Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides

Lior Artzi, Edward A. Bayer and Sarah Moraïs

Abstract | Cellulosomes are multienzyme complexes that are produced by anaerobic cellulolytic bacteria for the degradation of lignocellulosic biomass. They comprise a complex of scaffoldin, which is the structural subunit, and various enzymatic subunits. The intersubunit interactions in these multienzyme complexes are mediated by cohesin and dockerin modules. Cellulosome-producing bacteria have been isolated from a large variety of environments, which reflects their prevalence and the importance of this microbial enzymatic strategy. In a given species, cellulosomes exhibit intrinsic heterogeneity, and between species there is a broad diversity in the composition and configuration of cellulosomes. With the development of modern technologies, such as genomics and proteomics, the full protein content of cellulosomes and their expression levels can now be assessed and the regulatory mechanisms identified. Owing to their highly efficient organization and hydrolytic activity, cellulosomes hold immense potential for application in the degradation of biomass and are the focus of much effort to engineer an ideal microorganism for the conversion of lignocellulose to valuable products, such as biofuels.

Cellulose
A crystalline polysaccharide comprising very long, linear, unbranched chains of pure glucose (up to 15,000 units) that are connected by β-1,4-linkages and are resistant to simple enzymatic hydrolysis.

Hemicellulose
A mixture of branched polysaccharides that are relatively easily hydrolysable. Hemicelluloses include, among other polysaccharides, xylans, xyloglucans, glucuronoxylans, arabinoxylans, arabinans, mannans, glucomannans and β-glucans.

Department of Biomolecular Sciences, The Weizmann Institute of Science, 234 Herzl Street, Rehovot 7610001, Israel.

Correspondence to E.A.B. and S.M.
ed.bayer@weizmann.ac.il; sarahv@weizmann.ac.il

doi:10.1038/nrmicro.2016.164
Published online 12 Dec 2016

The plant cell wall is a recalcitrant network of polysaccharides that are highly resistant to enzymatic degradation and is composed of cellulose, which is the main component, hemicellulose (such as xylans), the non-polysaccharide aromatic polymer lignin, and other polysaccharides (such as pectin) and proteins[1–3].

Different species of cellulolytic bacteria and fungi can hydrolyse lignocellulose in plant cell walls[4]. In the early 1980s, the cellulosome complex was discovered in the highly cellulolytic thermophilic anaerobe *Clostridium thermocellum*[5,6]. Cellulosomes are multienzyme complexes that are produced by a select number of anaerobic bacteria. Cellulosome-producing bacterial species have been identified in different ecosystems, including forest and pasture soils, hot spring pools, sewage sludge, compost piles and the microbiota of both vertebrates and invertebrates[1,7,8] (TABLE 1). The energy levels in anaerobic bacteria limit the production of enzymes, and, to overcome this limitation, anaerobic bacteria developed the cellulosome — a highly efficient, highly organized, cell surface enzymatic system[4] that enables enzyme recycling and the direct assimilation of hydrolytic products. Furthermore, cellulosomes physically separate individual cellulose microfibrils from larger particles, which results in better access to cellulose surfaces[9,10]. Cellulosome-producing bacteria inhabit complex environments that comprise communities of numerous cellulolytic and saccharolytic microorganisms and are characterized by particularly recalcitrant polysaccharide substrates. These highly sophisticated bacteria provide a major polymer-degrading function in the lignocellulosic ecosystem and release ample amounts of soluble saccharides and other degradation products for accompanying non-cellulolytic 'satellite' strains[11].

Cellulosomes are composed of two main types of building block: dockerin-containing enzymes or other types of ancillary protein, and cohesin-containing structural proteins, which are termed scaffoldins. Cohesins and dockerins are complementary modules that bind tightly to each other. The specificity characteristics of the cohesin–dockerin interaction dictate the integration of the enzymes into the complex as well as its final architecture[8,11]. Scaffoldins can also contain a dockerin module for binding to other scaffoldins and a carbohydrate-binding module (CBM) for targeting the complex and its enzymes to appropriate sites on the plant cell wall substrate. Cellulosomes can be attached to the bacterial cell surface or can be released as cell-free cellulosomes[12,13] (FIG. 1).

In this Review, we discuss the cellulosome-producing bacteria that are currently known and their role in the environment, the composition of different cellulosomes and the structural characterization of key cellulosome modules. In addition, we discuss the enzyme diversity, function and regulation of cellulosomes, and their various applications, including the degradation of biomass for potential biofuel production.

Table 1 | **Cellulosome-producing bacteria**

Species	Temperature	Source	Largest anchoring scaffoldin*	Largest adaptor scaffoldin*	Largest primary scaffoldin*	Largest cellulosome complex‡	Data accession number§
Highly structured cellulosomes							
Acetivibrio cellulolyticus	Mesophile	Sewage sludge	3	4	7	96	GCA_000179595.2
Pseudobacteroides cellulosolvens	Mesophile	Sewage sludge	10	–	11	110	NZ_LGTC00000000.1
Clostridium alkalicellulosi	Mesophile	Soda lake	2	4	10	40	Ga0025046
Clostridium clariflavum	Thermophile	Thermophilic methanogenic bioreactor	4	5	8	160	NC_016627.1
Clostridium straminisolvens	Thermophile	Cellulose-degrading bacterial community	NA	NA	NA	NA	GCF_000521465.1
Clostridium thermocellum	Thermophile	Horse manure, hot springs, sewage and soil	7	–	9	63	NC_009012.1
Ruminococcus champanellensis	Mesophile	Human gut	1	2	7	11	NC_021039.1
Ruminococcus flavefaciens	Mesophile	Rumen	1	9	2	14	NZ_ACOK00000000.1
Simple cellulosomes							
Clostridium acetobutylicum	Mesophile	Soil	–	–	5	5	CP002660
Clostridium sp. BNL1100	Mesophile	Corn stover	–	–	6	6	CP003259
Clostridium bornimense	Mesophile	Biogas reactor	–	–	5	5	HG917868.1, HG917869.1
Clostridium cellobioparum	Mesophile	Rumen	NA	NA	NA	NA	JHYD01000000
Clostridium cellulolyticum	Mesophile	Compost	–	–	8	8	CP001348
Clostridium cellulovorans	Mesophile	Wood fermenter	–	–	9	9	CP002160
Clostridium josui	Mesophile	Compost	–	–	6	6	JAGE00000000.1
Clostridium papyrosolvens	Mesophile	Paper mill	–	–	6	6	GCA_000421965.1
Clostridium saccharoperbutylacetonicum	Mesophile	Soil	–	–	2	2	CP004121.1
Clostridium termitidis	Mesophile	Termite gut	–	–	2	5	AORV00000000.1
Ruminococcus bromii	Mesophile	Human gut	2	–	1	2	FP929051.1

NA, not available (or the data are too poor to make a proper estimate). *Number of cohesins. ‡Estimated number of enzymes. §From the NCBI Assembly Database.

Cellulosome-producing microorganisms

Cellulosome-producing bacteria originate from different anaerobic habitats (TABLE 1), belong to different genera and species, and can be thermophilic or mesophilic. The prototypical cellulosome-producing bacterium is the thermophile *C. thermocellum*, which has the most-studied and best-characterized cellulosome system.

C. thermocellum has eight scaffoldin genes[14] and its highly structured cellulosome complex contains up to 63 enzymes. The mesophilic bacterium *Acetivibrio cellulolyticus* contains an even more complicated cellulosome system that has 16 scaffoldins. The concept of an 'adaptor scaffoldin' (see below) was first defined in this bacterium[15]. Adaptor scaffoldins enable cellulosomes to incorporate even more enzymes into one complex, and the *A. cellulolyticus* cellulosome can thus integrate 96 enzymatic subunits. The thermophilic bacterium *Clostridium clariflavum* has a similar cellulosome system to that of *A. cellulolyticus*, with a record number of 160 enzymes in a single complex[16]. *Ruminococcus flavefaciens* was isolated from the cow rumen and produces a cellulosomal system that has

Lignin
A complex organic polymer that consists of various crosslinked aromatic alcohols. Lignin can be covalently linked to hemicellulose by ferulic acid side chains. Lignin confers structural support and turgidity to the plant cell wall.

an enormous number of components, including 222 dockerin-containing proteins that are divided into six different subtypes[17]. Recently, the cellulolytic bacterium *Ruminococcus champanellensis*, which was discovered in the human gut[18–21], was also observed to produce an elaborate cellulosome system[19,22].

All bacterial species that are known to produce simple cellulosome systems are mesophiles (TABLE 1); these bacteria secrete a relatively small complex that is based on a single scaffoldin that incorporates up to nine enzymatic subunits, according to the number of cohesin modules on the scaffoldin. Such bacteria include *Clostridium cellulolyticum*[23], *Clostridium cellulovorans*[24], *Clostridium josui*[1], *Clostridium papyrosolvens*[25], *Ruminococcus albus* and *Ruminococcus bromii*[26,27]. *R. albus* was isolated from the cow rumen and contains numerous dockerin-containing proteins, but only one cohesin sequence was found in its genome. The paucity of scaffoldins in this species would argue against an authentic cellulosome in this cellulolytic bacterium, and why this bacterium produces dockerin modules remains a mystery[28]. Intriguingly, *R. bromii*, which was isolated from the human gut, has the ability to degrade starch, but not cellulose. Some of its starch-degrading enzymes contain dockerin modules that were shown to interact with cohesins identified in the genome. In this case, the 'cellulosome' complexes can be referred to as amylosomes[26,27].

The possible existence of cellulosome complexes in anaerobic fungi (namely in the genera *Piromyces*, *Orpinomyces* and *Neocallimastix*) should be noted. However, the apparent lack of authentic cohesins and bona fide dockerins in the fungal systems would argue that these complexes differ markedly from bacterial cellulosomes[29].

Initially, enzymes and scaffoldins were discovered by seeking specific protein functions or by searching for genes that are adjacent to known cellulosomal genes, but methodical approaches for the identification of full cellulosomal systems in a particular microorganism were lacking[15,30–32]. Recently, genome-sequencing methods have improved remarkably, and systematic metagenomic approaches for genome and protein analysis are now common, thus facilitating the discovery of new, complete or fragmented cellulosome systems (BOX 1). These analyses can be applied to newly discovered bacteria to search for new cellulosomes or for further characterization of cellulosomes that are produced by well-known cellulosomal bacterial species[14,16,19,28,33].

Recent deep-sequencing analysis of cellulosomal elements in fibre-rich rumen metagenomes[34] revealed their remarkable phylogenetic diversity and identified many unknown proteins from uncultured bacteria and from the Bacteroidetes phylum, such as species in the *Prevotella* and *Bacteroides* genera, that were not documented before. In addition, uncommon cohesin sequences could not be classified into one of the three previously known types, which suggests that type classification is much more diverse than previously considered. Importantly, the dockerin-containing proteins had functions that were associated with catabolic processes as well as other microbial interactions, which suggests a broader role for the cellulosomal machinery than only fibre degradation.

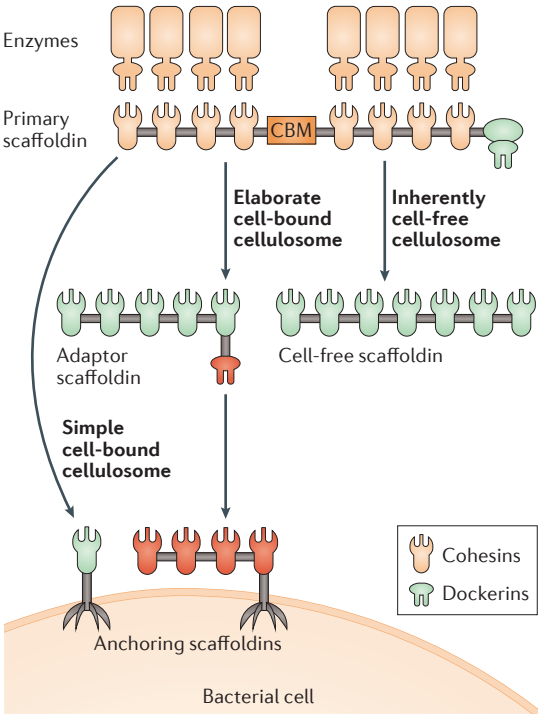

Figure 1 | **Types of cellulosome system.** A primary scaffoldin can be bound directly to the cell surface through an anchoring scaffoldin or, in a more elaborate system, through an intermediary adaptor scaffoldin. Cellulosomes can also exist in an inherently cell-free state through attachment to an appropriate free scaffoldin. In *Clostridium clariflavum*, direct attachment to the cell surface generates cell-bound cellulosomes comprising multiples of eight enzymes, depending on the number of cohesins on the anchoring scaffoldin. Successive assembly of primary, adaptor and anchoring scaffoldins amplifies the number of enzyme subunits, which results in a surface-bound cellulosome complex of 160 enzymes in this particular bacterium. The inherently cell-free cellulosome of *C. clariflavum* can reach up to 56 enzymatic subunits in a single complex, which is secreted from the cell and disperses in solution. CBM, carbohydrate-binding module.

Interestingly, cohesin-encoding and/or dockerin-encoding genes have also been identified in the genomes of non-cellulolytic microorganisms. However, only one or two cohesin modules per protein were detected. For example, the archaeon *Archaeoglobus fulgidus*[35,36] contains a gene that encodes one cohesin and a tandem gene that encodes both a cohesin and a dockerin[37]. Nevertheless, *A. fulgidus* lacks identifiable cellulosomal glycoside hydrolases. Several cohesin and dockerin modules were also found in the non-cellulolytic bacterium *Clostridium perfringens*, which is an opportunistic pathogen that inhabits the human and animal gastrointestinal tracts. Interestingly, these cohesin-containing and dockerin-containing proteins are all putative glycoside hydrolases, which suggests a role for these proteins in the degradation of mammalian polysaccharides[38].

Lignocellulose
A general term that refers to the composite of cellulose, hemicellulose, pectin and lignin, and indicates the plant cell wall and biomass.

Modules
Independently folding portions of proteins that have independent functions.

Thermophile
A bacterium or fungus that lives in an environment with relatively high temperatures. The optimum temperature for the growth of thermophiles is usually between 45 °C and 70 °C. Hyperthermophilic microorganisms thrive at temperatures that exceed 70 °C.

Mesophiles
Bacteria and fungi that live at moderate temperatures. The optimum temperature for the growth of mesophiles is usually between 20 °C and 45 °C.

Glycoside hydrolases
Enzymes that hydrolyse the glycosidic linkage between two carbohydrates or between a carbohydrate and a non-carbohydrate group.

Box 1 | **Contribution of omics studies to the cellulosome field**

Genomics entails the complete sequencing of single genomes, usually of pure, cultivated microbial clones. Metagenomics is a broader approach, in which the genomic content of all of the microorganisms that are present in a desired ecosystem is explored. Metagenomic studies enable the identification and quantification of microbial species that inhabit the natural environment from which the sample was taken, thus identifying bacterial species that cannot be cultivated in the laboratory. In addition, new genes can be discovered that encode novel enzymes or other putative proteins that are important for metabolic processes[140].

Metagenomic studies have been carried out on samples from the termite hindgut[141], cow rumen[34,142,143], gut microbiomes[144], enriched thermophilic cellulose-degrading sludge[145] and mangroves[146], in which numerous characterized and uncharacterized genes that encode polysaccharide-hydrolysing and cohesin-containing and dockerin-containing proteins were detected. These studies showed that the microbial diversity in each environment is enormous[142]. In addition, enzymes that have a role in the degradation of biomass are different in each environment.

For finding specific genes of interest, such as carbohydrate-active enzyme and cellulosomal genes, a gene-centric approach can be used[140]. This approach emphasizes the functionality of each community, based on the abundance of genes that are related to the community[140]. The most active genes that are involved in the degradation of biomass can thus be identified, which would help select genes of interest for further biochemical investigation. Metagenomic data benefit from metatranscriptomics and metaproteomics (the full mRNA and protein content in an environmental sample) to determine which enzymes or genes are most expressed and active.

Proteomic approaches were used to identify key cellulosomal components. Whole-genome sequencing has enabled analyses of the composition of cellulosomes by mass spectrometry methods that have become more sophisticated and accurate over the years. Cellulosomes that have been studied in this way include those from *Clostridium thermocellum*[81,82], *Clostridium cellulovorans*[147,148], *Clostridium cellulolyticum*[149], *Ruminococcus flavefaciens*[150], *Clostridium clariflavum*[46] and *Clostridium termitidis*[151]. Cellulosome complexes were purified from the extracellular medium by different methods. The cellulosomes were reported to be released from the bacterial cell wall during late stationary phase, which enabled their isolation from the used growth medium[6,43,81]. In some studies, different cellulosome populations were isolated from the same growth medium. For example, the cellulosome of *C. cellulolyticum* was fractionated by ion-exchange chromatography, which resulted in six fractions, each comprising differing enzyme profiles, and the catalytic activity of each fraction consequently diverged[152]. For bacterial species that have more complex cellulosome systems, such as *C. clariflavum*, gel filtration chromatography resulted in two major fractions that were distinct in both scaffoldin and enzyme content[46].

Although the cellulosomes of each bacterial species are generally divergent, the most abundant enzymes usually belong to similar glycoside hydrolase families. The exoglucanase glycoside hydrolase 48 (GH48) is one of the most highly expressed enzymes in every cellulosome described to date. The GH9 and GH5 enzyme families (usually cellulases) are also very prominent in cellulosomes, as are GH10, GH11 and GH26 hemicellulases. Many other families are commonly found in cellulosomes, although in much lower abundance.

In addition to proteomics, transcriptomics has been applied to detect cellulosome-related gene expression during the growth of *C. thermocellum*[153] and for understanding the effect of different conditions and substrates on gene expression[154].

Some were identified as toxins and were shown to form two-enzyme or enzyme–toxin complexes that promote virulence[36]. Overall, cohesins and/or dockerins were identified in the genomes of 40% of the archaea and 14% of the bacteria, but only in a few primitive eukaryotes[36]. There is evidence of extensive horizontal gene transfer between bacteria and other microorganisms[36]. The distribution of these modules among so many non-cellulosomal microorganisms indicates broad use of the strong cohesin–dockerin interaction for different, currently undefined functions. Importantly, several cohesin-containing scaffoldins seem to be characteristic only of cellulosome-producing bacteria, which indicates that the cellulosome paradigm may be the exception rather than the rule for the use of these modules.

Cellulosome composition and structure

Cellulosomes can be divided into two major groups: simple cellulosomes and highly structured cellulosomes (TABLE 1). Highly structured cellulosomes are composed of more than one scaffoldin and, consequently, contain many more enzymes in a single assembly than simple cellulosomes. By contrast, simple cellulosomes are based on a single primary scaffoldin.

Scaffoldins. Three major types of scaffoldin — primary, anchoring and adaptor scaffoldins — can form cellulosomes, although not every species contains all types. The most important scaffoldin is the primary scaffoldin[39], which is usually the most highly expressed scaffoldin and contains numerous cohesins that interact with dockerin-containing enzymes[23,24,31]. The primary scaffoldin usually contains a cellulose-binding CBM that targets the complex to its substrate. For simple cellulosomes, the mechanism by which the single primary scaffoldin is attached to the cell surface is unknown[40]. However, cell surface attachment was suggested by electron micrographs that showed protuberances on the cell surface. Furthermore, the cells were shown to attach to cellulose fibres[40], and, for *C. cellulovorans*, cell attachment of cellulosomes has been proposed to occur through the enzyme EngE, which interacts, through its dockerin module, with the cohesin of the primary scaffoldin, cellulose-binding protein A (CbpA). EngE also contains a peptidoglycan-binding S-layer homology (SLH) domain that could anchor the scaffoldin to the cell surface[41].

Highly structured cellulosomes contain several scaffoldins and many enzymes. In these cellulosomes, the primary scaffoldin sometimes contains a specialized

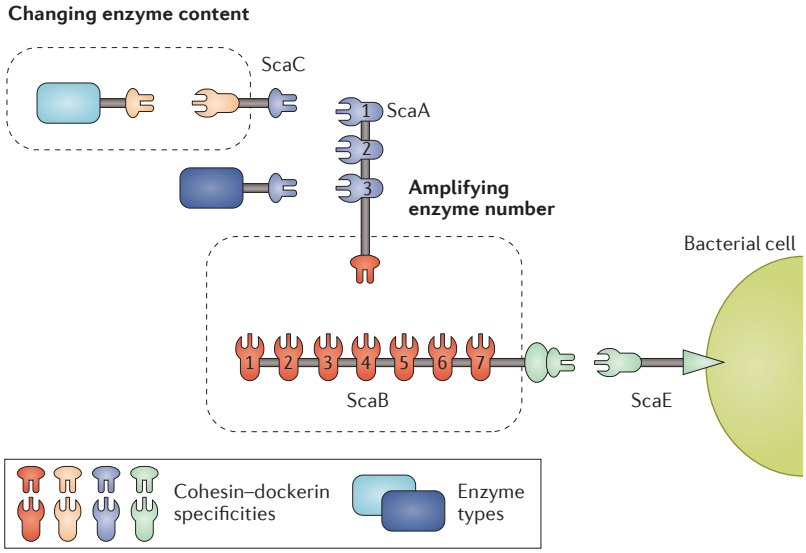

Figure 2 | **Alternative roles of adaptor scaffoldins.** The cellulosome of *Ruminococcus flavefaciens* is covalently linked to the cell surface through a sortase recognition motif that is located at the carboxyl terminus of the anchoring scaffoldin E (ScaE). In strain 17 (REF. 45), the ScaE cohesin binds to the polyvalent adaptor scaffoldin ScaB, the seven cohesins of which bind to the primary ScaA dockerin, which acts to amplify the number of enzymes on the ScaB scaffoldin. The ScaA cohesins bind either directly to a dockerin-containing enzyme or to the dockerin of the monovalent adaptor scaffoldin ScaC, of which its single cohesin of alternative specificity changes the composition of the enzymes integrated onto the ScaA scaffoldin. The system is characterized by four different cohesin–dockerin specificities, which are colour-coded.

dockerin that mediates cell surface attachment by interacting with the cohesin, or cohesins, of an anchoring scaffoldin. Anchoring scaffoldins interact with the cell surface through specialized anchoring modules, either non-covalently through SLH domains or covalently through sortase motifs[42,43]. The more complex cellulosomes contain adaptor scaffoldins that either connect two scaffoldins or a scaffoldin and an enzyme. These scaffoldins may have a regulatory role in determining the assembly and composition of a cellulosome complex, depending on the available substrate (FIG. 2). Monovalent (single cohesin) adaptor scaffoldins can change the type of enzyme that is integrated into a cellulosome and can be regarded as a 'switch' that changes the cohesin specificity of the primary scaffoldin[22,44]. Depending on the substrate, different enzymes with different activities can thus be integrated into the cellulosomal complex. By contrast, polyvalent adaptor scaffoldins (containing several cohesins) can act as a platform for the expansion of the cellulosome complex and the integration of multiple enzymes[14,15], thus enabling more efficient substrate hydrolysis. Polyvalent adaptor scaffoldins have been found in the cellulosomes of *A. cellulolyticus*[15] and *C. clariflavum*[16], whereas monovalent forms have been identified in *R. flavefaciens*[44,45] and *R. champanellensis*[19].

Most of the cellulosomes that have been described to date are cell-anchored. However, recently, evidence for inherent cell-free scaffoldins was reported in *C. thermocellum*, *C. clariflavum* and *A. cellulolyticus* (FIG. 1). The secretion of cellulosomes was verified experimentally for *C. thermocellum*[13] and *C. clariflavum*[46]. The expression of cell-free versus cell-anchored cellulosomes has yet to be examined quantitatively. Cell-free cellulosomes are composed of different combinations of scaffoldins compared with cell-anchored cellulosomes, which suggests that their expression is different. Cell-free scaffoldins lack an appropriate anchoring domain or motif that could tether the complex to the cell surface. In addition, it has long been recognized that anchored cellulosomes detach from the cell surface during the later stages of growth[47].

Enzymes. The cellulosome of *C. thermocellum* was initially discovered owing to its ability to adhere to and hydrolyse cellulose[5,6], and numerous cellulases were found in its cellulosome. In addition to cellulases, other polysaccharide-degrading, carbohydrate-active cellulosomal enzymes were subsequently identified; most notably, xylanases, pectinases, mannanases and xyloglucanases. Plant cell wall-degrading enzymes are particularly diverse and complex. They include glycoside hydrolases, carbohydrate esterases and polysaccharide lyases. These enzymes are broadly grouped according to their functionality, and are classified into families[48] according to the primary sequence of their catalytic module, its consequent structure and mechanism of action.

Carbohydrate-active enzymes characteristically act synergistically to hydrolyse resistant plant-derived substrates. Synergism may emanate from different modes of action towards the same substrate. For example, an exoglucanase could cleave more chain ends if an endoglucanase concomitantly hydrolyses the substrate, thereby producing additional chain ends[49]. Synergy could also result from the hydrolysis of two intermingled substrates, in which the action of one enzyme could make the concealed substrate accessible for the action of the second enzyme; for example, the cooperation of cellulases and xylanases. In addition, one enzyme could decrease the product inhibition of another enzyme thereby restoring its activity, such as the effect of β-glucosidases on cellulases[50]. Synergism between different glycoside hydrolases is often observed[51]. In this respect, different types of synergistic glycoside hydrolases and/or carbohydrate esterases can be found in the same cellulosomal protein, forming highly active multifunctional enzymes[7,52].

Intriguingly, all known cellulosome-producing bacteria characteristically produce a single glycoside hydrolase 48 (GH48) exoglucanase, which is usually expressed in very large amounts and is crucial for enzymatic activity[46,53,54]. By contrast, they generally produce an extensive repertoire of family 9 glycoside hydrolases. Recently, the entire set of 13 GH9 enzymes in *C. cellulolyticum* was characterized[55]; these enzymes exhibit different activities, distinct abilities to bind to cellulosic substrates and diverse synergies with the major Cel48A exoglucanase, independently of their modular organization. The GH9 enzymes of *R. champanellensis* were also examined for different activities on different cellulolytic substrates, and distinct synergies with the Cel48A exoglucanase were observed[22]. These reports

Cellulases
A group of enzymes that catalyse cleavage of the cellulose chain.

Carbohydrate esterases
Enzymes that deacetylate substituted saccharides or alkyl (for example, ethyl) groups of hemicelluloses.

Polysaccharide lyases
Enzymes that cleave polysaccharide chains that contain uronic acid.

Exoglucanase
A cellulase that hydrolyses the cellulose chain only at one of its free termini (reducing or non-reducing) and then degrades the substrate in a processive manner.

Endoglucanase
A cellulase that can hydrolyse the glycosidic bond at any site along a cellulose chain.

β-Glucosidases
Cellulases that hydrolyse only terminal, non-reducing glucose units from short cellodextrins.

suggest the importance of enzyme diversity, especially for the GH9 enzymes, for efficient cellulose degradation. Other glycoside hydrolases, such as GH5, GH10, GH11 and GH43, are common components of cellulosome systems, thereby providing bacteria with a powerful and diverse enzymatic apparatus for the enhanced hydrolysis of plant wall polysaccharides.

In addition to dockerin-containing carbohydrate-active enzymes, other dockerin-containing proteins are present in cellulosomes, such as serpins[56], proteases[57] and expansin-like proteins[58,59]. These proteins have unique functions that are uncommon to cellulosomes, and their diverse roles may contribute to physiological processes in bacteria, to the assembly and regulation of cellulosome components and/or indirectly to the degradation of biomass.

Cohesin and dockerin modules. The cohesin–dockerin interaction is the fundamental basis for the assembly of cellulosome complexes. This non-covalent interaction is one of the strongest known in nature[60–62] and is very difficult to dissociate[63]. Intriguingly, the force that is required to break the interaction between a cohesin–dockerin pair was estimated to be half of the force required to break a covalent bond[64,65]. Cellulosome complexes can contain dozens of interconnected components that are reinforced by these strong interactions. Such a strong bond between the cohesin and dockerin modules is required for maintaining the assembly and stability of the complex under adverse environmental conditions.

Three types of cohesin–dockerin interaction have been described that are based on the sequences of each cohesin–dockerin pair and their binding partners. Type I interactions occur between dockerin-containing catalytic subunits and cohesins of the primary scaffoldin. Type II interactions occur between two scaffoldins (usually anchoring scaffoldins and primary scaffoldins), but there are exceptions[46]. Curiously, *Bacteroides cellulosolvens* is the only known bacterium to have the opposite interaction pattern, whereby its enzymes contain type II dockerins, whereas the scaffoldins contain type I dockerins[66]. Type III interactions are observed in ruminococcal cellulosomes; these interactions are distinct from the type I and type II interactions that are observed in *Clostridium* spp. (REF. 28).

Structural studies of cohesin and dockerin modules identified crucial interface residues that are responsible for the strong intermodular binding and specificity, and that are characteristic for cellulosome architecture. The type I cohesin module is usually composed of approximately 150 amino acids, and is constructed of two nine-stranded β-sheets that are arranged in jelly-roll topology[67,68]. All three cohesin types interact with their dockerins through β-strands 5, 6, 3 and 8 of their β-sheets[67,69]. The type III cohesin of *R. flavefaciens* ScaE has a very similar topology to that of type I cohesins and interacts with its dockerin module through β-strands 5, 6, 3 and 8, similar to that of the type I interaction. The type III cohesin of ScaE contains two 'β-flaps' between β-strands 4 and 8, similar to those of type II cohesins, but also has a prominent 13-residue α-helix that is enveloped by an extensive amino-terminal loop that is not found in other cohesin types[70].

The type I dockerin module contains approximately 70 amino acids and folds into two tandem repeats, each of which comprises a distinctive Ca^{2+}-binding loop and α-helix. Characteristic 'recognition residues' have been noted at helix positions 10, 11, 17 and 18, which are generally conserved in both repeats[71]. Surprisingly, the dockerin typically interacts with the cohesin through either of the two α-helix repeats, but not both[67]. A dual-binding mode was thus proposed for the type I interaction, whereby either of the two symmetrical repeats can bind to the cohesin surface, but in 180° rotation[68]. This implies that dockerin-containing proteins can be incorporated into the complex in two different orientations that may either avert steric clashes in large, multicomponent cellulosomes and/or promote conformational alterations in the position of the parent enzyme during degradation of the cellulosic substrate.

In regard to the structure of the type I dockerin module itself, recent structural studies have resolved a long-standing enigma in the field. It had previously been inferred that the dockerin undergoes conformational changes following cohesin binding. However, new evidence now favours an inherent cohesin-primed conformation of the dockerin without cohesin-induced alterations to its structure[72].

In addition to type I cohesin–dockerin interactions, crystal structures of type II and type III complexes were also solved[69,70]. Interestingly, the type II interaction in *C. thermocellum* does not have a dual-binding mode, and the two helices interact with the cohesin through several interactions at the cohesin–dockerin interface. However, despite initial evidence that the type I and type II interactions reflect dual-binding and single-binding modes, respectively, it is now clear that the mode of binding is not strictly indicative of the modular type[73]. By contrast, the conserved or divergent nature of the recognition residues in the two repeated segments determines the binding mode of a given dockerin.

Type III interactions can also be of either single-binding or dual-binding mode. In fact, the ruminococcal type III cohesins and dockerins are highly diverse, and, in some cases, the sequence of the second calcium-binding dockerin loop is severely distorted. The type III dockerin of the CttA protein from *R. flavefaciens* contains two additional helices, but interacts with the ScaE cohesin in a manner similar to that of the type I interaction[70]. The specialized atypical type III dockerins that have extra helices are rare and contain three unusual sequence inserts that act as structural buttresses to support the extended stalk-like neighbouring X module. The latter module probably maintains the parent protein at a fixed distance from the cell surface and thus requires the additional physical reinforcement that is provided by the inserts.

As more cellulosome-producing bacteria and their cohesins and dockerins are sequenced, our views on the sequence characteristics of these modules have changed. Although cohesin–dockerin specificity of a given type

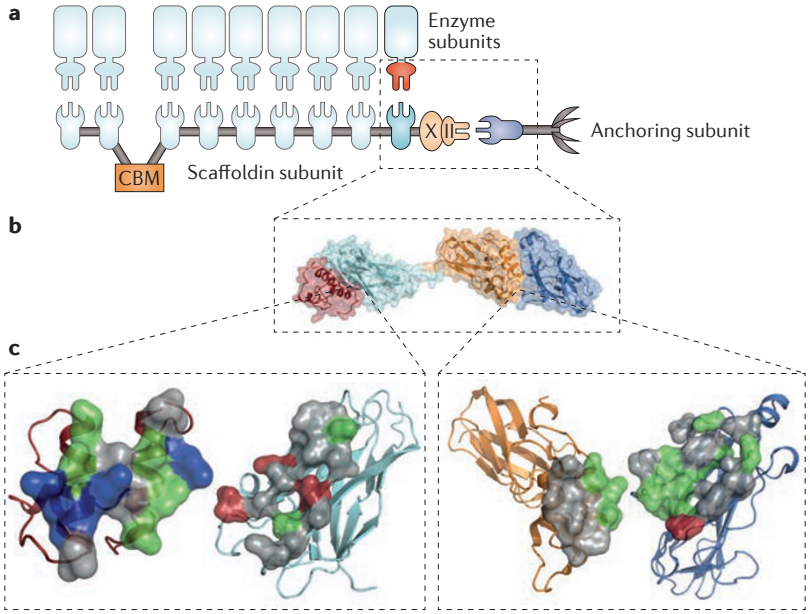

Figure 3 | **Crystal structure of a *Clostridium thermocellum* cellulosome fragment.** **a** | Schematic representation of a basic cellulosome assembly in *Clostridium thermocellum* that comprises the primary scaffoldin, a monovalent anchoring scaffoldin and nine dockerin-containing enzyme subunits. **b** | The cellulosome fragment includes the type II cohesin module from the anchoring protein scaffoldin dockerin binding protein A (SdbA; blue), a tri-modular portion of the scaffoldin subunit, comprising the ninth type I cohesin (cyan) and X module–type II dockerin pair (orange), and an enzyme-derived type I dockerin module (red). **c** | Structural depiction of the type I (left) and type II (right) interfaces between the cohesin and dockerin modules. The surfaces of interface residues are shown and amino acid residues are colour-coded red for acidic, blue for basic, green for polar and grey for hydrophobic. The character of the type II interaction is clearly more hydrophobic, as indicated by the dominance of grey areas, than the type I interface, which also presents extensive hydrophilic and electrostatic interactions. CBM, carbohydrate-binding module.

is usually preserved within a species, there are exceptions. Similarly, specificity between species is not always observed. Broad interspecies recognition is common among the simple cellulosome systems of mesophilic clostridia. Similarly, some cross-species overlap has been observed with type II interactions[37].

The success at determining the structures of individual cohesins and dockerins and their combined complexes has generated ambitious attempts to crystallize larger portions of cellulosomal components. However, these efforts have proved problematic, and only isolated crystal structures of larger cellulosome fragments have been described, the most extensive of which included three different proteins comprising five separate modules[74] (FIG. 3).

Carbohydrate-binding modules. The major scaffoldin-borne CBM positions the cellulosomal enzymes, and presumably the bacterial cell itself, upon the cellulosic substrate. Primary scaffoldins generally contain a family 3 CBM (CBM3) that binds selectively to cellulose[75] and has a crucial role in its degradation[31,75,76].

Recently, a putative cell-free scaffoldin, ScaM, was found to be a part of the cellulosome system of *A. cellulolyticus*, which contains three type I cohesins and two family 2 CBMs[14]. Similar CBM2-containing scaffoldins were discovered in the genome of *C. clariflavum*[16]. These scaffoldins are the only known examples that contain CBMs from a family other than CBM3. Family 2 CBMs are usually associated with free, non-cellulosomal enzymes and are divided into two subfamilies, one of which binds to cellulose, whereas the other binds to xylan[77]. Family 2 CBMs are of interest, as they seem to be present only on cell-free scaffoldins and not on cell-anchored scaffoldins, which suggests that they have a different role to family 3 CBMs.

CBMs are important for cellulosome function, but not only as a part of the scaffoldin protein. Although the known scaffoldins contain CBMs that belong only to two families, many cellulosomal enzymes have CBMs from various families that exhibit different carbohydrate specificities. The enzyme-borne CBMs are thought to position the catalytic modules for optimal hydrolysis[78].

Regulation of cellulosomal components

The number of genes that encode dockerin-containing enzymes is much larger than the number of cohesins that are available on the primary scaffoldin, which suggests a fine-tuned regulation of the enzymatic components that is dependent on the available carbon source. The carbon source is an important factor that determines the composition of a cellulosome, in regard to both its integrated enzymatic subunits and overall architecture, as revealed through proteomics studies of several cellulosomal bacteria that were grown on different growth media[46,79–81]. For example, the quantity of the GH48 exoglucanase was reported to increase when a bacterium is grown on cellulose compared with cellobiose[46,82]. A similar pattern could be observed for other cellulases, whereas other glycoside hydrolases showed the opposite response. In addition, following cultivation on hemicellulosic substrates, the expression levels of cellulosomal enzymes changed compared with growth on cellulose.

Cellulosomal genes are expressed at higher levels in the presence of complex natural polymers than in the presence of simple oligosaccharides, and, at high growth rate, cellulosomal genes are downregulated[83]. This implies a mechanism of substrate sensing that coordinates the expression of relevant cellulosomal genes. In *C. thermocellum* some cellulosomal genes are regulated by several anti-sigma factors and alternative sigma factors[84–86]. In the absence of substrate, the transmembrane anti-sigma factor is attached to an alternative sigma factor in the cell. Each anti-sigma factor also has an exocellular CBM-like component that contacts the polysaccharide substrate in the external medium. Binding of an appropriate substrate to this CBM results in a conformational change in the anti-sigma factor that releases the alternative sigma factor, which then interacts with RNA polymerase, thereby initiating transcription of cellulosomal genes.

Similar gene pairs that encode sigma factors and anti-sigma factors are present in related bacteria that produce highly structured cellulosomes, such as *C. clariflavum, A. cellulolyticus* and *B. cellulosolvens*[1], as well as in non-cellulosomal bacteria[87]. This sophisticated system enables the sensing of the status of the plant

Box 2 | **Minicellulosomes and designer cellulosomes**

Minicellulosomes comprise a truncated scaffoldin, in which cohesin specificity is very similar or identical to the native protein, and the precise positional integration of dockerin-containing enzymes cannot be achieved. This enables the incorporation of multiple copies of a single enzyme or the uncontrolled (but not necessarily random[155]) incorporation of multiple enzymes. By contrast, artificial designer cellulosomes are characterized by a chimeric scaffoldin that contains cohesins with divergent specificity and originates from different bacterial species, thus enabling the specific and controlled incorporation of desired enzymes.

Studies using artificial cellulosomes have shown that complexed enzymes are more efficient in substrate degradation than mixtures of free enzymes[156,157]. Enzyme targeting to the substrate through the scaffoldin-borne carbohydrate-binding module (CBM), the physical proximity effect of the enzymatic components and synergy among different glycoside hydrolases that are specific for different parts of the lignocellulose substrate[158], were defined as major factors for cellulosome efficiency[159].

Designer cellulosomes have been used as tools for studying and understanding the structure–function relationships of native cellulosomes. In this context, the role of linker length between the different modules has been studied. The length of intermodular linkers is essential to avoid steric hindrance[160] and to allow flexibility[161], but is of relatively minor importance for the enzymes[162]. The modular architecture of cellulosomes has also been investigated. The overall flexibility of the scaffoldin enables efficient degradation in conventional cellulosomes[163], and a negative influence of multiple CBMs in the cellulosome was observed. The position of a specific enzyme relative to the others could also have a role in specific cases[164]. The position of the dockerin in the chimeric enzymes also seems to be an important parameter[162].

In many of these studies, the simple but potent free enzymatic system of the non-cellulosomal bacterium *Thermobifida fusca* was converted to the cellulosomal mode by fusing dockerin modules to the enzymes. Unlike cellulosome-producing bacteria, the *T. fusca* cellulase system contains a limited number of well-characterized cellulases (only seven) and other glycoside hydrolases, thus simplifying the decision of which enzymes to study. This system has since acted as an excellent model for the comparison of free-enzyme systems with engineered designer cellulosomes[157,162].

Very recently, the apparent limitation of designer cellulosomes in terms of scaffoldin size[157] was overcome through the use of adaptor scaffoldins, which extend the platform scaffoldin and enable the integration of more enzymes (FIG. 4). Octavalent designer cellulosomes were thus produced that were 70% as efficient in degrading wheat straw than natural *C. thermocellum* cellulosomes[92]. These artificial designer cellulosomes, which are comparable in size to natural cellulosomes, were correctly assembled, and evidence for their stability was documented. This strategy paves the way for increased enzyme numbers and types within a single cellulosomal complex, in which all components are inserted in a controlled and precise manner.

wall polysaccharides in the extracellular milieu during the degradation process and regulates the production of the cellulosomal enzymes that are necessary to respond to newly exposed substrates and newly produced substrate intermediates.

In *C. cellulolyticum*, carbon catabolite repression and a two-component system regulate the core and accessory enzymes in cellulosomes[88]. This mechanism resembles the sigma factor–anti-sigma factor mechanism in the ability of the available substrate to determine gene expression and thereby optimize cellulosome function. Recently, a new mechanism of transcriptional regulation of cellulosomes has been reported in *C. cellulolyticum*[89]. Selective RNA processing and stabilization were reported to be responsible for regulating the stoichiometry of transcripts that encode the cellulosomal components in the *cip-cel* operon. This suggests substrate-independent regulation of cellulosome composition that would complement the action of specific carbon substrates at the transcriptional level by producing pre-optimized cellulosomes. This memory mechanism at the RNA level would certainly be crucial for correct cellulosome function and could account for the respective adaptation and competition strategies in cellulolytic communities.

Carbon catabolite repression
A regulatory mechanism that, in the presence of a preferred substrate, prevents the expression of genes that are required for the utilization of secondary sources of carbon.

Two-component system
A signal transduction pathway that involves a sensor in the extracellular environment and an intracellular response regulator.

Applications of cellulosomes

Minicellulosomes and designer cellulosomes. Artificial complexes that mimic cellulosomes were proposed more than two decades ago[11], and since, they have been produced extensively, both *in vitro* and *in vivo*. Minicellulosomes and designer cellulosomes have been used both as tools for the study of cellulosome action and as potential replacements for, or extensions of, native cellulosomes for nanobiotechnological applications, notably for the production of biofuels from cellulosic biomass (BOX 2).

To broaden cellulosome diversity and increase substrate degradation, 'external' enzymes, such as β-glucosidases, lytic polysaccharide monooxygenases (LPMOs) or expansins, have been incorporated into designer cellulosomes. This incorporation complemented the complex with novel enzymatic activities that generally resulted in an enhancement of overall activity[50,59,90]. Integration of a laccase increased cellulase activity, thereby paving the way for the combined cellulosome-mediated degradation of both lignin and cellulose[91]. In addition, adaptor scaffoldins have been used to amplify the number of enzymes that can be incorporated into designer cellulosome complexes[92] (FIG. 4).

Cellulosome architecture has also inspired the design of other complexes, including self-assembled 12-enzyme and 18-enzyme complexes[93,94], cellulases that are covalently bound to nanospheres[95], cellulases that are bound to streptavidin and inorganic particles[96], and cellulases that are bound to a DNA scaffold[97]. These studies increased the number of enzymes in a single complex, demonstrating that the value of the proximity effect in cellulosomes can be transferred to novel platforms. Nevertheless, other important parameters, such as structural flexibility, substrate targeting and/or the potential control of enzyme composition, could not be reproduced in these complexes.

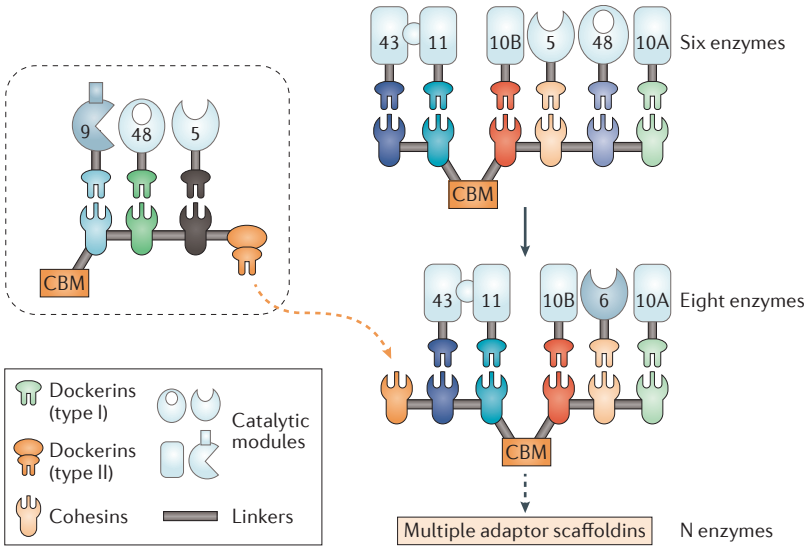

Figure 4 | **State-of-the-art designer cellulosome technology.** The limited size of recombinant designer cellulosomes[157] can be overcome through the use of an adaptor scaffoldin in combination with the sequential incorporation of enzymes. The incorporation of two additional enzymes, the family 6 and family 9 glycoside hydrolases (coloured), was thus achieved[92]. This strategy can be further extended to enable the inclusion of additional numbers and types of enzymes into the cellulosome complex. Cohesin–dockerin interactions are colour-coded in the figure, with each pair having a different specificity. In the shorthand notation for the enzymes, the numbers, 5, 6, 9, 10, 11, 43 and 48 refer to the respective glycoside hydrolase family of the catalytic modules. CBM, carbohydrate-binding module.

Recombinant designer cellulosome-producing microorganisms. Minicellulosomes and designer cellulosomes, which are displayed on microbial cell surfaces, have been considered for use in industrial applications, notably for the production of biofuels from plant-derived biomass. In this context, the development of consolidated bioprocessing (CBP) has been proposed[98], in which a single microbial culture, co-culture or mixed culture is used to produce biofuels directly from lignocellulosic biomass in a single bioreactor. Several microorganisms have been engineered to display minicellulosomes or designer cellulosomes on their cell surface for the enzymatic conversion of lignocellulose to sugar and lactic acid. These organisms include *Bacillus subtilis*[99], *Pichia pastoris*[100] and *Lactobacillus plantarum*[101]. In most cases, clustering of the enzymes in cellulosome complexes resulted in more efficient degradation of plant cell wall substrates. An additional benefit of surface-displayed cellulosomes is the possible recovery of enzymes by centrifugation of the cells and their recycling as biocatalysts, which might decrease costs[102].

One example of CBP was through the use of *Clostridium acetobutylicum*, which was supplemented with a functional cellulosome for production of acetone or butanol[103]. Similarly, *Saccharomyces cerevisiae*[104,105] was engineered to display cellulosomes for the direct production of ethanol from biomass. Likewise, *Corynebacterium glutamicum* was transformed with a mini-cellulosome for the production of amino acids[106]. The cellulosome-lacking, butanol producer *Clostridium beijerinckii* has also been proposed for the integration of cellulosomal genes[107].

Taken together, these studies demonstrate that cellulosomal components are active and can be functionally assembled in a more manageable foreign organism that would then be more suitable for the production of valuable products from biomass waste.

Designer cellulosome technology can also be used with alternative metabolic pathways for the enhancement of reaction times in enzymatic cascades. Substrate channelling in such enzymatic complexes can increase the efficiency of a reaction[108], in particular by accelerating rate-limiting steps that are controlled by enzymes that have low activities[109].

Thermostable cellulosomes. Thermophilic microorganisms and enzymatic processes are very attractive for industrial applications. Their advantages include increased stability during extended periods of time, increased reaction rates, increased process flexibility, higher rates of diffusion, higher substrate solubility and lower contamination risks[110,111]. In addition, the high growth temperature could facilitate the recovery of valuable products, such as ethanol[112].

Two recent studies[113,114] investigated the stability of designer cellulosomes at increased temperatures (>50 °C) *in vitro* and showed increased hydrolytic performance of thermophilic cellulosomes compared with conventional designer cellulosomes[114]. However, an ideal thermophilic microorganism that can naturally convert plant biomass into biofuels has not been identified. Engineering a cellulosome-producing thermophilic microorganism for the production of biofuels is the preferred strategy for the development of thermophilic CBP, as engineering a solventogenic thermophile to hydrolyse lignocellulose requires the introduction of more enzymes to achieve complete biomass degradation[110,112].

Two candidates for thermophilic CBP are *C. thermocellum* and *C. clariflavum*, which are naturally ethanologenic and can be adapted to tolerate relatively high concentrations of ethanol[115,116]. It has also been proposed to engineer the metabolic pathways[98] of *C. thermocellum* for the production of isobutanol[117] and ethanol[118]. In parallel, co-culture of *C. thermocellum* with a solventogenic bacterium is also being extensively explored, as it might decrease production costs for the conversion of lignocellulose into biofuels[119]. The lack of pentose metabolism in *C. thermocellum*[120] can be remedied by concomitant growth with an appropriate thermophile, such as *Thermoanaerobacter* spp.[110]. *C. clariflavum*, as a relatively newly described cellulosomal species, has not been used for the production of solvents. Nevertheless, *C. clariflavum* strain 4-2a, with its ability to utilize pentoses[121], is a particularly attractive candidate for CBP. However, some issues remain problematic for the effective application of CBP, such as enzyme inactivation by alcohol and solvents[122], and the costs of culturing under anaerobic conditions[123].

Recent advances in genetic engineering of these particularly recalcitrant anaerobic bacteria and the development of genetic tools such as targetrons, which are gene-targeting vectors that are derived from

mobile group II introns, have been demonstrated in *C. thermocellum*[124]. In addition, CRISPR–Cas9, which has already been used in the mesophile *C. cellulolyticum*[125], could be adapted to thermophilic anaerobes.

Additional applications of cellulosomal components in biotechnology. Owing to the modular nature of cellulosomes, cellulosomal components have been proposed for use in many biotechnological applications[4], especially together with other affinity systems (such as protein A, antibodies and lectins). These hybrid biomolecules could be used for various purposes, such as immunoassays and blotting, microarray technology, drug delivery, localization and cytochemistry, isolation and immobilization, affinity chromatography, and cell separation[11]. Indeed, the high-fidelity, high-affinity cohesin–dockerin interaction[69] could be used as a partner in other affinity-based applications, such as those that involve avidin–biotin[126]. For example, a biosensor that is based on the cohesin–dockerin interaction was recently developed[127]. Nevertheless, only a few of the many possible research applications have been explored, and there is a large potential for future innovation.

A major limitation in some of these techniques is the near-irreversible cohesin–dockerin binding[63,128], with dissociation occurring only in the presence of denaturants and at high temperatures. Nevertheless, engineering of the *C. thermocellum* dockerin has decreased its affinity for cohesin, which enabled its use as an affinity tag for protein purification[129,130].

In contrast to the cohesin–dockerin interaction, the affinity of CBMs for different polysaccharides has been explored more extensively[131]. Fusion proteins that contain a CBM have been used as bioreactors[132], plant growth modulators[132], as affinity tags for protein purification[133], microarray tools for the study of protein interactions[37], biosensors[134] and to study cellulosic substrates[135,136]. In addition, an array of quantum dots using cellulosomal components has been developed[137].

Outlook

When cellulosomes were discovered more than 30 years ago, we initially expected that cellulosome-producing bacteria would be prevalent in nature. However, the following years have demonstrated that cellulosome-producing bacterial species are specialized and rare, but they clearly have a pivotal role in processing intractable polysaccharides that are derived from plant cell walls in lignocellulosic ecosystems. Since their discovery, cellulosomes have been identified, owing to genome sequencing, in specialized anaerobic bacteria. Only relatively few species have evolved to produce cellulosomes, but they have a key polymer-degrading function in environments characterized by decomposing plant matter.

The development of metagenomics will probably identify additional cellulosome-producing bacteria. Structures of individual cellulosomal components have been extensively investigated, as well as the lignocellulosic enzymatic activities that can be incorporated into these complexes. The composition of cellulosomes is dependent on the carbon source and other regulatory factors, and its diverse nature within given cellulosome systems has been investigated by transcriptomic and proteomic studies.

Future research should be dedicated to the function of cellulosomes, their role in natural microbiomes and their importance in a given ecosystem. In parallel, studies should continue to focus on further discovery of novel cellulosome-producing bacteria from various ecosystems, including terrestrial, aquatic and marine ecosystems, and vertebrate and invertebrate microbiota. There is a need for investigations into cellulosome secretion and assembly processes, and comparative cellulosome regulation in various species.

The high-resolution structural determination of a complete cellulosome seems a romantic but impracticable goal[138]. Global structural analysis of complete cellulosomes has been hindered by their inherent compositional heterogeneity. Heterogeneity in enzyme content, the random incorporation of subunits into the complex, glycosylation of the elongated linker segments, alternative conformations that reflect the dual-binding mode, the general dynamic nature of the complex and its enormous size, would all argue against its crystallization for X-ray structural determination. Thus, crystal structures of a complete native complex seem to be an implausible dream, and other approaches of lower resolution and/or reliability, such as electron microscopy, small-angle X-ray scattering[78] and advanced computational techniques[139], will probably have to suffice[68].

In addition, further exploration into the functions and roles of non-cellulosomal cohesin-containing and dockerin-containing proteins is warranted. It is intriguing to consider the various functions of these complementary high-affinity protein modules in nature. They are produced almost exclusively by archaea and bacteria. Their role in cellulosome structure and function is clear, but what other roles do they have in non-cellulosomal systems?

Finally, designer cellulosomes have proven an exceptional tool for studying the structure and function of cellulosomes. They can potentially be used for various applications, the most challenging of which would be the production of biofuels by CBP. In addition, their applications in areas other than the conventional degradation of biomass, such as the enhancement of metabolic pathways and the enhancement or extension of affinity interactions, should prove valuable. Future research will continue to broaden both the basic and applied aspects of cellulosome-based technology.

1. Bayer, E. A., Shoham, Y. & Lamed, R. in *The Prokaryotes. Prokaryotic Physiology and Biochemistry* (eds Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. & Thomson, F.) 215–265 (Springer, 2013).
A comprehensive chapter that details cellulosomes and non-cellulosomal cellulase systems, and the bacteria that produce them.
2. Wei, H. *et al.* Natural paradigms of plant cell wall degradation. *Curr. Opin. Biotechnol.* **20**, 330–338 (2009).
3. Albersheim, P., Darvill, A., Roberts, K., Sederoff, R. & Staehelin, A. *Plant Cell Walls: From Chemistry to Biology* (Garland Science, 2011).
4. Bayer, E. A., Belaich, J.-P., Shoham, Y. & Lamed, R. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* **58**, 521–554 (2004).
5. Lamed, R., Setter, E. & Bayer, E. A. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *J. Bacteriol.* **156**, 828–836 (1983).
Together with reference 6, this article represents the first study in which the cellulosome was discovered.

6. Bayer, E. A., Kenig, R. & Lamed, R. Adherence of *Clostridium thermocellum* to cellulose. *J. Bacteriol.* **156**, 818–827 (1983).
7. Himmel, M. E. *et al.* Microbial enzyme systems for biomass conversion: emerging paradigms. *Biofuels* **1**, 323–341 (2010).
8. Doi, R. H. & Kosugi, A. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat. Rev. Microbiol.* **2**, 541–551 (2004).
9. Ding, S.-Y. *et al.* How does plant cell wall nanoscale architecture correlate with enzymatic digestibility? *Science* **338**, 1055–1060 (2012).
In this study, advanced real-time imaging techniques demonstrate that free fungal cellulases degrade cell walls by mechanisms that differ considerably from those of cellulosomes.
10. Resch, M. G. *et al.* Fungal cellulases and complexed cellulosomal enzymes exhibit synergistic mechanisms in cellulose deconstruction. *Energy Environ. Sci.* **6**, 1858–1867 (2013).
The combination of a free-enzyme system and purified cellulosomes was found to act with exceptionally high synergy on cellulosic substrates.
11. Bayer, E. A., Morag, E. & Lamed, R. The cellulosome — a treasure-trove for biotechnology. *Trends Biotechnol.* **12**, 379–386 (1994).
This is the first paper to propose designer cellulosomes and to define the terms cohesin, dockerin and scaffoldin.
12. Hamberg, Y. *et al.* Elaborate cellulosome architecture of *Acetivibrio cellulolyticus* revealed by selective screening of cohesin–dockerin interactions. *PeerJ* **2**, e636 (2014).
13. Xu, Q. *et al.* Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. *Sci. Adv.* **2**, e1501254 (2016).
14. Dassa, B. *et al.* Genome-wide analysis of *Acetivibrio cellulolyticus* provides a blueprint of an elaborate cellulosome system. *BMC Genomics* **13**, 210 (2012).
15. Xu, Q. *et al.* The cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell surface anchoring protein. *J. Bacteriol.* **185**, 4548–4557 (2003).
16. Artzi, L. *et al.* Cellulosomics of the cellulolytic thermophile *Clostridium clariflavum*. *Biotechnol. Biofuels* **7**, 100 (2014).
17. Rincon, M. T. *et al.* Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD-1. *PLoS ONE* **5**, e12476 (2010).
18. Chassard, C., Delmas, E., Robert, C. & Bernalier-Donadille, A. The cellulose-degrading microbial community of the human gut varies according to the presence or absence of methanogens. *FEMS Microbiol. Ecol.* **74**, 205–213 (2010).
19. Ben David, Y. *et al.* Ruminococcal cellulosome systems from rumen to human. *Environ. Microbiol.* **17**, 3407–3426 (2015).
20. Cann, I., Bernardi, R. C. & Mackie, R. I. Cellulose degradation in the human gut: *Ruminococcus champanellensis* expands the cellulosome paradigm. *Environ. Microbiol.* **18**, 307–310 (2016).
21. Michel, G. Ruminococcal cellulosomes: molecular Lego to deconstruct microcrystalline cellulose in human gut. *Environ. Microbiol.* **17**, 3113–3115 (2015).
22. Moraïs, S. *et al.* 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 (2016).
23. Pagès, S. *et al.* Sequence analysis of scaffolding protein CipC and ORFXp, a new cohesin-containing protein in *Clostridium cellulolyticum*: comparison of various cohesin domains and subcellular localization of ORFXp. *J. Bacteriol.* **181**, 1801–1810 (1999).
24. Shoseyov, O., Takagi, M., Goldstein, M. A. & Doi, R. H. Primary sequence analysis of *Clostridium cellulovorans* cellulose binding protein A. *Proc. Natl Acad. Sci. USA* **89**, 3483–3487 (1992).
25. Pohlschröder, M., Leschine, S. B. & Canale-Parola, E. Multicomplex cellulase–xylanase system of *Clostridium papyrosolvens* C7. *J. Bacteriol.* **176**, 70–76 (1994).
26. Ze, X. *et al.* Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic Firmicutes bacterium *Ruminococcus bromii*. *mBio* **6**, e01058-15 (2015).
27. Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* **6**, 1535–1543 (2012).
28. Dassa, B. *et al.* Rumen cellulosomics: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. *PLoS ONE* **9**, e99221 (2014).
29. Gilmore, S. P., Henske, J. K. & O'Malley, M. Driving biomass breakdown through engineered cellulosomes. *Bioengineered* **6**, 204–208 (2015).
30. Ding, S. Y., Bayer, E. A., Steiner, D., Shoham, Y. & Lamed, R. A novel cellulosomal scaffoldin from *Acetivibrio cellulolyticus* that contains a family 9 glycosyl hydrolase. *J. Bacteriol.* **181**, 6720–6729 (1999).
31. Gerngross, U. T., Romaniec, M. P. M., Kobayashi, T., Huskisson, N. S. & Demain, A. L. Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal S_L-protein reveals an unusual degree of internal homology. *Mol. Microbiol.* **8**, 325–334 (1993).
32. Fujino, T., Béguin, P. & Aubert, J. P. 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 (1993).
33. Izquierdo, J. A. *et al.* Complete genome sequence of *Clostridium clariflavum* DSM 19732. *Stand. Genomic Sci.* **6**, 104–115 (2012).
34. Bensoussan, L. *et al.* Broad phylogeny and functionality of cellulosomal components in the bovine rumen microbiome. *Environ. Microbiol.* http://dx.doi.org/10.1111/1462-2920.13561 (2016).
This study details the ultra-deep sequencing of the fibre-adherent rumen microbiome, which reveals that the cellulosomal machinery is conserved and widely used. Dockerin-containing proteins are not restricted to fibre degradation per se and mediate other catabolic processes as well as microbial interactions.
35. Bayer, E. A., Coutinho, P. M. & Henrissat, B. Cellulosome-like sequences in *Archaeoglobus fulgidus*: an enigmatic vestige of cohesin and dockerin domains. *FEBS Lett.* **463**, 277–280 (1999).
36. Peer, A., Smith, S. P., Bayer, E. A., Lamed, R. & Borovok, I. Noncellulosomal cohesin- and dockerin-like modules in the three domains of life. *FEMS Microbiol. Lett.* **291**, 1–16 (2009).
37. Haimovitz, R. *et al.* Cohesin–dockerin microarray: diverse specificities between two complementary families of interacting protein modules. *Proteomics* **8**, 968–979 (2008).
38. Adams, J. J., Gregg, K., Bayer, E. A., Boraston, A. B. & Smith, S. P. Structural basis of *Clostridium perfringens* toxin complex formation. *Proc. Natl Acad. Sci. USA* **105**, 12194–12199 (2008).
39. Zverlov, V. V., Klupp, M., Krauss, J. & Schwarz, W. H. Mutations in the scaffoldin gene, *cipA*, of *Clostridium thermocellum* with impaired cellulosome formation and cellulose hydrolysis: insertions of a new transposable element, IS*1447*, and implications for cellulase synergism on crystalline cellulose. *J. Bacteriol.* **190**, 4321–4327 (2008).
This study shows that mutants of *Clostridium thermocellum* that lack the major scaffoldin gene contain the enzymes in free form, and the mutated bacteria exhibit decreased hydrolytic levels on crystalline cellulose substrates.
40. Desvaux, M. *Clostridium cellulolyticum*: model organism of mesophilic cellulolytic clostridia. *FEMS Microbiol. Rev.* **29**, 741–764 (2005).
41. Kosugi, A., Murashima, K., Tamaru, Y. & Doi, R. H. Cell-surface-anchoring role of N-terminal surface layer homology domains of *Clostridium cellulovorans* EngE. *J. Bacteriol.* **184**, 884–888 (2002).
42. Rincon, M. T. *et al.* Unconventional mode of attachment of the *Ruminococcus flavefaciens* cellulosome to the cell surface. *J. Bacteriol.* **187**, 7569–7578 (2005).
43. Lemaire, M., Ohayon, H., Gounon, P., Fujino, T. & Béguin, P. 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 (1995).
44. Rincón, M. T. *et al.* 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 (2004).
45. Jindou, S. *et al.* Conservation and divergence in cellulosome architecture between two strains of *Ruminococcus flavefaciens*. *J. Bacteriol.* **188**, 7971–7976 (2006).
46. Artzi, L., Morag, E., Barak, Y., Lamed, R. & Bayer, E. A. *Clostridium clariflavum*: key cellulosome players are revealed by proteomic analysis. *mBio* **6**, e00411-15 (2015).
47. Bayer, E. A. & Lamed, R. Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *J. Bacteriol.* **167**, 828–836 (1986).
48. Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495 (2014).
49. Eriksson, T., Karlsson, J. & Tjerneld, F. A model explaining declining rate in hydrolysis of lignocellulose substrates with cellobiohydrolase I (Cel7A) and endoglucanase I (Cel7B) of *Trichoderma reesei*. *Appl. Biochem. Biotechnol.* **101**, 41–60 (2002).
50. Gefen, G., Anbar, M., Morag, E., Lamed, R. & Bayer, E. A. 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 (2012).
51. Béguin, P., Aubert, J.-P. & Beguin, P. The biological degradation of cellulose. *FEMS Microbiol. Rev.* **13**, 25–58 (1994).
52. Xu, Q., Luo, Y., Ding, S. & Himmel, M. E. Multifunctional enzyme systems for plant cell wall degradation. *Compr. Biotechnol.* **3**, 15–25 (2011).
53. Morag, E., Halevy, I., Bayer, E. A. & Lamed, R. Isolation and properties of a major cellobiohydrolase from the cellulosome of *Clostridium thermocellum*. *J. Bacteriol.* **173**, 4155–4162 (1991).
54. Ravachol, J. *et al.* Combining free and aggregated cellulolytic systems in the cellulosome-producing bacterium *Ruminiclostridium cellulolyticum*. *Biotechnol. Biofuels* **8**, 114 (2015).
The study details the transformation of a cellulase gene and its subsequent expression by a cellulosome-producing bacterium. Its integration into the cellulosome induced the release of regular cellulosomal components, notably the exoglucanase Cel48F, thereby impairing the activity of the complex on crystalline cellulose.
55. Ravachol, J., Borne, R., Tardif, C., de Philip, P. & Fierobe, H.-P. Characterization of all family-9 glycoside hydrolases synthesized by the cellulosome-producing bacterium *Clostridium cellulolyticum*. *J. Biol. Chem.* **289**, 7335–7348 (2014).
56. Kang, S., Barak, Y., Lamed, R., Bayer, E. A. & Morrison, M. The functional repertoire of prokaryote cellulosomes includes the serpin superfamily of serine proteinase inhibitors. *Mol. Microbiol.* **60**, 1344–1354 (2006).
57. Levy-Assaraf, M. *et al.* Crystal structure of an uncommon cellulosome-related protein module from *Ruminococcus flavefaciens* that resembles papain-like cysteine peptidases. *PLoS ONE* **8**, e56138 (2013).
58. Artzi, L., Morag, E., Shamshoum, M. & Bayer, E. A. Cellulosomal expansin: functionality and incorporation into the complex. *Biotechnol. Biofuels* **9**, 61 (2016).
59. Chen, C. *et al.* Integration of bacterial expansin-like proteins into cellulosome promotes the cellulose degradation. *Appl. Microbiol. Biotechnol.* **100**, 2203–2212 (2016).
60. Stahl, S. W. *et al.* Single-molecule dissection of the high-affinity cohesin–dockerin complex. *Proc. Natl Acad. Sci. USA* **109**, 20431–20436 (2012).
61. Gunnoo, M. *et al.* Nanoscale engineering of designer cellulosomes. *Adv. Mater.* **28**, 5619–5647 (2016).
62. Valbuena, A. *et al.* On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc. Natl Acad. Sci. USA* **106**, 13791–13796 (2009).
63. Bhat, K. & Wood, T. M. The cellulase of the anaerobic bacterium *Clostridium thermocellum*: isolation, dissociation, and reassociation of the cellulosome. *Carbohydr. Res.* **227**, 293–300 (1992).
64. Schoeler, C. *et al.* Ultrastable cellulosome–adhesion complex tightens under load. *Nat. Commun.* **5**, 5635 (2014).
65. Schoeler, C. *et al.* Mapping mechanical force propagation through biomolecular complexes. *Nano Lett.* **15**, 7370–7376 (2015).
66. Xu, Q. *et al.* Architecture of the *Bacteroides cellulosolvens* cellulosome: description of a cell surface-anchoring scaffoldin and a family 48 cellulase. *J. Bacteriol.* **186**, 968–977 (2004).

67. Carvalho, A. L. *et al.* Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex. *Proc. Natl Acad. Sci. USA* **100**, 13809–13814 (2003). **This study presents, for the first time, the crystal structure of a cohesin–dockerin pair and suggests a dual mode of binding.**
68. Smith, S. P. & Bayer, E. A. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr. Opin. Struct. Biol.* **23**, 686–694 (2013).
69. Adams, J. J., Pal, G., Jia, Z. & Smith, S. P. 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 (2006).
70. Salama-Alber, O. *et al.* Atypical cohesin–dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J. Biol. Chem.* **288**, 16827–16838 (2013).
71. Pagès, S. *et al.* 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 (1997).
72. Chen, C. *et al.* Revisiting the NMR solution structure of the Cel48S type-I dockerin module from *Clostridium thermocellum* reveals a cohesin-primed conformation. *J. Struct. Biol.* **188**, 188–193 (2014).
73. Nash, M. A., Smith, S. P., Fontes, C. & Bayer, E. A. Single versus dual-binding conformations in cellulosomal cohesin–dockerin complexes. *Curr. Opin. Struct. Biol.* **40**, 89–96 (2016).
74. Currie, M. A. *et al.* Scaffoldin conformation and dynamics revealed by a ternary complex from the *Clostridium thermocellum*. *J. Biol. Chem.* **287**, 26953–26961 (2012).
75. Morag, E. *et al.* 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 (1995).
76. Poole, D. M. *et al.* Identification of the cellulose binding domain of the cellulosome subunit S1 from *Clostridium thermocellum*. *FEMS Microbiol. Lett.* **99**, 181–186 (1992).
77. Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J. & Williamson, M. P. The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *J. Biol. Chem.* **275**, 41137–41142 (2000).
78. Hammel, M. *et al.* Structural basis of cellulosome efficiency explored by small angle X-ray scattering. *J. Biol. Chem.* **280**, 38562–38568 (2005).
79. Han, S. O., Yukawa, H., Inui, M. & Doi, R. H. Regulation of expression of cellulosomal cellulase and hemicellulase genes in *Clostridium cellulovorans*. *J. Bacteriol.* **185**, 6067–6075 (2003).
80. Dror, T. W., Rolider, A., Bayer, E. A., Lamed, R. & Shoham, Y. Regulation of expression of scaffoldin-related genes in *Clostridium* thermocellum. *J. Bacteriol.* **185**, 5109–5116 (2003).
81. Raman, B. *et al.* Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PLoS ONE* **4**, e5271 (2009).
82. Dykstra, A. B. *et al.* 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 (2014).
83. Dror, T. W. *et al.* Regulation of the cellulosomal *celS* (Cel48A) gene of *Clostridium thermocellum* is growth-rate dependent. *J. Bacteriol.* **185**, 3042–3048 (2003).
84. Muñoz-Gutiérrez, I. *et al.* Decoding biomass-sensing regulons of *Clostridium thermocellum* alternative sigma-I factors in a heterologous *Bacillus subtilis* host system. *PLoS ONE* **11**, e0146316 (2015).
85. Kahel-Raifer, H. *et al.* The unique set of putative membrane-associated anti-σ factors in *Clostridium thermocellum* suggests a novel extracellular carbohydrate-sensing mechanism involved in gene regulation. *FEMS Microbiol. Lett.* **308**, 84–93 (2010).
86. Nataf, Y. *et al. Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proc. Natl Acad. Sci. USA* **107**, 18646–18651 (2010). **Together with reference 85, this article proposes an extensive regulatory system that enables cellulosome-producing bacteria to sense the status of impermeant polysaccharides in the external milieu.**
87. Xu, J. *et al.* A genomic view of the human–*Bacteroides thetaiotaomicron* symbiosis. *Science* **299**, 2074–2076 (2003).
88. Xu, C. *et al.* Structure and regulation of the cellulose degradome in *Clostridium cellulolyticum*. *Biotechnol. Biofuels* **6**, 73 (2013).
89. Xu, C. *et al.* Cellulosome stoichiometry in *Clostridium cellulolyticum* is regulated by selective RNA processing and stabilization. *Nat. Commun.* **6**, 6900 (2015).
90. Arfi, Y., Shamshoum, M., Rogachev, I., Peleg, Y. & Bayer, E. A. Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation. *Proc. Natl Acad. Sci. USA* **111**, 9109–9114 (2014).
91. Davidi, L. *et al.* Toward combined delignification and saccharification of wheat straw by a laccase-containing designer cellulosome. *Proc. Natl Acad. Sci. USA* **113**, 10854–10859 (2016). **This study shows the integration of a lignin-active enzyme into a designer cellulosome, which leads to the enhanced degradation of plant-derived cellulose and xylan. Together with reference 90, this article shows that oxidative enzymes can be incorporated into cellulosomes.**
92. Stern, J., Moraïs, S., Lamed, R. & Bayer, E. A. Adaptor scaffoldins: an original strategy for extended designer cellulosomes, inspired from nature. *mBio* **7**, e00083-16 (2016). **This study details a recent breakthrough in designer cellulosome technology, in which an adaptor scaffoldin was used to increase the number of enzymes that can be integrated into an artificial designer cellulosome.**
93. Moraïs, S. *et al.* Enhanced cellulose degradation by nano-complexed enzymes: synergism between a scaffold-linked exoglucanase and a free endoglucanase. *J. Biotechnol.* **147**, 205–211 (2010).
94. Mitsuzawa, S. *et al.* The rosettazyme: a synthetic cellulosome. *J. Biotechnol.* **143**, 139–144 (2009).
95. Blanchette, C., Lacayo, C. I., Fischer, N. O., Hwang, M. & Thelen, M. P. Enhanced cellulose degradation using cellulase–nanosphere complexes. *PLoS ONE* **7**, e42116 (2012).
96. Kim, D.-M. *et al.* A nanocluster design for the construction of artificial cellulosomes. *Catal. Sci. Technol.* **2**, 499 (2012).
97. Mori, Y. *et al.* Aligning an endoglucanase Cel5A from *Thermobifida fusca* on a DNA scaffold: potent design of an artificial cellulosome. *Chem. Commun. (Camb.).* **49**, 6971–6973 (2013).
98. Biswas, R., Zheng, T., Olson, D. G., Lynd, L. R. & Guss, A. M. Elimination of hydrogenase active site assembly blocks H_2 production and increases ethanol yield in *Clostridium thermocellum*. *Biotechnol. Biofuels* **8**, 20 (2015).
99. Anderson, T. D. *et al.* Assembly of minicellulosomes on the surface of *Bacillus subtilis*. *Appl. Environ. Microbiol.* **77**, 4849–4858 (2011).
100. Ou, J. & Cao, Y. Incorporation of *Nasutitermes takasagoensis* endoglucanase into cell surface-displayed minicellulosomes in *Pichia pastoris* X33. *J. Microbiol. Biotechnol.* **24**, 1178–1188 (2014).
101. Moraïs, S., Shterzer, N., Lamed, R., Bayer, E. A. & Mizrahi, I. A combined cell-consortium approach for lignocellulose degradation by specialized *Lactobacillus plantarum* cells. *Biotechnol. Biofuels* **7**, 112 (2014).
102. Garvey, M., Klose, H., Fischer, R., Lambertz, C. & Commandeur, U. Cellulases for biomass degradation: comparing recombinant cellulase expression platforms. *Trends Biotechnol.* **31** 581–593 (2013).
103. Willson, B. J. *et al.* Biotechnology for biofuels production of a functional cell wall-anchored minicellulosome by recombinant clostridium acetobutylicum ATCC 824. *Biotechnol. Biofuels* **9**, 109 (2016). **This study details recent advances in the conversion of the solventogenic bacterium *Clostridium acetobutylicum* to a cellulose degrader.**
104. Liang, Y., Si, T., Ang, E. L. & Zhao, H. Engineered pentafunctional minicellulosome for simultaneous saccharification and ethanol fermentation in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **80**, 6677–6684 (2014).
105. Fan, L. H. *et al.* Biotechnology for biofuels engineering yeast with bifunctional minicellulosome and cellodextrin pathway for co-utilization of cellulose-mixed sugars. *Biotechnol. Biofuels* **9**, 137 (2016).
106. Hyeon, J. E., Jeon, W. J., Whang, S. Y. & Han, S. O. Production of minicellulosomes for the enhanced hydrolysis of cellulosic substrates by recombinant *Corynebacterium glutamicum*. *Enzyme Microb. Technol.* **48**, 371–377 (2011).
107. Bayer, E. A., Lamed, R. & Himmel, M. E. The potential of cellulases and cellulosomes for cellulosic waste management. *Curr. Opin. Biotechnol.* **18**, 237–245 (2007).
108. You, C. & Zhang, Y. H. P. Self-assembly of synthetic metabolons through synthetic protein scaffolds: one-step purification, co-immobilization, and substrate channeling. *ACS Synth. Biol.* **2**, 102–110 (2013). **In this study, designer cellulosome-inspired complexes that contain a cascade of three enzymes were shown to exhibit strong substrate channelling and a consequent enhancement of reaction rates.**
109. You, C. & Zhang, Y.-H. P. Annexation of a high-activity enzyme in a synthetic three-enzyme complex greatly decreases the degree of substrate channeling. *ACS Synth. Biol.* **3**, 380–386 (2014).
110. Blumer-Schuette, S. E. *et al.* Thermophilic lignocellulose deconstruction. *FEMS Microbiol. Rev.* **38**, 393–448 (2014). **A detailed review of thermophilic enzymatic systems — both cellulosomal and non-cellulosomal.**
111. Bhalla, A. *et al.* Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresour. Technol.* **158**, 581–593 (2013).
112. Thomas, L., Joseph, A. & Gottumukkala, L. D. Xylanase and cellulase systems of *Clostridium sp.*: an insight on molecular approaches for strain improvement. *Bioresour. Technol.* **158**, 343–350 (2014).
113. Galanopoulou, A. P. *et al.* Insights into the functionality and stability of designer cellulosomes at elevated temperatures. *Appl. Microbiol. Biotechnol.* **100**, 8731–8743 (2016).
114. Moraïs, S. *et al.* Enhancement of cellulosome-mediated deconstruction of cellulose by improving enzyme thermostability. *Biotechnol. Biofuels.* **9**, 164 (2016).
115. Brown, S. D. *et al.* Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*. *Proc. Natl Acad. Sci. USA* **108**, 13752–13757 (2011).
116. Williams, T. I., Combs, J. C., Lynn, B. C. & Strobel, H. J. Proteomic profile changes in membranes of ethanol-tolerant *Clostridium thermocellum*. *Appl. Microbiol. Biotechnol.* **74**, 422–432 (2007). 117. Lin, P. P. *et al.* Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. *Metab. Eng.* **31**, 44–52 (2015).
118. Papanek, B., Biswas, R., Rydzak, T. & Guss, A. M. Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum*. *Metab. Eng.* **32**, 49–54 (2015).
119. Du, R., Li, S., Zhang, X., Fan, C. & Wang, L. Using a microorganism consortium for consolidated bioprocessing cellulosic ethanol production. *Biofuels* **2**, 569–575 (2011).
120. Wiegel, J., Mothershed, C. P. & Puls, J. Differences in xylan degradation by various noncellulolytic thermophilic anaerobes and *Clostridium thermocellum*. *Appl. Environ. Microbiol.* **49**, 656–659 (1985).
121. Izquierdo, J. A., Pattathil, S., Guseva, A., Hahn, M. G. & Lynd, L. R. Comparative analysis of the ability of *Clostridium clariflavum* strains and *Clostridium thermocellum* to utilize hemicellulose and unpretreated plant material. *Biotechnol. Biofuels* **7**, 136 (2014).
122. Podkaminer, K. K., Shao, X., Hogsett, D. A. & Lynd, L. R. Enzyme inactivation by ethanol and development of a kinetic model for thermophilic simultaneous saccharification and fermentation at 50 °C with *Thermoanaerobacterium saccharolyticum* ALK2. *Biotechnol. Bioeng.* **108**, 1268–1278 (2011).
123. Lambertz, C. *et al.* Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. *Biotechnol. Biofuels* **7**, 1–15 (2014).
124. Hong, W. *et al.* The contribution of cellulosomal scaffoldins to cellulose hydrolysis by *Clostridium thermocellum* analyzed by using thermotargetrons. *Biotechnol. Biofuels* **7**, 80 (2014).
125. Xu, T. *et al.* Efficient genome editing in *Clostridium cellulolyticum* via CRISPR–Cas9 nickase. *Appl. Environ. Microbiol.* **81**, 4423–4431 (2015).
126. Wilchek, M., Bayer, E. A. & Livnah, O. Essentials of biorecognition: the (strept)avidin-biotin system as a model for protein–protein and protein–ligand interaction. *Immunol. Lett.* **103**, 27–32 (2006).

127. Hyeon, J. E., Kang, D. H. & Han, S. O. Signal amplification by a self-assembled biosensor system designed on the principle of dockerin–cohesin interactions in a cellulosome complex. *Analyst* **139**, 4790–4793 (2014).
128. Mori, Y. Purification and characterization of an endoglucanase from the cellulosome (multicomponent cellulase complex) of *Clostridium thermocellum*. *Biosci. Biotechnol. Biochem.* **56**, 1199–1203 (1992).
129. Sakka, K. *et al.* Analysis of cohesin–dockerin interactions using mutant dockerin proteins. *FEMS Microbiol. Lett.* **314**, 75–80 (2011).
130. Demishtein, A., Karpol, A., Barak, Y., Lamed, R. & Bayer, E. A. Characterization of a dockerin-based affinity tag: application for purification of a broad variety of target proteins. *J. Mol. Recognit.* **23**, 525–535 (2010).
131. Levy, I. & Shoseyov, O. Cellulose-binding domains: biotechnological applications. *Biotechnol. Adv.* **20**, 191–213 (2002).
132. Shpigel, E. *et al.* Immobilization of recombinant heparinase I fused to cellulose-binding domain. *Biotechnol. Bioeng.* **5**, 17–23 (1999).
133. Hyeon, J. E. *et al.* Production of functional agarolytic nano-complex for the synergistic hydrolysis of marine biomass and its potential application in carbohydrate-binding module-utilizing one-step purification. *Process Biochem.* **47**, 877–881 (2012).
134. Hussack, G. *et al.* Multivalent anchoring and oriented display of single-domain antibodies on cellulose. *Sensors (Basel).* **9**, 5351–5367 (2009).
135. Gourlay, K., Arantes, V. & Saddler, J. N. Use of substructure-specific carbohydrate binding modules to track changes in cellulose accessibility and surface morphology during the amorphogenesis step of enzymatic hydrolysis. *Biotechnol. Biofuels* **5**, 51 (2012).
136. Gao, S., You, C., Renneckar, S., Bao, J. & Zhang, Y.-H. P. New insights into enzymatic hydrolysis of heterogeneous cellulose by using carbohydrate-binding module 3 containing GFP and carbohydrate-binding module 17 containing CFP. *Biotechnol. Biofuels* **7**, 24 (2014).
137. Ding, S.-Y. *et al.* Ordered arrays of quantum dots using cellulosomal proteins. *Indust. Biotechnol.* **1**, 198–206 (2005).
138. Bayer, E. A. *et al.* in *Biotechnology of Lignocellulose Degradation and Biomass Utilization* (eds Sakka, K. *et al.*) 183–205 (Ito Print Publishing Division, 2009).
139. Currie, M. A. *et al.* Small angle X-ray scattering analysis of *Clostridium thermocellum* cellulosome N-terminal complexes reveals a highly dynamic structure. *J. Biol. Chem.* **288**, 7978–7985 (2013).
140. Hugenholtz, P. & Tyson, G. W. Microbiology: metagenomics. *Nature* **455**, 481–483 (2008).
141. Warnecke, F. *et al.* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**, 560–565 (2007).
142. Brulc, J. M. *et al.* Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl Acad. Sci. USA* **106**, 1948–1953 (2009).
143. Hess, M. *et al.* Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**, 463–467 (2011).
144. Morrison, M., Pope, P. B., Denman, S. E. & McSweeney, C. S. Plant biomass degradation by gut microbiomes: more of the same or something new? *Curr. Opin. Biotechnol.* **20**, 358–363 (2009).
145. Xia, Y., Ju, F., Fang, H. H. P. & Zhang, T. Mining of novel thermo-stable cellulolytic genes from a thermophilic cellulose-degrading consortium by metagenomics. *PLoS ONE* **8**, e53779 (2013).
146. Simões, M. F. *et al.* Soil and rhizosphere associated fungi in gray mangroves (*Avicennia marina*) from the red sea — a metagenomic approach. *Genomics Proteomics Bioinformatics* **13**, 310–320 (2015).
147. Morisaka, H. *et al.* Profile of native cellulosomal proteins of *Clostridium cellulovorans* adapted to various carbon sources. *AMB Express* **2**, 37 (2012).
148. Esaka, K., Aburaya, S., Morisaka, H., Kuroda, K. & Ueda, M. Exoproteome analysis of *Clostridium cellulovorans* in natural soft-biomass degradation. *AMB Express* **5**, 2 (2015).
149. Blouzard, J.-C. *et al.* Modulation of cellulosome composition in *Clostridium cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. *Proteomics* **10**, 541–554 (2010).
150. Vodovnik, M. *et al.* Expression of cellulosome components and type IV pili within the extracellular proteome of *Ruminococcus flavefaciens* 007. *PLoS ONE* **8**, e65333 (2013).
151. Munir, R. I. *et al.* Quantitative proteomic analysis of the cellulolytic system of *Clostridium termitidis* CT1112 reveals distinct protein expression profiles upon growth on α-cellulose and cellobiose. *J. Proteomics* **125**, 41–53 (2015).
152. Fendri, I. *et al.* The cellulosomes from *Clostridium cellulolyticum*: identification of new components and synergies between complexes. *FEBS J.* **276**, 3076–3086 (2009).
153. Raman, B., McKeown, C. K., Rodriguez, M., Brown, S. D. & Mielenz, J. R. Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation. *BMC Microbiol.* **11**, 134 (2011).
154. Wilson, C. M. *et al.* Global transcriptome analysis of *Clostridium thermocellum* ATCC 27405 during growth on dilute acid pretreated *Populus* and switchgrass. *Biotechnol. Biofuels* **6**, 179 (2013).
155. Borne, R., Bayer, E. A., Pagès, S., Perret, S. & Fierobe, H.-P. Unraveling enzyme discrimination during cellulosome assembly independent of cohesin–dockerin affinity. *FEBS J.* **280**, 5764–5779 (2013).
156. Fierobe, H. P. *et al.* 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 (2005).
One of the first studies to demonstrate the benefits of designer cellulosomes. In one example, a mixed designer cellulosome that contained two cellulases and a xylanase was constructed.
157. Moraïs, S. *et al.* Deconstruction of lignocellulose into soluble sugars by native and designer cellulosomes. *mBio* **3**, e00508-12 (2012).
158. Jeon, S. D., Yu, K. O., Kim, S. W. & Han, S. O. A celluloytic complex from *Clostridium cellulovorans* consisting of mannanase B and endoglucanase E has synergistic effects on galactomannan degradation. *Appl. Microbiol. Biotechnol.* **90**, 565–572 (2011).
159. Fierobe, H.-P. *et al.* Degradation of cellulose substrates by cellulosome chimeras: substrate targeting versus proximity of enzyme components. *J. Biol. Chem.* **277**, 49621–49630 (2002).
160. Vazana, Y. *et al.* A synthetic biology approach for evaluating the functional contribution of designer cellulosome components to deconstruction of cellulosic substrates. *Biotechnol. Biofuels* **6**, 182 (2013).
161. Molinier, A.-L. *et al.* Synergy, structure and conformational flexibility of hybrid cellulosomes displaying various inter-cohesins linkers. *J. Mol. Biol.* **405**, 143–157 (2011).
162. Caspi, J. *et al.* Effect of linker length and dockerin position on conversion of a *Thermobifida fusca* endoglucanase to the cellulosomal mode. *Appl. Environ. Microbiol.* **75**, 7335–7342 (2009).
163. Mingardon, F., Chantal, A., Tardif, C., Bayer, E. A. & Fierobe, H.-P. Exploration of new geometries in cellulosome-like chimeras. *Appl. Environ. Microbiol.* **73**, 7138–7149 (2007).
164. Stern, J. *et al.* Significance of relative position of cellulases in designer cellulosomes for optimized cellulolysis. *PLoS ONE* **10**, e0127326 (2015).

Acknowledgements
The authors are grateful for long-term collaborations with R. Lamed (co-discoverer of the cellulosome concept), M. Wilchek, E. Morag, Y.I Shoham, I. Borovok, I. Mizrahi, Y. Barak, the late F. Frolow, O. Livnah, S. P. Smith, D. Wilson, M. Himmel, S.-Y. Ding, Y. Bomble, H. Flint, B. White, J.-P. Belaich, A. Belaich, H.-P. Fierobe, B. Henrissat, M. Carrion-Vazquez, M. Czjzek, S. Jindou, M. Morrison, H. Gilbert, C. Fontes, H. Gaub, M. Nash, R. Bernardi, K. Schulten, and the many superb students, technicians and postdoctoral fellows who have contributed to the described research throughout the years. The authors specifically thank J. Weinstein for the design and preparation of figure 3 and B. Dassa for pre-publication of bioinformatic data. E.A.B. is currently supported by the following funding agencies: the Israel Science Foundation (ISF; grant 1349); the Israeli Center of Research Excellence (I-CORE Center; grant 152/11) managed by the ISF; by the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel; the European Union, Area NMP.2013.1.1-2: Self-assembly of naturally occurring nanosystems: CellulosomePlus Project number 604530; and European Union Horizon 2020 contract: Sustainable production of next generation biofuels from waste streams: WASTE2FUELS. The authors also acknowledge the Dana and Yossie Hollander Center for Structural Proteomics, the Leona M. and Harry B. Helmsley Charitable Trust, the Brazilian Friends of the Weizmann Institute of Science and the Jewish Community Endowment Fund. E.A.B. is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry.

Competing interests statement
The authors declare no competing interests.

DATABASES
NCBI Assembly Database:
http://www.ncbi.nlm.nih.gov/assembly/