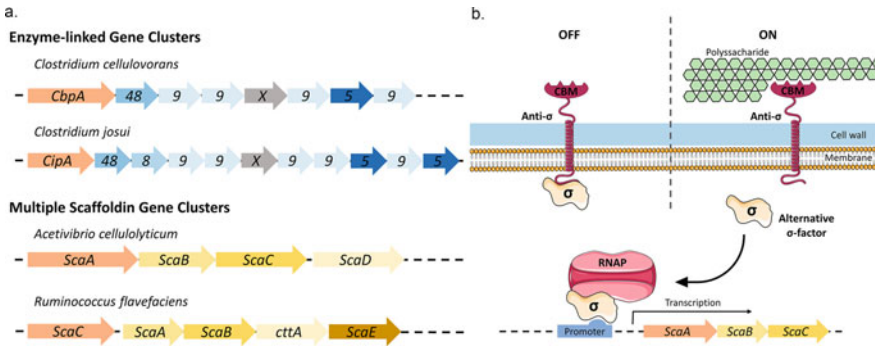

Fig. 9.5 Transcription regulation of cellulosomal genes. **a**. Examples of gene clusters present in the genome of two simple cellulosome producing bacteria and two complex cellulosome producing bacteria. **b**. Transcription regulation via a substrate sensitive anti sigma factor and alternative sigma factor mechanism. The CBM-like extracellular domain of the anti-sigma factor (in purple) recognizes the appropriate polyssacharide (green hexagons) and changes conformation releasing the alternative sigma factor (**σ**) inside the bacterial cell. The alternative sigma factor binds to the RNA polymerase (RNAP) which will then be able to recognize the appropriate promoter and initiate transcription of the cellulosomal gene. This figure was made using Servier Medical ART (SMART)

multiple scaffoldin clusters (Fig. 9.5a). As seen in the supra mentioned examples, the repertoire of dockerin bearing enzymes encoded by each cellulosome producing system far exceeds the number of cohesins available, even on some of the largest assemblies. For example, *R. flavefaciens* FD-1 can gather up to 14 enzymes on a single cellulosome while its genome encodes over 220 dockerin containing proteins (Rincon et al. 2010). Even *B. cellulosolvens,* whose largest assembly can theoretically hold up to 110 enzymes, possesses over 200 dockerin-bearing ORFs in its genome (Zhivin et al. 2017). One of the few species whose cellulosome could eventually attach its full repertoire of dockerin-bearing proteins (78 in total) simultaneously on a single unit is *C. clariflavum*, with its theoretical 160 cohesin-bearing assembly (Artzi et al. 2014). This suggests that cellulosome composition is not preestablished and is probably influenced by external factors. Growth rate seems to influence the expression of cellulosomal genes, with high growth rates resulting in down-regulation of the components (Dror et al. 2003a). More importantly though, is that the expression of these genes can change according to the nature of the available carbon source, resulting in a diversity of possible cellulosomal configurations within the same system (Dror et al. 2003b; Han et al. 2003; Raman et al. 2009; Artzi et al. 2015). It was previously observed that the expression levels of cellulases, such as the highly prevalent GH48 exoglucanase, are increased when the bacteria are grown on cellulose compared to when cultivated on cellobiose, while the opposite is true for other glycoside hydrolases. Enzymatic profiles also seem to differ in the presence of hemicellulosic polysaccharides when compared to those of cultures grown on crystalline cellulose (Dykstra et al. 2014; Artzi et al. 2015).

It is evident that a substrate sensing mechanism is operating within cellulosome producing organisms, in order to adjust the expression of cellulosomal components to

the available carbon sources. Some of the bacteria producing complex cellulosomes, such as *H. thermocellum*, *C. clariflavum*, *A. cellulolyticus* and *B. cellulosolvens*, seem to regulate their cluster genes by a pathway involving a set of transmembrane and RNA polymerase interacting proteins called anti-sigma factors and alternative sigma factors, respectively (Kahel-Raifer et al. 2010; Nataf et al. 2010; Muñoz-Gutiérrez et al. 2016). This is a common mechanism found in bacteria, in which the sigma factor (or alternative sigma factor) interacts with an RNA polymerase, promoting gene transcription, while the anti-sigma factor suppresses transcription by sequestering the sigma factor (Paget 2015). The ingenuity of the cellulosome regulating mechanism is in the extracellular portion of the anti-sigma factor involved, which has CBM-like domains capable of recognizing specific polysaccharides in the surrounding medium. When a transmembrane anti-sigma factor binds to its cognate substrate it changes conformation, releasing the alternative sigma factor in the intracellular space, which will in turn interact with the RNA polymerase, allowing promoter recognition and initiating transcription (Fig. 9.5b). This allows the bacteria to optimize its resources, by recognizing the polysaccharides in the extracellular medium and reacting accordingly, changing cellulosome composition and architecture as new substrates are exposed during the degradation process. A substrate independent transcriptional regulation mechanism has also been reported in *C. cellulolyticum*, which regulates cellulosomal stoichiometry by selective RNA processing and stabilization (Xu et al. 2015). This would allow the production of pre-optimized cellulosomes which could be crucial for correct cellulolytic function.

Applications for Cellulosomes

Due to its building-block nature, it is no surprise that the cellulosome has inspired numerous ingenious biotechnological solutions. These mainly explore the highly specific cohesin-dockerin interaction to promote reaction mechanisms whose components function in close proximity. The development of mini-cellulosomes and designer cellulosomes represent one of the most popular and successful studies in the field of cellulosomics (Bayer et al. 1994). The in vitro and in vivo construction of these cellulosome derivatives have been used both in the study of cellulosomal function and in trying to improve the efficiency of native cellulosomes, for various biotechnological purposes. The main difference between mini-cellulosomes and designer cellulosomes lies in the nature of their cohesins. The first are built on truncated versions of the native scaffoldins, containing cohesins similar to those found in the original complexes. Designer cellulosomes, on the other hand, have a chimeric scaffoldin composed of cohesins with distinct specificities that can originate from multiple species, allowing the incorporation of specific enzymes with control over position and number of copies (Fierobe et al. 2002, 2005; Moraïs et al. 2012). By designing tailor-made scaffoldins, several authors have been able to answer numerous questions regarding cellulosome assembly and function. These include the dynamics of enzyme incorporation into scaffoldins (Borne et al. 2013), the effects

of enzyme relative position (Stern et al. 2015), the importance of scaffoldin-borne CBMs for catalytic efficiency (Mingardon et al. 2007) and the importance of intermodular linkers for flexibility and steric accommodation (Caspi et al. 2009; Molinier et al. 2011; Vazana et al. 2013), among other findings.

One of the most extensively explored applications of designer cellulosomes has been its use in the conversion of lignocellulosic biomass into fermentable sugar, which is still a major bottleneck in biofuel production (Hasunuma et al. 2013). For this purpose, cellulosomes are used as a blueprint for the design, construction and exploitation of tailor-made catalytic multi-protein complexes integrating fungal and bacterial enzymes from non-aggregating systems, displaying particular promise in biomass saccharification. These enzymes include some not commonly found in cellulosomes, such as lytic polysaccharide monooxygenases (LPMOs), expansins and laccases, which have successfully helped to improve cellulosome efficiency with the introduction of novel complementary enzymatic activities (Gefen et al. 2012; Arfi et al. 2014; Chen et al. 2016; Davidi et al. 2016). Notably, there have been some efforts to obtain genetically engineered microorganisms, capable of producing and displaying mini-cellulosomes or designer cellulosomes on their cell surface. The objective is to create bacteria suitable for consolidated bioprocessing (CBP), meaning that they can be used to ferment plant cell walls into fuels in a one step process, using a single bioreactor, without the addition of enzymes (Hyeon et al. 2011; Moraïs et al. 2014; Ou and Cao 2014; Liang et al. 2014; Willson et al. 2016).

Other innovative applications have been exploring the properties of cellulosomal components, mainly taking advantage of the ability for CBMs to recognize polysaccharides and the high-affinity and specificity of cohesin-dockerin interactions. These include microarray technologies (Haimovitz et al. 2008; Israeli-Ruimy et al. 2017), affinity protein-purification systems (Demishtein et al. 2010) and molecular biosensors (Hyeon et al. 2014). To surpass the limitation posed by the near irreversible cohesin-dockerin interaction, a *H. thermocellum* dockerin with a lower affinity for its cohesin partner was successfully engineered, which enabled its use as an affinity tag for protein purification (Demishtein et al. 2010). The protein purification system consists of a *H. thermocellum* CBM fused to a cohesin that can be immobilized in a cellulose column and is capable of recognizing the dockerin tag in a recombinant protein. The protein is then eluted with EDTA, which disrupts the cohesin-dockerin interaction. Besides proving a very efficient and robust alternative system for affinity chromatography, the affinity tag was also shown to have little effect on the properties of the proteins tested, including enzymes. Furthermore, the relatively inexpensive costs of cellulose-based affinity columns together with their reusable nature and high capacity make this system very attractive for affinity protein purification (Karpol et al. 2009; Demishtein et al. 2010). Another curious cohesin-dockerin based technology consists of a high sensitivity and selectivity self-assembling biosensor that could potentially be used to detect multiple specific molecules in a complex mixture, such as blood serum (Hyeon et al. 2014).

It has been shown that the proximity effect promoted by cellulosomes can also be transferred to other platforms. By drawing inspiration from cellulosomal assembly, other cellulolytic macromolecular structures have been designed. These include

self-assembling circular multi-enzyme complexes (Mitsuzawa et al. 2009), cellulase coated nanospheres (Blanchette et al. 2012), streptavidin clustered cellulases (Kim et al. 2012) and cellulases that are bound to DNA scaffolds (Mori et al. 2013). Nevertheless, none of these platforms seem to possess the same degree of structural flexibility or substrate targeting capabilities which accounts for much of the cellulosome's efficiency.

The utilisation of enzymes and cellulosomes are also being considered as valuable alternatives for the usage of agro-wastes and organic pollutants as a renewable resource, reducing the consequent environmental pollution (Bayer et al. 2007; Karmakar and Ray 2011) and as feed additives to improve digestibility of animal diets (Costa et al. 2014). Even so, only a few of the many possible research applications have been explored and, as our knowledge about these multi-enzyme complexes increases, so does the potential for future innovation.

Concluding Remarks

There is no doubt that the study of cellulosomes has provided considerable insights on how recalcitrant polysaccharides are processed in nature, highlighting the pivotal role that anaerobic microorganisms colonizing lignocellulosic environments have in the carbon cycle. In addition, deciphering cellulosome function led to the discovery of many new biocatalysts and carbohydrate interacting molecules (Venditto et al. 2016) with many potential biotechnological applications, with the conversion of cellulosic biomass into biofuels being the most explored (Bayer et al. 2010; Arora et al. 2015; Kahn et al. 2019). The strong and highly-specific interaction between cohesin and dockerin modules has also broaden our knowledge about protein-protein interactions and provided innovative tools for the development of new affinity based technologies, with applications in research and healthcare (Haimovitz et al. 2008; Demishtein et al. 2010; Hyeon et al. 2014).

There is still much to be explored in the field of cellulosomics. Although cellulosome producing microorganisms are relatively rare, new systems will no doubt continue to emerge at increasingly faster rates. The massification of next generation sequencing technologies, metagenomic approaches, high-throughput methodologies and automation have streamlined otherwise very time-consuming protocols and are allowing the identification of cellulosomal components in a large variety of organisms, including some non-cellulolytic species (Adams et al. 2008; Peer et al. 2009; Dassa et al. 2014; Ben David et al. 2015; Haitjema et al. 2017). It would be extremely interesting to understand how wide cellulosome diversity can extend and what other functions could the cohesin-dockerin interaction potentially have in nature, beyond cellulosomal assembly. Regulation of cellulosomal component expression and particularly its dependence on substrate availability has been somewhat elucidated through transcriptomic and proteomic studies, but there is still a lack of information regarding their secretion and the assembly processes (Blouzard et al. 2010; Kahel-Raifer et al. 2010). Fungal cellulosomes are also still very poorly understood, especially regarding

the nature of the cohesin-dockerin interactions supporting their assembly. Clearly defined cohesin modules are yet to be described and whether glycosylation is important for dockerin binding remains to be established (Nagy et al. 2007). Given the extensive repertoire of dockerin-bearing fungal enzymes, understanding the fungal cellulosome system is an objective worth pursuing, as it could provide answers to many of the challenges found in biomass conversion (Haitjema et al. 2017).

One particularly challenging task that has been eluding researchers for many years is the determination of a high-resolution structure of a complete cellulosome. Such structure could shed light into the interactions between the several cellulosomal components and between the cellulosome and the substrate. It could also provide the answer to whether the dual-binding mode actually serves to better accommodate the enzymes in the complex, avoiding steric hindrance. Unfortunately, the highly flexible nature of cellulosomes and the heterogeneity in enzyme content make them impractical crystallization subjects. Cryo-electron microscopy could prove useful to that end, but for now only low-resolution small-angle X-ray scattering structures are available (Hammel et al. 2005; Currie et al. 2013).

The development of novel designer cellulosomes is one of the most prolific strategies in the field of cellulosomics. Besides providing several tools for the study of cellulosomes, many studies have been developing new complexes with improved biochemical properties destined for biotechnological purposes (Arora et al. 2015; Kahn et al. 2019; Gilmore et al. 2020), including the production of biofuels by CBP. It is to be expected that new cellulosome based technology will continue to emerge, in not only the field of biofuel production, but also in those where affinity-based approaches should prove useful. To that end, it is essential to further our knowledge about the mechanics behind these intricate systems.

References

Adams JJ, Webb BA, Spencer HL, Smith SP (2005) Structural characterization of type II dockerin module from the cellulosome of Clostridium thermocellum: calcium-induced effects on conformation and target recognition. Biochemistry 44:2173–2182. https://doi.org/10.1021/bi048039u

Adams JJ, Pal G, Jia Z, Smith SP (2006) Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. Proc Natl Acad Sci USA 103:305–310. https://doi.org/10.1073/pnas.0507109103

Adams JJ, Gregg K, Bayer EA et al (2008) Structural basis of Clostridium perfringens toxin complex formation. Proc Nat Acad Sci 105:12194–12199. https://doi.org/10.1073/pnas.0803154105

Albersheim P, Darvill A, Roberts K et al (2011) Plant cell walls: from chemistry to biology. Garland Science

Arfi Y, Shamshoum M, Rogachev I et al (2014) Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation. Proc Natl Acad Sci USA 111:9109–9114. https://doi.org/10.1073/pnas.1404148111

Armenta S, Moreno-Mendieta S, Sánchez-Cuapio Z et al (2017) Advances in molecular engineering of carbohydrate-binding modules: engineering of carbohydrate-binding modules. Proteins: Struct, Funct, Bioinf 85:1602–1617. https://doi.org/10.1002/prot.25327

Arora R, Behera S, Sharma NK, Kumar S (2015) Bioprospecting thermostable cellulosomes for efficient biofuel production from lignocellulosic biomass. Bioresources and Bioprocessing 2. https://doi.org/10.1186/s40643-015-0066-4

Artzi L, Dassa B, Borovok I et al (2014) Cellulosomics of the cellulolytic thermophile Clostridium clariflavum. Biotechnol Biofuels 7:100. https://doi.org/10.1186/1754-6834-7-100

Artzi L, Morag E, Barak Y et al (2015) Clostridium clariflavum: key cellulosome players are revealed by proteomic analysis. MBio 6:e00411–e00415. https://doi.org/10.1128/mBio.00411-15

Artzi L, Bayer EA, Moraïs S (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol 15:83–95. https://doi.org/10.1038/nrmicro.2016.164

Bayer EA, Kenig R, Lamed R (1983) Adherence of Clostridium thermocellum to cellulose. J Bacteriol 156:818–827

Bayer EA, Morag E, Lamed R (1994) The cellulosome–a treasure-trove for biotechnology. Trends Biotechnol 12:379–386. https://doi.org/10.1016/0167-7799(94)90039-6

Bayer EA, Coutinho PM, Henrissat B (1999) Cellulosome-like sequences in archaeoglobus fulgidus: an enigmatic vestige of cohesin and dockerin domains. FEBS Lett 463:277–280

Bayer EA, Belaich J-P, Shoham Y, Lamed R (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 58:521–554. https://doi.org/10.1146/annurev.micro.57.030502.091022

Bayer EA, Lamed R, Himmel ME (2007) The potential of cellulases and cellulosomes for cellulosic waste management. Curr Opin Biotechnol 18:237–245. https://doi.org/10.1016/j.copbio.2007.04.004

Bayer EA, Lamed R, White BA et al (2010) Conversion of agricultural residues to bioethanol: the roles of cellulases and cellulosomes. In: Blaschek HP, Ezeji TC, Scheffran J (eds) Biofuels from agricultural wastes and byproducts. Wiley-Blackwell, Oxford, UK, pp 67–96

Ben David Y, Dassa B, Borovok I et al (2015) Ruminococcal cellulosome systems from rumen to human. Environ Microbiol 17:3407–3426. https://doi.org/10.1111/1462-2920.12868

Bensoussan L, Moraïs S, Dassa B et al (2017) Broad phylogeny and functionality of cellulosomal components in the bovine rumen microbiome. Environ Microbiol 19:185–197. https://doi.org/10.1111/1462-2920.13561

Berg Miller ME, Antonopoulos DA, Rincon MT et al (2009) Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of ruminococcus flavefaciens FD-1. PLoS ONE 4:e6650. https://doi.org/10.1371/journal.pone.0006650

Blanchette C, Lacayo CI, Fischer NO et al (2012) Enhanced cellulose degradation using cellulase-nanosphere complexes. PLoS ONE 7:e42116. https://doi.org/10.1371/journal.pone.0042116

Blouzard J-C, Coutinho PM, Fierobe H-P et al (2010) Modulation of cellulosome composition in Clostridium cellulolyticum: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. Proteomics 10:541–554. https://doi.org/10.1002/pmic.200900311

Bolam DN, Ciruela A, McQueen-Mason S et al (1998) Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem J 331(Pt 3):775–781

Boraston AB, McLean BW, Kormos JM et al (1999) Carbohydrate-binding modules: diversity of structure and function. In: Gilbert HJ, Davies GJ, Henrissat B, Svensson B (eds) Recent advances in carbohydrate bioengineering. Royal Society of Chemistry, Cambridge, pp 202–211

Boraston AB, Kwan E, Chiu P et al (2003) Recognition and hydrolysis of noncrystalline cellulose. J Biol Chem 278:6120–6127. https://doi.org/10.1074/jbc.M209554200

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382:769–781. https://doi.org/10.1042/bj20040892

Borne R, Bayer EA, Pagès S et al (2013) Unraveling enzyme discrimination during cellulosome assembly independent of cohesin-dockerin affinity. FEBS J 280:5764–5779. https://doi.org/10.1111/febs.12497

Brás JLA, Alves VD, Carvalho AL et al (2012) Novel Clostridium thermocellum type I cohesin-dockerin complexes reveal a single binding mode. J Biol Chem 287:44394–44405. https://doi.org/10.1074/jbc.M112.407700

Brás JLA, Pinheiro BA, Cameron K et al (2016) Diverse specificity of cellulosome attachment to the bacterial cell surface. Sci Rep 6:38292. https://doi.org/10.1038/srep38292

Bule P, Alves VD, Leitão A, et al (2016) Single-binding mode integration of hemicellulose degrading enzymes via adaptor scaffoldins in Ruminococcus flavefaciens cellulosome. J Biol Chem jbc.M116.761643. https://doi.org/10.1074/jbc.M116.761643

Bule P, Alves VD, Israeli-Ruimy V et al (2017) Assembly of Ruminococcus flavefaciens cellulosome revealed by structures of two cohesin-dockerin complexes. Sci Rep 7:759. https://doi.org/10.1038/s41598-017-00919-w

Bule P, Cameron K, Prates JAM et al (2018a) Structure–function analyses generate novel specificities to assemble the components of multi-enzyme bacterial cellulosome complexes. J Biol Chem jbc.RA117.001241. https://doi.org/10.1074/jbc.RA117.001241

Bule P, Pires VM, Fontes CM, Alves VD (2018b) Cellulosome assembly: paradigms are meant to be broken! Curr Opin Struct Biol 49:154–161. https://doi.org/10.1016/j.sbi.2018.03.012

Bule P, Pires VMR, Alves VD, et al (2018c) Higher order scaffoldin assembly in Ruminococcus flavefaciens cellulosome is coordinated by a discrete cohesin-dockerin interaction. Scientific Reports 8:. https://doi.org/10.1038/s41598-018-25171-8

Cameron K, Najmudin S, Alves VD et al (2015) Cell-surface attachment of bacterial multienzyme complexes involves highly dynamic protein-protein anchors. J Biol Chem 290:13578–13590. https://doi.org/10.1074/jbc.M114.633339

Cann I, Bernardi RC, Mackie RI (2016) Cellulose degradation in the human gut: Ruminococcus champanellensis expands the cellulosome paradigm. Environ Microbiol 18:307–310. https://doi.org/10.1111/1462-2920.13152

Cantarel BL, Coutinho PM, Rancurel C et al (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37:D233–D238. https://doi.org/10.1093/nar/gkn663

Carpita N, Ralph J, McCann M (2015) Cell walls. In: Buchanan B, Gruissem W, Jones R (eds) Biochemistry & molecular biology of plants, 2nd edn. Wiley Blackwell, Oxford, UK, pp 45–110

Carvalho AL, Dias FMV, Prates JAM et al (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. Proc Natl Acad Sci USA 100:13809–13814. https://doi.org/10.1073/pnas.1936124100

Caspi J, Barak Y, Haimovitz R et al (2009) Effect of linker length and dockerin position on conversion of a thermobifida fusca endoglucanase to the cellulosomal mode. Appl Environ Microbiol 75:7335–7342. https://doi.org/10.1128/AEM.01241-09

Chakraborty S, Fernandes VO, Dias FMV et al (2015) Role of pectinolytic enzymes identified in Clostridium thermocellum cellulosome. PLoS ONE 10:e0116787. https://doi.org/10.1371/journal.pone.0116787

Chen C, Cui Z, Song X et al (2016) Integration of bacterial expansin-like proteins into cellulosome promotes the cellulose degradation. Appl Microbiol Biotechnol 100:2203–2212. https://doi.org/10.1007/s00253-015-7071-6

Cho W, Jeon SD, Shim HJ et al (2010) Cellulosomic profiling produced by Clostridium cellulovorans during growth on different carbon sources explored by the cohesin marker. J Biotechnol 145:233–239. https://doi.org/10.1016/j.jbiotec.2009.11.020

Costa M, Fernandes VO, Ribeiro T et al (2014) Construction of GH16 β-glucanase mini-cellulosomes to improve the nutritive value of barley-based diets for broilers. J Agric Food Chem 62:7496–7506. https://doi.org/10.1021/jf502157y

Cragg SM, Beckham GT, Bruce NC et al (2015) Lignocellulose degradation mechanisms across the Tree of Life. Curr Opin Chem Biol 29:108–119. https://doi.org/10.1016/j.cbpa.2015.10.018

Currie MA, Adams JJ, Faucher F et al (2012) Scaffoldin conformation and dynamics revealed by a ternary complex from the Clostridium thermocellum cellulosome. J Biol Chem 287:26953–26961. https://doi.org/10.1074/jbc.M112.343897

Currie MA, Cameron K, Dias FMV et al (2013) Small angle X-ray scattering analysis of Clostridium thermocellum cellulosome N-terminal complexes reveals a highly dynamic structure. J Biol Chem 288:7978–7985. https://doi.org/10.1074/jbc.M112.408757

Dassa B, Borovok I, Lamed R et al (2012) Genome-wide analysis of acetivibrio cellulolyticus provides a blueprint of an elaborate cellulosome system. BMC Genom 13:210. https://doi.org/10.1186/1471-2164-13-210

Dassa B, Borovok I, Ruimy-Israeli V et al (2014) Rumen cellulosomics: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. PLoS ONE 9:e99221. https://doi.org/10.1371/journal.pone.0099221

Dassa B, Borovok I, Lombard V et al (2017) Pan-cellulosomics of mesophilic clostridia: variations on a theme. Microorganisms 5:74. https://doi.org/10.3390/microorganisms5040074

Davidi L, Moraïs S, Artzi L et al (2016) Toward combined delignification and saccharification of wheat straw by a laccase-containing designer cellulosome. Proc Natl Acad Sci USA 113:10854–10859. https://doi.org/10.1073/pnas.1608012113

Demishtein A, Karpol A, Barak Y et al (2010) Characterization of a dockerin-based affinity tag: application for purification of a broad variety of target proteins. J Mol Recognit 23:525–535. https://doi.org/10.1002/jmr.1029

Doi RH, Kosugi A (2004) Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol 2:541–551. https://doi.org/10.1038/nrmicro925

Dror TW, Morag E, Rolider A et al (2003a) Regulation of the cellulosomal CelS (cel48A) gene of Clostridium thermocellum is growth rate dependent. J Bacteriol 185:3042–3048

Dror TW, Rolider A, Bayer EA et al (2003b) Regulation of expression of scaffoldin-related genes in Clostridium thermocellum. J Bacteriol 185:5109–5116

Dykstra AB, St Brice L, Rodriguez M et al (2014) Development of a multipoint quantitation method to simultaneously measure enzymatic and structural components of the Clostridium thermocellum cellulosome protein complex. J Proteome Res 13:692–701. https://doi.org/10.1021/pr400788e

Fan Z, Yuan L (2010) Production of multifunctional chimaeric enzymes in plants: a promising approach for degrading plant cell wall from within. Plant Biotechnol J 8:308–315. https://doi.org/10.1111/j.1467-7652.2009.00484.x

Fanutti C, Ponyi T, Black GW et al (1995) The Conserved Noncatalytic 40-Residue Sequence in Cellulases and Hemicellulases from Anaerobic Fungi Functions as a Protein Docking Domain. J Biol Chem 270:29314–29322. https://doi.org/10.1074/jbc.270.49.29314

Fierobe H-P, Bayer EA, Tardif C et al (2002) Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. J Biol Chem 277:49621–49630. https://doi.org/10.1074/jbc.M207672200

Fierobe H-P, Mingardon F, Mechaly A et al (2005) Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. J Biol Chem 280:16325–16334. https://doi.org/10.1074/jbc.M414449200

Fontes CMGA, Gilbert HJ (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem 79:655–681. https://doi.org/10.1146/annurev-biochem-091208-085603

Fujimoto Z, Jackson A, Michikawa M et al (2013) The structure of a *streptomyces avermitilis* α-l-rhamnosidase reveals a novel carbohydrate-binding module cbm67 within the six-domain arrangement. J Biol Chem 288:12376–12385. https://doi.org/10.1074/jbc.M113.460097

Fujino T, Béguin P, Aubert JP (1993) Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. J Bacteriol 175:1891–1899

Gefen G, Anbar M, Morag E et al (2012) Enhanced cellulose degradation by targeted integration of a cohesin-fused β-glucosidase into the Clostridium thermocellum cellulosome. Proc Natl Acad Sci USA 109:10298–10303. https://doi.org/10.1073/pnas.1202747109

Gerwig GJ, Kamerling JP, Vliegenthart JF et al (1993) The nature of the carbohydrate-peptide linkage region in glycoproteins from the cellulosomes of Clostridium thermocellum and Bacteroides cellulosolvens. J Biol Chem 268:26956–26960

Gilbert HJ, Knox JP, Boraston AB (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Curr Opin Struct Biol 23:669–677. https://doi.org/10.1016/j.sbi.2013.05.005

Gilkes NR, Warren RA, Miller RC, Kilburn DG (1988) Precise excision of the cellulose binding domains from two Cellulomonas fimi cellulases by a homologous protease and the effect on catalysis. J Biol Chem 263:10401–10407

Gilmore SP, Henske JK, O'Malley MA (2015) Driving biomass breakdown through engineered cellulosomes. Bioengineered 6:204–208. https://doi.org/10.1080/21655979.2015.1060379

Gilmore SP, Lillington SP, Haitjema CH et al (2020) Designing chimeric enzymes inspired by fungal cellulosomes. Synthetic and Systems Biotechnology 5:23–32. https://doi.org/10.1016/j.synbio.2020.01.003

Gold ND, Martin VJJ (2007) Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis. J Bacteriol 189:6787–6795. https://doi.org/10.1128/JB.00882-07

Guillén D, Sánchez S, Rodríguez-Sanoja R (2009) Carbohydrate-binding domains: multiplicity of biological roles. Appl Microbiol Biotechnol 85:1241–1249. https://doi.org/10.1007/s00253-009-2331-y

Hagen LH, Brooke CG, Shaw C et al (2020) Proteome specialization of anaerobic fungi during ruminal degradation of recalcitrant plant fiber. Microbiology

Haimovitz R, Barak Y, Morag E et al (2008) Cohesin-dockerin microarray: diverse specificities between two complementary families of interacting protein modules. Proteomics 8:968–979. https://doi.org/10.1002/pmic.200700486

Haitjema CH, Gilmore SP, Henske JK et al (2017) A parts list for fungal cellulosomes revealed by comparative genomics. Nature Microbiology 2:17087. https://doi.org/10.1038/nmicrobiol.2017.87

Hamberg Y, Ruimy-Israeli V, Dassa B et al (2014) Elaborate cellulosome architecture of Acetivibrio cellulolyticus revealed by selective screening of cohesin-dockerin interactions. PeerJ 2:e636. https://doi.org/10.7717/peerj.636

Hammel M, Fierobe H-P, Czjzek M et al (2004) Structural insights into the mechanism of formation of cellulosomes probed by small angle X-ray scattering. J Biol Chem 279:55985–55994. https://doi.org/10.1074/jbc.M408979200

Hammel M, Fierobe H-P, Czjzek M et al (2005) Structural basis of cellulosome efficiency explored by small angle X-ray scattering. J Biol Chem 280:38562–38568. https://doi.org/10.1074/jbc.M503168200

Han SO, Yukawa H, Inui M, Doi RH (2003) Regulation of expression of cellulosomal cellulase and hemicellulase genes in Clostridium cellulovorans. J Bacteriol 185:6067–6075

Hasunuma T, Okazaki F, Okai N et al (2013) A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. Bioresour Technol 135:513–522. https://doi.org/10.1016/j.biortech.2012.10.047

Henrissat B, Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol 7:637–644. https://doi.org/10.1016/S0959-440X(97)80072-3

Henrissat B, Davies GJ (2000) Glycoside Hydrolases and Glycosyltransferases. Families, Modules, and Implications for Genomics. Plant Physiol 124:1515–1519. https://doi.org/10.1104/pp.124.4.1515

Himmel ME, Bayer EA (2009) Lignocellulose conversion to biofuels: current challenges, global perspectives. Curr Opin Biotechnol 20:316–317. https://doi.org/10.1016/j.copbio.2009.05.005

Himmel ME, Xu Q, Luo Y et al (2010) Microbial enzyme systems for biomass conversion: emerging paradigms. Biofuels 1:323–341. https://doi.org/10.4155/bfs.09.25

Hong W, Zhang J, Feng Y et al (2014) The contribution of cellulosomal scaffoldins to cellulose hydrolysis by Clostridium thermocellum analyzed by using thermotargetrons. Biotechnol Biofuels 7:80. https://doi.org/10.1186/1754-6834-7-80

Hyeon JE, Jeon WJ, Whang SY, Han SO (2011) Production of minicellulosomes for the enhanced hydrolysis of cellulosic substrates by recombinant corynebacterium glutamicum. Enzyme Microb Technol 48:371–377. https://doi.org/10.1016/j.enzmictec.2010.12.014

Hyeon JE, Kang DH, Han SO (2014) Signal amplification by a self-assembled biosensor system designed on the principle of dockerin-cohesin interactions in a cellulosome complex. Analyst 139:4790–4793. https://doi.org/10.1039/c4an00856a

Ichikawa S, Karita S, Kondo M, Goto M (2014) Cellulosomal carbohydrate-binding module from Clostridium josui binds to crystalline and non-crystalline cellulose, and soluble polysaccharides. FEBS Lett 588:3886–3890. https://doi.org/10.1016/j.febslet.2014.08.032

Israeli-Ruimy V, Bule P, Jindou S et al (2017) Complexity of the Ruminococcus flavefaciens FD-1 cellulosome reflects an expansion of family-related protein-protein interactions. Scientific Reports 7:42355. https://doi.org/10.1038/srep42355

Jørgensen H, Kristensen JB, Felby C (2007) Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuels, Bioprod Biorefin 1:119–134. https://doi.org/10.1002/bbb.4

Kahel-Raifer H, Jindou S, Bahari L et al (2010) The unique set of putative membrane-associated anti-σ factors in Clostridium thermocellum suggests a novel extracellular carbohydrate-sensing mechanism involved in gene regulation: biomass-sensing anti-σ factors in C. thermocellum. FEMS Microbiol Lett 308:84–93. https://doi.org/10.1111/j.1574-6968.2010.01997.x

Kahn A, Moraïs S, Galanopoulou AP, et al (2019) Creation of a functional hyperthermostable designer cellulosome. Biotechnol Biofuels 12:. https://doi.org/10.1186/s13068-019-1386-y

Kang S, Barak Y, Lamed R et al (2006) The functional repertoire of prokaryote cellulosomes includes the serpin superfamily of serine proteinase inhibitors. Mol Microbiol 60:1344–1354. https://doi.org/10.1111/j.1365-2958.2006.05182.x

Karmakar M, Ray RR (2011) Current trends in research and application of microbial cellulases. Res J Microbiol 6:41–53. https://doi.org/10.3923/jm.2011.41.53

Karpol A, Kantorovich L, Demishtein A et al (2009) Engineering a reversible, high-affinity system for efficient protein purification based on the cohesin-dockerin interaction. J Mol Recognit 22:91–98. https://doi.org/10.1002/jmr.926

Kataeva IA, Uversky VN, Brewer JM et al (2004) Interactions between immunoglobulin-like and catalytic modules in Clostridium thermocellum cellulosomal cellobiohydrolase CbhA. Protein Eng Des Sel 17:759–769. https://doi.org/10.1093/protein/gzh094

Kim D-M, Nakazawa H, Umetsu M et al (2012) A nanocluster design for the construction of artificial cellulosomes. Catal Sci Technol 2:499. https://doi.org/10.1039/c2cy00371f

Kleine J, Liebl W (2006) Comparative characterization of deletion derivatives of the modular xylanase XynA of Thermotoga maritima. Extremophiles 10:373–381. https://doi.org/10.1007/s00792-006-0509-0

Kruus K, Lua AC, Demain AL, Wu JH (1995) The anchorage function of CipA (CelL), a scaffolding protein of the Clostridium thermocellum cellulosome. Proc Natl Acad Sci USA 92:9254–9258

Kuhad RC, Gupta R, Singh A (2011) Microbial cellulases and their industrial applications. Enzyme Res 2011:1–10. https://doi.org/10.4061/2011/280696

Kurokawa J, Hemjinda E, Arai T et al (2001) Sequence of the Clostridium thermocellum mannanase gene man26B and characterization of the translated product. Biosci Biotechnol Biochem 65:548–554

Lamed R, Setter E, Bayer EA (1983) Characterization of a cellulose-binding, cellulase-containing complex in Clostridium thermocellum. J Bacteriol 156:828–836

Leibovitz E, Béguin P (1996) A new type of cohesin domain that specifically binds the dockerin domain of the Clostridium thermocellum cellulosome-integrating protein CipA. J Bacteriol 178:3077–3084

Lemaire M, Ohayon H, Gounon P et al (1995) OlpB, a new outer layer protein of Clostridium thermocellum, and binding of its S-layer-like domains to components of the cell envelope. J Bacteriol 177:2451–2459

Levy-Assaraf M, Voronov-Goldman M, Rozman Grinberg I et al (2013) Crystal structure of an uncommon cellulosome-related protein module from Ruminococcus flavefaciens that resembles papain-like cysteine peptidases. PLoS ONE 8:e56138. https://doi.org/10.1371/journal.pone.0056138

Liang Y, Si T, Ang EL, Zhao H (2014) Engineered pentafunctional minicellulosome for simultaneous saccharification and ethanol fermentation in Saccharomyces cerevisiae. Appl Environ Microbiol 80:6677–6684. https://doi.org/10.1128/AEM.02070-14

Lombard V, Bernard T, Rancurel C et al (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem J 432:437–444. https://doi.org/10.1042/BJ20101185

Lombard V, Golaconda Ramulu H, Drula E et al (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178

Mazmanian SK (1999) Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. Science 285:760–763. https://doi.org/10.1126/science.285.5428.760

Mesnage S (2000) Bacterial SLH domain proteins are non-covalently anchored to the cell surface *via* a conserved mechanism involving wall polysaccharide pyruvylation. The EMBO J 19:4473–4484. https://doi.org/10.1093/emboj/19.17.4473

Mingardon F, Chanal A, Tardif C et al (2007) Exploration of new geometries in cellulosome-like chimeras. Appl Environ Microbiol 73:7138–7149. https://doi.org/10.1128/AEM.01306-07

Mitsuzawa S, Kagawa H, Li Y et al (2009) The rosettazyme: a synthetic cellulosome. J Biotechnol 143:139–144. https://doi.org/10.1016/j.jbiotec.2009.06.019

Molinier A-L, Nouailler M, Valette O et al (2011) Synergy, structure and conformational flexibility of hybrid cellulosomes displaying various inter-cohesins linkers. J Mol Biol 405:143–157. https://doi.org/10.1016/j.jmb.2010.10.013

Morag E, Halevy I, Bayer EA, Lamed R (1991) Isolation and properties of a major cellobiohydrolase from the cellulosome of Clostridium thermocellum. J Bacteriol 173:4155–4162

Morag E, Lapidot A, Govorko D et al (1995) Expression, purification, and characterization of the cellulose-binding domain of the scaffoldin subunit from the cellulosome of Clostridium thermocellum. Appl Environ Microbiol 61:1980–1986

Moraïs S, Morag E, Barak Y et al (2012) Deconstruction of lignocellulose into soluble sugars by native and designer cellulosomes. MBio 3. https://doi.org/10.1128/mBio.00508-12

Moraïs S, Shterzer N, Lamed R et al (2014) A combined cell-consortium approach for lignocellulose degradation by specialized lactobacillus plantarum cells. Biotechnol Biofuels 7:112. https://doi.org/10.1186/1754-6834-7-112

Moraïs S, Ben David Y, Bensoussan L et al (2016a) Enzymatic profiling of cellulosomal enzymes from the human gut bacterium, ruminococcus champanellensis, reveals a fine-tuned system for cohesin-dockerin recognition. Environ Microbiol 18:542–556. https://doi.org/10.1111/1462-2920.13047

Moraïs S, Cockburn DW, Ben-David Y et al (2016b) Lysozyme activity of the ruminococcus champanellensis cellulosome. Environ Microbiol 18:5112–5122. https://doi.org/10.1111/1462-2920.13501

Mori Y, Ozasa S, Kitaoka M et al (2013) Aligning an endoglucanase Cel5A from thermobifida fusca on a DNA scaffold: potent design of an artificial cellulosome. Chem Commun (Camb) 49:6971–6973. https://doi.org/10.1039/c3cc42614a

Mosbah A, Belaïch A, Bornet O et al (2000) Solution structure of the module X2_1 of unknown function of the cellulosomal scaffolding protein CipC of Clostridium cellulolyticum. J Mol Biol 304:201–217. https://doi.org/10.1006/jmbi.2000.4192

Muñoz-Gutiérrez I, Ortiz de Ora L, Rozman Grinberg I et al (2016) Decoding biomass-sensing regulons of Clostridium thermocellum alternative sigma-I factors in a heterologous bacillus subtilis host system. PLoS ONE 11:e0146316. https://doi.org/10.1371/journal.pone.0146316

Nagy T, Tunnicliffe RB, Higgins LD et al (2007) Characterization of a double dockerin from the cellulosome of the anaerobic fungus Piromyces equi. J Mol Biol 373:612–622. https://doi.org/10.1016/j.jmb.2007.08.007

Nataf Y, Bahari L, Kahel-Raifer H et al (2010) Clostridium thermocellum cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. Proc Natl Acad Sci USA 107:18646–18651. https://doi.org/10.1073/pnas.1012175107

Ou J, Cao Y (2014) Incorporation of Nasutitermes takasagoensis endoglucanase into cell surface-displayed minicellulosomes in Pichia pastoris X33. J Microbiol Biotechnol 24:1178–1188

Pagès S, Bélaïch A, Bélaïch JP et al (1997) Species-specificity of the cohesin-dockerin interaction between Clostridium thermocellum and Clostridium cellulolyticum: prediction of specificity determinants of the dockerin domain. Proteins 29:517–527

Paget M (2015) Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. Biomolecules 5:1245–1265. https://doi.org/10.3390/biom5031245

Paterson GK, Mitchell TJ (2004) The biology of Gram-positive sortase enzymes. Trends in Microbiol 12:89–95. https://doi.org/10.1016/j.tim.2003.12.007

Peer A, Smith SP, Bayer EA et al (2009) Noncellulosomal cohesin—and dockerin-like modules in the three domains of life. FEMS Microbiol Lett 291:1–16. https://doi.org/10.1111/j.1574-6968.2008.01420.x

Perret S, Bélaich A, Fierobe H-P et al (2004) Towards designer cellulosomes in Clostridia: mannanase enrichment of the cellulosomes produced by Clostridium cellulolyticum. J Bacteriol 186:6544–6552. https://doi.org/10.1128/JB.186.19.6544-6552.2004

Pinheiro BA, Gilbert HJ, Sakka K et al (2009) Functional insights into the role of novel type I cohesin and dockerin domains from Clostridium thermocellum. Biochem J 424:375–384. https://doi.org/10.1042/BJ20091152

Pohlschröder M, Leschine SB, Canale-Parola E (1994) Multicomplex cellulase-xylanase system of Clostridium papyrosolvens C7. J Bacteriol 176:70–76. https://doi.org/10.1128/JB.176.1.70-76.1994

Poole D (1992) Identification of the cellulose-binding domain of the cellulosome subunit S1 from Clostridium thermocellum YS. FEMS Microbiol Lett 99:181–186. https://doi.org/10.1016/0378-1097(92)90022-G

Raghothama S, Eberhardt RY, Simpson P et al (2001) Characterization of a cellulosome dockerin domain from the anaerobic fungus piromyces equi. Nat Struct Biol 8:775–778. https://doi.org/10.1038/nsb0901-775

Raman B, Pan C, Hurst GB et al (2009) Impact of pretreated Switchgrass and biomass carbohydrates on Clostridium thermocellum ATCC 27405 cellulosome composition: a quantitative proteomic analysis. PLoS ONE 4:e5271. https://doi.org/10.1371/journal.pone.0005271

Ravachol J, Borne R, Tardif C et al (2014) Characterization of all family-9 glycoside hydrolases synthesized by the cellulosome-producing bacterium Clostridium cellulolyticum. J Biol Chem 289:7335–7348. https://doi.org/10.1074/jbc.M113.545046

Ravachol J, Borne R, Meynial-Salles I et al (2015) Combining free and aggregated cellulolytic systems in the cellulosome-producing bacterium ruminiclostridium cellulolyticum. Biotechnol Biofuels 8:114. https://doi.org/10.1186/s13068-015-0301-4

Ribeiro DO, Viegas A, Pires VMR et al (2019) Molecular basis for the preferential recognition of β1,3-1,4-glucans by the family 11 carbohydrate-binding module from Clostridium thermocellum. The FEBS J. https://doi.org/10.1111/febs.15162

Rincon MT, Martin JC, Aurilia V et al (2004) ScaC, an adaptor protein carrying a novel cohesin that expands the dockerin-binding repertoire of the ruminococcus flavefaciens 17 cellulosome. J Bacteriol 186:2576–2585. https://doi.org/10.1128/JB.186.9.2576-2585.2004

Rincon MT, Cepeljnik T, Martin JC et al (2005) Unconventional mode of attachment of the ruminococcus flavefaciens cellulosome to the cell surface. J Bacteriol 187:7569–7578. https://doi.org/10.1128/JB.187.22.7569-7578.2005

Rincon MT, Dassa B, Flint HJ et al (2010) Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, ruminococcus flavefaciens FD-1. PLoS ONE 5:e12476. https://doi.org/10.1371/journal.pone.0012476

Sabathé F, Bélaïch A, Soucaille P (2002) Characterization of the cellulolytic complex (cellulosome) of Clostridium acetobutylicum. FEMS Microbiol Lett 217:15–22

Salama-Alber O, Jobby MK, Chitayat S et al (2013) Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. J Biol Chem 288:16827–16838. https://doi.org/10.1074/jbc.M113.466672

Schneewind O, Model P, Fischetti VA (1992) Sorting of protein a to the staphylococcal cell wall. Cell 70:267–281. https://doi.org/10.1016/0092-8674(92)90101-H

Schubot FD, Kataeva IA, Chang J et al (2004) Structural basis for the exocellulase activity of the cellobiohydrolase cbha from Clostridium thermocellum †. Biochemistry 43:1163–1170. https://doi.org/10.1021/bi030202i

Shimon LJ, Pagès S, Belaich A et al (2000) Structure of a family IIIa scaffoldin CBD from the cellulosome of Clostridium cellulolyticum at 2.2 A resolution. Acta Crystallogr D Biol Crystallogr 56:1560–1568

Simpson PJ, Xie H, Bolam DN et al (2000) The Structural Basis for the Ligand Specificity of Family 2 Carbohydrate-binding Modules. J Biol Chem 275:41137–41142. https://doi.org/10.1074/jbc.M006948200

Sleytr UB, Messner P, Pum D, Sára M (1993) Crystalline bacterial cell surface layers. Mol Microbiol 10:911–916. https://doi.org/10.1111/j.1365-2958.1993.tb00962.x

Slutzki M, Reshef D, Barak Y et al (2015) Crucial roles of single residues in binding affinity, specificity, and promiscuity in the cellulosomal cohesin-dockerin interface. J Biol Chem 290:13654–13666. https://doi.org/10.1074/jbc.M115.651208

Smith SP, Bayer EA, Czjzek M (2017) Continually emerging mechanistic complexity of the multienzyme cellulosome complex. Curr Opin Struct Biol 44:151–160. https://doi.org/10.1016/j.sbi.2017.03.009

Stahl SW, Nash MA, Fried DB et al (2012) Single-molecule dissection of the high-affinity cohesin-dockerin complex. Proc Natl Acad Sci USA 109:20431–20436. https://doi.org/10.1073/pnas.1211929109

Steenbakkers PJ, Li XL, Ximenes EA et al (2001) Noncatalytic docking domains of cellulosomes of anaerobic fungi. J Bacteriol 183:5325–5333

Stern J, Kahn A, Vazana Y et al (2015) Significance of relative position of cellulases in designer cellulosomes for optimized cellulolysis. PLoS ONE 10:e0127326. https://doi.org/10.1371/journal.pone.0127326

Tamaru Y, Doi RH (2001) Pectate lyase A, an enzymatic subunit of the Clostridium cellulovorans cellulosome. Proc Natl Acad Sci USA 98:4125–4129. https://doi.org/10.1073/pnas.071045598

Tormo J, Lamed R, Chirino AJ et al (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. EMBO J 15:5739–5751

Tsai S-L, Goyal G, Chen W (2010) Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production. Appl Environ Microbiol 76:7514–7520. https://doi.org/10.1128/AEM.01777-10

Vazana Y, Barak Y, Unger T et al (2013) A synthetic biology approach for evaluating the functional contribution of designer cellulosome components to deconstruction of cellulosic substrates. Biotechnol Biofuels 6:182. https://doi.org/10.1186/1754-6834-6-182

Venditto I, Luis AS, Rydahl M et al (2016) Complexity of the Ruminococcus flavefaciens cellulosome reflects an expansion in glycan recognition. Proc Natl Acad Sci USA 113:7136–7141. https://doi.org/10.1073/pnas.1601558113

Voronov-Goldman M, Yaniv O, Gul O et al (2015) Standalone cohesin as a molecular shuttle in cellulosome assembly. FEBS Lett 589:1569–1576. https://doi.org/10.1016/j.febslet.2015.04.013

Wang Y, Leng L, Islam MK et al (2019) Substrate-related factors affecting cellulosome-induced hydrolysis for lignocellulose valorization. Int J Mol Sci 20:3354. https://doi.org/10.3390/ijms20133354

Willson BJ, Kovács K, Wilding-Steele T et al (2016) Production of a functional cell wall-anchored minicellulosome by recombinant Clostridium acetobutylicum ATCC 824. Biotechnol Biofuels 9:109. https://doi.org/10.1186/s13068-016-0526-x

Wilson Catriona A, Wood Thomas M (1992) The anaerobic fungus Neocallimastix frontalis: isolation and properties of a cellulosome-type enzyme fraction with the capacity to solubilize hydrogen-bond-ordered cellulose. Appl Microbiol Biotechnol 37. https://doi.org/10.1007/BF00174216

Xu Q, Gao W, Ding S-Y et al (2003) The cellulosome system of Acetivibrio cellulolyticus includes a novel type of adaptor protein and a cell surface anchoring protein. J Bacteriol 185:4548–4557

Xu Q, Barak Y, Kenig R et al (2004a) A novel Acetivibrio cellulolyticus anchoring scaffoldin that bears divergent cohesins. J Bacteriol 186:5782–5789. https://doi.org/10.1128/JB.186.17.5782-5789.2004

Xu Q, Bayer EA, Goldman M et al (2004b) Architecture of the Bacteroides cellulosolvens cellulosome: description of a cell surface-anchoring scaffoldin and a family 48 cellulase. J Bacteriol 186:968–977

Xu C, Huang R, Teng L et al (2015) Cellulosome stoichiometry in Clostridium cellulolyticum is regulated by selective RNA processing and stabilization. Nat Commun 6:6900. https://doi.org/10.1038/ncomms7900

Xu Q, Resch MG, Podkaminer K et al (2016) Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities. Sci Adv 2:e1501254. https://doi.org/10.1126/sciadv.1501254

You C, Zhang X-Z, Sathitsuksanoh N et al (2012) Enhanced microbial utilization of recalcitrant cellulose by an ex vivo cellulosome-microbe complex. Appl Environ Microbiol 78:1437–1444. https://doi.org/10.1128/AEM.07138-11

Ze X, Duncan SH, Louis P, Flint HJ (2012) Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. The ISME J 6:1535–1543. https://doi.org/10.1038/ismej.2012.4

Ze X, Ben David Y, Laverde-Gomez JA et al (2015) Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic firmicutes bacterium ruminococcus bromii. mBio 6. https://doi.org/10.1128/mBio.01058-15

Zhivin O, Dassa B, Moraïs S, et al (2017) Unique organization and unprecedented diversity of the Bacteroides (Pseudobacteroides) cellulosolvens cellulosome system. Biotechnology for Biofuels 10. https://doi.org/10.1186/s13068-017-0898-6