For reprint orders, please contact reprints@future-science.com

Microbial enzyme systems for biomass conversion: emerging paradigms

Biofuels (2010) 1(2), 323–341

Michael E Himmel[1,2], Qi Xu[1,2], Yonghua Luo[1], Shi-You Ding[1,2], Raphael Lamed[3] & Edward A Bayer[4†]

The contemporary relevance of biofuels as an attractive replacement for liquid fossil fuels has rekindled global interest in the conversion of cellulosic biomass – the most abundant renewable source of carbon and energy on our planet. In order to achieve efficient systems for such a formidable substrate, we take guidance from the native enzyme systems of the microbes that have evolved to rid the natural environment of plant-derived wastes. These cellulolytic bacteria and fungi employ a diversity of contrasting but complementary mechanisms for the hydrolysis of cellulose and other related complex plant cell wall polysaccharides. This review covers various known microbial approaches for attacking the recalcitrance problem in the conversion of cellulosic biomass to soluble sugars en route to a biofuels-based society.

Biofuels derived from lignocellulose, the most abundant source of organic material on our planet, are an attractive alternative to current petroleum-based fuels, due to their potential for sustainability as well as reduction of greenhouse gas emissions. Plants have evolved mechanisms over millennia to protect the structural forms of polysaccharides from which their cell walls are comprised. Therefore, only a small fraction of microorganisms possess the ability to degrade cellulose efficiently. Fungi and bacteria are the dominant microorganisms responsible for lignocellulose degradation in the biosphere. These microbes show significant diversity in their surviving environments and can be found in mesophilic as well as thermophilic ecosystems where plant matter is abundant, such as forest and pasture soils, hot spring pools and decaying plant debris (Table 1) [1]. In these decay communities, degradation of the plant cell wall is accomplished by complex suites of hydrolytic enzymes that all cellulolytic microbes secrete outside of their cell wall. In this article, we will review the major paradigms of microbial enzyme systems for biomass conversion.

Lignocellulosic biomass: an abundant resource

Lignocellulosic biomass can be defined as crop residues, short rotation transgenic trees (e.g., poplars), woody grasses (e.g., switchgrass), forestry waste, construction waste, waste from pulp and paper production, agricultural residues and municipal solid waste. A recent study by Oak Ridge National Laboratory (ORNL) determined that 1.3 billion tons of lignocellulosic biomass was theoretically available in the USA alone [2]. Although some challenges remain for biomass-based biofuels processes, the sheer availability of biomass does not appear to be a limiting factor. Rather, the high cost of conversion is recognized as the major deterrent for commercialization. Crop residues consist primarily of plant tissues that are composed of various types of cell walls and these plant tissues pose significant resistance to chemical, enzymatic and microbial deconstruction.

Plant cell wall biomass represents the structural forms of monosaccharides, as opposed to starch and amylopectin, which represent the storage forms [3].

†Author for correspondence
[1]Biosciences Center, National Renewable Energy Laboratory, Golden, CO 80401, USA
[2]Bioenergy Science Center (BESC), Oak Ridge National Laboratory, TN 37830, USA
[3]Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv 69978, Israel
[4]Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel
Tel.: +972 8 934 2373; Fax: +972 8 946 8256; E-mail: ed.bayer@weizmann.ac.il

10.4155/BFS.09.25 © 2010 Future Science Ltd ISSN 1759-7269

Table 1. Representative cellulolytic microbes isolated from diverse natural ecosystems.

Bacteria		Fungi	
Species	**Source**	**Species**	**Source**
Aerobes (free, noncomplexed cellulases)			
Mesophilic bacteria		***Mesophilic fungi***	
Bacillus brevis[§]	Termite gut	*Aspergillus nidulans*[‡], *A. niger*	Soil, wood rot
Cellulomonas fimi[§]	Soil	*Agaricus bisporus*	Compost
Cellvibrio japonicus	Soil	*Coprinus truncorum*	Soil, compost
Cytophaga hutchinsonii[‡]	Soil, compost	*Geotrichum candidum*	Soil, compost
Paenibacillus polymyxa	Compost	*Penicillium chrysogenum*	Soil, wood rot
Pseudomonas fluorescens[‡], *P. putida*	Soil, sludge	*Phanerochaete chrysosporium*	Compost
Saccharophagus degradans[‡]	Rotting marsh grass	*Trichocladium canadense*	Soil
Sorangium cellulosum	Soil	*Hypocrea jecorina*[‡¶]	Soil, rotting canvas
Thermophilic bacteria		***Thermophilic fungi***	
Acidothermus cellulolyticus[‡]	Hot spring	*Chaetomium thermophilum*	Soil
Thermobifida fusca[‡]	Compost	*Corynascus thermophilus*	Mush compost
		Paecilomyces thermophila	Soil, compost
		Thielavia terrestris	Soil
Anaerobes (complexed or free, noncomplexed cellulases)			
Mesophilic bacteria		***Mesophilic fungi***	
Acetivibrio cellulolyticus	Sewage	Neocallimastix patriciarum	Rumen
Bacteroides cellulosolvens	Sewage	Orpinomyces joyonii	Rumen
Clostridium cellulolyticum[*]	Compost	Orpinomyces PC-2	Rumen
Clostridium cellulovorans	Wood fermenter	Piromyces equi	Rumen
Clostridium josui	Compost	Piromyces E2	Feces
Clostridium papyrosolvens[‡]	Mud (freshwater)		
Clostridium phytofermentans[‡]	Soil		
Fibrobacter succinogenes	Rumen		
Prevotella ruminicola	Rumen		
Ruminococcus albus[‡]	Rumen		
Ruminococcus flavefaciens[‡]	Rumen		
Thermophilic bacteria			
Anaerocellum thermophilum[‡]	Hot spring		
Caldicellulosiruptor saccharolyticus	Hot spring		
Clostridium thermocellum[‡]	Sewage, soil, manure		
Clostridium stercorarium	Compost		
Thermotoga maritima[‡]	Mud (marine)		
Rhodothermus marinus	Hot spring		

[‡]Microbial strains whose genome sequencing has been completed.
[§]Most Cellulomonas and Bacillus strains are facultative anaerobes that can also grow anaerobically.
[¶]Hypocree jecorina was originally named Trichoderma reesei.

The predominant polysaccharide in the cell walls of plants is cellulose, followed by the hemicelluloses and then the pectins. In general, the most abundant weight fraction cell wall type in plant tissue is the secondary cell wall, produced after the cell has stopped growing. Secondary cell walls contain structural polysaccharides, strengthened further with polymeric lignin covalently cross-linked to hemicellulose.

Plant cellulose is a linear unbranched polymeric chain, consisting solely of β-(1,4)-linked D-glucose residues. The single chains can reach lengths of over 10,000 glucose units. Cellulose is synthesized by cellulose synthase 'rosettes', which contain 36 enzyme units located in the cell membrane. Immediately after synthesis, the cellodextrin chains are directly deposited into the cell wall as elementary fibrils that coalesce to form successively larger microfibrils and in some cases, macrofibrils [4]. Plant cellulose is comprised of cellulose Iα and cellulose Iβ. Cellulose Iα is a triclinic form with one chain per unit cell and is of higher energy compared with the more stable monoclinic Iβ form [5,6]. The Iα form is more susceptible to hydrolysis, but plant cell wall cellulose consists mainly of the stable Iβ form. In cellulose, all of the glucosyl hydroxyl groups are in the equatorial position, whereas all of the axial positions are occupied by nonpolar (and nonhydrogen-bonding) aliphatic protons. Consequently, the 'sides' of the elementary microfibrils are polar and hydrogen bonding and the 'tops and bottoms' are hydrophobic. The relatively extended solution structure of cellodextrins permits them to aggregate with a regular crystalline packing, matching up hydrophobic faces as well as allowing hydrogen bond formation between chains [7]. Currently, the detailed mechanisms of cellulase attack on these cellulose allomorphs and their respective crystal faces remain unknown.

In addition to cellulose, the plant cell wall matrix contains two additional major types of cell wall polysaccharides; the hemicelluloses and the pectins. Unlike cellulose, both are synthesized in the Golgi apparatus, delivered to the cell membrane via small vesicles and secreted into the cell wall. Hemicelluloses are generally complex, branched carbohydrate polymers that are formed from different monomeric sugars attached through different linkages. Carbohydrate substituents and noncarbohydrate components occur in hemicelluloses on either the main chain or on the carbohydrate branches. The complex structures of the hemicelluloses are thought to confer a wide range of biophysical and biomechanical properties on the plant tissues in which they occur, as well as on products made from these tissues. The principal pentose sugar in the major plant cell wall hemicellulose is β-D-xylopyranose, which has only the position 2- and 3-carbons available for *O*-linked substitution by substituent sugars when β-(1,4) linked as in xylan. Hemicelluloses also include xyloglucan, arabinoxylan and glucomannans, which contain other sugars including the pentose arabinose and the aldohexoses glucose, mannose and galactose. The hemicelluloses can also be esterified by acetylation and/or cross-linked to lignins via p-coumaroyl and feruloyl groups.

As a complex phenolic polymer, lignins exist widely in cell walls of plants and some algae. There are three types of functional groups in lignins, including p-hydroxyphenyl, guaiacyl and syringyl, from which the monolignols, 4-hydroxycinnamyl, coniferyl and sinapyl alcohols, respectively, are comprised. The complex and highly variable chemical heterogeneity of lignin is due to the diversity of substitution patterns and intermolecular linkages utilized during polymerization. Although lignins enable critical functions for the plant, including mechanical support, water transport and defense, lignin is also an undesirable component in the biomass conversion process, due to its ability to shield polysaccharides from enzymatic hydrolysis and generally impede diffusion into plant tissue by chemicals and enzymes. In native plant cell walls, lignins are covalently linked to hemicellulose, which forms a matrixing layer around the cellulose comprising the microfibril core that further hinders cellulolytic and hemicellulolytic enzymes. Moreover, many of the lignin degradation products are either inhibitory or generally detrimental to the plant cell wall polysaccharide-degrading enzymes. It is therefore necessary to take this into account when designing an enzymatic process for degradation of lignocellulosic biomass.

Key terms

Lignocellulosic biomass: Biomass feedstock containing cellulose, hemicelluloses, pectins and lignin but not starch; usually crop residues or wood waste

Cellulases: Enzymes that hydrolyze the β-(1,4) bond in cellulose

Hemicellulases: Enzymes that hydrolyze any portion of the various complex, branched polysaccharides known as hemicelluloses, such as the β-(1,4) bond of the xylan main chain or any of the various linkages in the side chains

Glycoside hydrolases: Enzymes that hydrolyze a glycosidic bond between two adjacent saccharide groups or between a carbohydrate and a noncarbohydrate moiety

Dominant paradigms for plant cell walls

Bacteria and fungi that decompose plant cell wall polysaccharides efficiently employ a multitude of remarkable enzymes to accomplish this particularly challenging feat. These highly specialized, often intricate enzyme systems include the **cellulases**, **hemicellulases** and other related **glycoside hydrolases**, as well as the polysaccharide lyases and the carbohydrate esterases. The selected microbes that have evolved to occupy different lignocellulolytic ecosystems utilize a surprisingly varied set of strategies to either compete or collaborate with other bacteria and fungi and to, thus, survive and thrive in their environment.

▪ Free enzyme systems

In cellulolytic bacteria and fungi, the cellulases all hydrolyze the same type of bond of the cellulose chain (i.e., the β-[1,4]-glucosidic bond). They do so, however, using different modes of action. The definitive enzymatic degradation of cellulose to glucose is generally accomplished by the synergistic action of three distinct classes of enzymes:

Key terms

Endoglucanases: Enzymes that hydrolyze the bonds in cellulose via random cleavage at internal sites of the cellulose chain

Exoglucanases: 'Processive' enzymes that hydrolyze the bonds in cellulose via sequential cleavage from either the reducing or nonreducing free chain ends

- The endo-β-(1,4)-glucanases or β-(1,4)-D-glucan-4-glucanohydrolases (EC 3.2.1.4), which act randomly on soluble and insoluble β-(1,4)-glucan substrates and are commonly measured by detecting the reducing groups released from carboxymethylcellulose;
- The exo-β-(1,4)-D-glucanases, including both the β-(1,4)-D-glucan glucohydrolases (EC 3.2.1.74), which liberate D-glucose from β-(1,4)-D-glucans and hydrolyze D-cellobiose slowly, and β-(1,4)-D-glucan cellobiohydrolase (EC 3.2.1.91), which liberates D-cellobiose in a 'processive' manner (successive cleavage of product) from β-(1,4)-glucans;
- The β-D-glucosidases or β-D-glucoside glucohydrolases (EC 3.2.1.21), which act to release D-glucose units from cellobiose and soluble cellodextrins, as well as an array of glycosides. The above classification scheme is not entirely rigid and a few enzymes have properties that do not fit one of the above definitions.

Free cellulases frequently bear a cellulose-binding carbohydrate-binding module (CBM) that delivers the catalytic module to the surface of its crystalline cellulosic substrate as shown schematically in **Figure 1** [8]. In aerobic fungi, the CBM is invariably from family 1, which is very small (~30–35 amino acid residues). The ancillary CBMs of bacterial cellulases are often from family 2 or 3, which are much larger than their fungal analogs, comprising approximately 100 and 150 amino acid residues, respectively. Despite the differences in size, these types of cellulose-binding CBMs all exhibit a planar array of aromatic side chains located on a relatively flat surface of the CBM molecule. These planar-strip residues are generally highly conserved and are believed to align with the hydrophobic faces of the glucose rings along the length of a single cellodextrin lying on the cellulose surface, thus providing the structural rationale for substrate binding of the CBM, the parent enzyme, the cellulosome (see below) or the entire microbial cell (in cases where a CBM is attached to the cell surface).

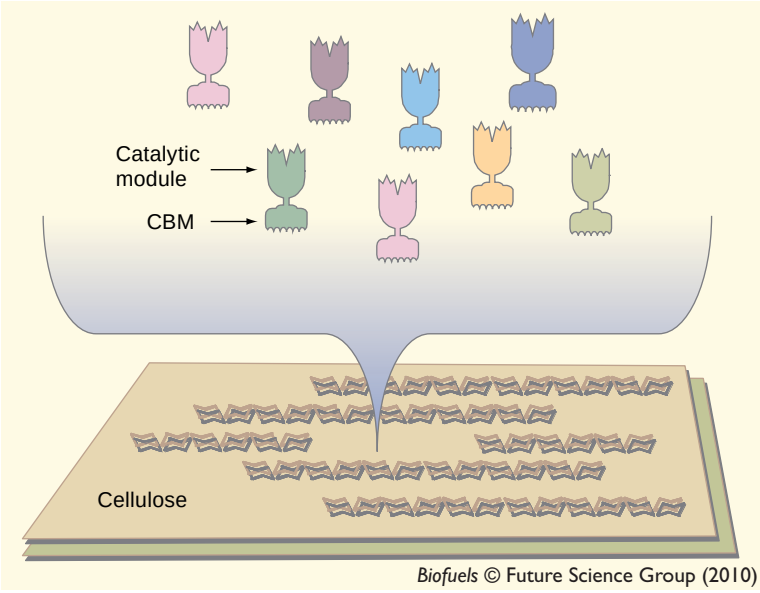

Figure 1. Interaction of free cellulase system with cellulosic substrates. The CBM of each enzyme delivers the catalytic module to the cellulosic substrate and the various individual enzymes act in a synergistic manner to degrade plant cell wall polysaccharides. Some CBMs are specific for noncrystalline or hemicellulose portions of the plant cell wall and their parent enzymes (and catalytic module) are directed towards the appropriate substrate (not shown in the figure).
CBM: Cellulose-binding module.

The cellulases and hemicellulases belong to the glycoside hydrolases (GHs), a large group of enzymes, which hydrolyze the glycosidic bond between two or more carbohydrates or between a carbohydrate and a noncarbohydrate moiety. Classification schemes have been based previously on substrate specificities of an enzyme, but this is inappropriate for glycoside hydrolases, since the same protein fold often harbors several types of specificities. A more appropriate classification scheme was instituted [9], based on the amino acid sequence and consequent fold of the protein. The various glycoside hydrolases form 115 different families to date and membership of a given enzyme into a known GH family provides insight into its comparative structural features within a family, its evolutionary relationships with other family members and its mechanism of action. A compendium of the glycoside hydrolases and related carbohydrate-active enzymes (CAZymes) can be found on the CAZy Website [201]. In addition to the GHs, the polysaccharide lyases are classified into 21 families and carbohydrate esterases into 16. The CBMs are also divided into families, currently numbering 59 on the CAZy database.

Structurally, the topology of the active sites differs between the **endoglucanases** and **exoglucanases**. The active sites of endoglucanases typically attain a cleft-like topology. Thus, a cellulose chain can be accessed in random fashion by an endoglucanase and bond cleavage can occur anywhere along the chain of the substrate. By contrast, the active site of the exoglucanases resembles a tunnel, formed by long loops of the protein molecule that fold over the active site residues [10]. Consequently, a single glycan chain is fed into one end of the tunnel-like active site, followed by subsequent bond cleavage inside the tunnel and release of cellobiose product from the other end [11,12]. Since the chain is still fixed within the active site tunnel, successive cleavage events can continue processively in a unidirectional manner along

the glucan chain [13,14]. It seems, however, that some differences in this mechanism may occur among the different types of exoglucanases [15], perhaps reflecting the length of the tunnel, the directionality of action (from nonreducing to reducing end or vice versa) and flexibility of the loops that form the tunnel.

Selected endoglucanases can also exhibit a distinctive type of processive action on the substrate [16]. For example, certain types of GH9 enzymes contain a subfamily of CBM3, termed CBM3c, which is fused to the catalytic module; the two modules are connected tightly via a characteristic linker segment. The GH9-CBM3c couple work in concert, such that the CBM3c serves to feed a single cellulose chain into the active site cleft of the endoglucanase, modifying its character from a simple endoglucanase, which acts randomly along the substrate, to one that acts successively along its chain.

In contrast to microbial degradation of cellulose, bacteria and fungi produce many different types of enzymes that act very efficiently on the various types of hemicellulose. The very complex structure of hemicelluloses, however, requires a very large number of enzymes for complete degradation. Hemicellulases can be placed into three general categories:

- The endo-acting enzymes attack polysaccharide chains internally;
- The exo-acting enzymes tend to act processively from either the reducing or nonreducing termini;
- The so-called 'accessory' enzymes required to hydrolyze hemicellulose in native plant tissue.

The endo-acting hemicellulases exhibit very little activity on short oligomers (i.e., degree of polymerization [DP] less than 3). In contrast, some exo-acting enzymes have preferences for short chain substrates (DP 2–4) although others prefer larger substrates (DP >4). The accessory enzymes include a variety of acetyl esterases and esterases that hydrolyze lignin-linked glycoside bonds, such as coumaric acid esterase and feruloyl acid esterase [17]. The complexity of hemicellulose structures requires a high degree of coordination between the enzymes involved in hemicellulose degradation. Synergism in enzyme action is defined as a mixture of enzymes that produce more end product than each could produce separately. In the cellulase system, for example, synergism can be easily shown for native and artificial combinations of endo- and exo-glucanases [18]. As may be expected for an analogous, yet more complex series of biopolymers, synergism has been demonstrated between β-xylanases and acetylxylan esterase [19], α-L-arabinofuranosidase [20] and β-glucuronidase [21]. Most enzymes have very specific requirements for tight substrate binding and precise transition state formation, which usually leads to high catalytic turnover rates.

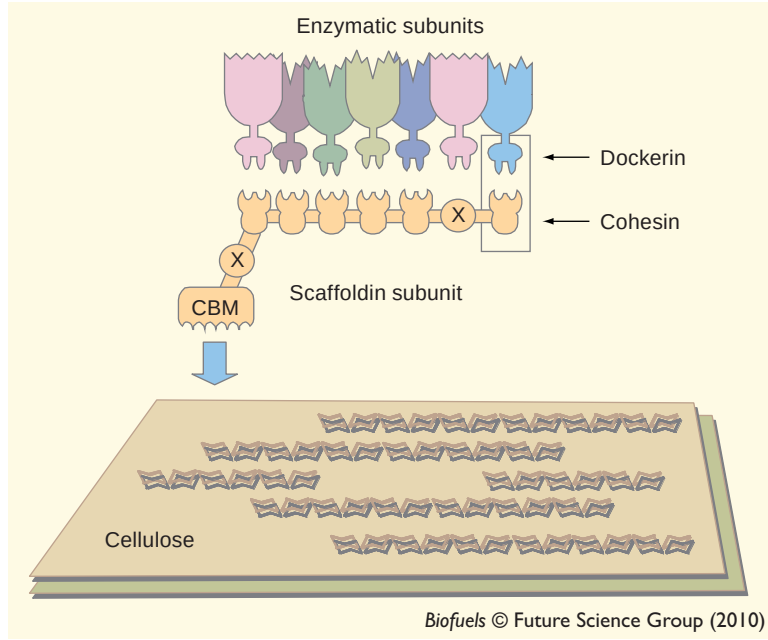

Figure 2. Interaction of the cellulosome with cellulosic substrates. Cellulosomal enzymes are integrated into the complex via the potent protein–protein interaction between the single dockerin of each enzymatic subunit and multiple cohesin modules on the scaffoldin subunit. The cohesins are separated on the scaffoldin polypeptide by distinctive linker segments. The entire cellulosome complex is delivered to the cellulosic substrate through the action of the lone CBM of the scaffoldin subunit. Scaffoldins can also carry one or more 'X' modules of unknown function. Some dockerin-bearing cellulosomal enzymes also contain CBMs that are selective for noncrystalline or hemicellulosic portions of the plant cell wall and would therefore be directed towards these substrates during the degradation process while tethered to the flexible scaffoldin subunit (not shown in the figure).
CBM: Cellulose binding module.

Cellulosomes

In general, the multienzyme cellulosome complex is composed of two major types of subunit: the noncatalytic scaffoldin(s) and the enzymes (Figure 2) [22–24]. The assembly of the enzymatic subunits into the cellulosome complex is facilitated by the high-affinity recognition between cohesin modules of the scaffoldin subunit and enzyme-borne dockerin modules. Scaffoldins usually contain multiple cohesin modules, thereby enabling numerous different enzymes to be assembled into the cellulosome complex (Table 2). In addition, a multiplicity of scaffoldins has been found in some species, which lends a higher level of complexity to cellulosome assembly. Theoretically, over 70 different dockerin-containing components can be assembled into the cellulosome of *Clostridium thermocellum* [25,26]. Since the scaffoldin subunit in this bacterium contains only nine cohesin modules, the varied collection of individual **cellulosomes** is immensely heterogeneous.

Table 2. Diversity of scaffoldins derived from various cellulosome-producing bacteria.

Species	Type of scaffoldin composition	Protein	Modular architecture	Mol. Wt. (kDa)
Clostridium cellulovorans	Single	CbpA	CBM3-X-Coh1-Coh2-X-Coh3-Coh4-Coh5-Coh6-Coh7-Coh8-X-X-Coh9	186
Clostridium cellulolyticum	Single	CipC	CBM3-X-Coh1-Coh2-Coh3-Coh4-Coh5-Coh6-Coh7-X-Coh8	155
Clostridium josui	Single	CipJ	CBM3-X-Coh1-Coh2-Coh3-Coh4-Coh5-Coh6	117
Clostridium acetobulyticum	Single	CipA	CBM3-X-X-Coh1-X-Coh2-X-Coh3-X-Coh4-X-Coh5	152
Bacteroides cellulosolvens	Multiple	(1) ScaA	(1) Coh1- Coh2-Coh3-Coh4-Coh5-CBM3-Coh6-Coh7-Coh8-Coh9-Coh10-Coh11-Doc	(1) 242
		(2) ScaB	(2) Coh1-Coh2-Coh3-Coh4-Coh5-Coh6-Coh7-Coh8-Coh9-Coh10-X-SLH	(2) 240
Clostridium thermocellum	Multiple	(1) CipA	(1) Coh1-Coh2-CBM3-Coh3-Coh4-Coh5-Coh6-Coh7-Coh8-Coh9-XDoc	(1) 194
		(2) OlpB	(2) Coh1-Coh2-Coh3-Coh4-Coh5-Coh6-Coh7-SLH	(2) 245
		(3) Orf2p	(3) Coh1-Coh2-SLH	(3) 245
		(4) SdbA	(4) Coh-X-SLH	(4) 66
Acetivibrio cellulolyticus	Multiple	(1) ScaA	(1) GH9-Coh1-Coh2-Coh3-CBM3-Coh4-Coh5-Coh6-Coh7-XDoc	(1) 197
		(2) ScaB	(2) Coh1-Coh2-Coh3-Coh4-Doc	(2) 97
		(3) ScaC	(3) Coh1-Coh2-Coh3-SLH	(3) 124
		(4) ScaD	(4) Coh1-Coh2-Coh3-SLH	(4) 89
Ruminococcus flavefaciens	Multiple	(1) ScaA	(1) X-Coh1-Coh2-Coh3-Doc	(1) 90
		(2) ScaB	(2) Coh1-Coh2-Coh3-Coh4-Coh5-Coh6-Coh7-X-Doc	(2) 178
		(3) ScaC	(3) Coh-Doc	(3) 26
		(4) ScaE	(4) Coh-Ssm	(4) 28

CBM3: Cellulose-binding family-3; Coh: Cohesin; Doc: Dockerin; GH9: Family-9 cellulase; SLH: S-layer homology module; Ssm: Sortase-signal motif (SPKTG); X: Module of unknown function; XDoc: Dyad of X-module and dockerin.

Another important scaffoldin-borne component is the cellulose-specific CBM, which functions as the major binding factor for specific recognition of cellulosic substrates. The CBM of the scaffoldin serves to deliver the entire complement of cellulosome enzymes collectively to the lignocellulosic substrate, thus fulfilling another important requirement for efficient degradation.

In many aspects, cellulosomal enzymes are very similar to their free counterparts, except their catalytic modules are attached to a dockerin rather than a CBM. The scaffoldin-based CBM serves as a single cellulose-targeting agent for all cellulosomal components. Members of the same families of cellulases and hemicellulases that are involved in the free enzyme systems also serve as cellulosomal enzymes – with some exceptions. In this context, the GH7 and GH45 cellulases that occur exclusively in fungi never appear in the cellulosomal context. Intriguingly, however, GH6 enzymes that occur both in fungi as well as some bacteria, have not been found in native cellulosome systems. With so much horizontal transfer of GH genes among the microbes in a given ecosystem, it is fascinating why members of selected cellulase families would be excluded from cellulosomal systems. Compared with free enzyme systems, the cellulosome brings the catalytic modules into close physical association with each other and, collectively, to the cellulose surface, thereby promoting their essential synergistic action by concentrating the enzymes with complementary functions at defined sites on the lignocellulosic substrate. Nevertheless, enzyme proximity within a complex could theoretically reduce synergism as a consequence of conformational restrictions imposed by the intricate quaternary structure of the cellulosome. This potential obstacle, however, is countered by the plasticity (or flexibility) of the cellulosome that provides an important clue to its functionality. Plasticity of the quaternary structure of the cellulosome is thus a major functional rationale for the observed synergy among its enzymatic subunits. The following properties of native cellulosomal components contribute to the observed plasticity of the complex:

- The scaffoldin(s) of the cellulosome contains intermodular linkers of various lengths, but usually very long [27];
- The linker between the catalytic module and dockerin of the parent cellulase provides fine-tuned positioning of the catalytic module on the substrate [28];
- The type I dockerin of the *C. thermocellum* cellulosomal cellulases can bind to the corresponding scaffoldin-borne cohesins in two alternative modes, thereby generating two opposing orientations for the associated catalytic module on the cohesin–dockerin interface [29].

The purported conformational 'flip-flop' thus generated by this dual mode of binding provides significant plasticity by further modulating the orientation of the catalytic subunits within the supramolecular cellulosome assembly, relative to the cellulosic substrate, as the cellulolytic process proceeds [30,31]. This dual mode of binding

Key term

Cellulosomes: Discrete high-molecular weight enzyme complexes secreted from anaerobic bacteria, which contain considerable diversity in content of glycoside hydrolases and other related plant cell wall-degrading enzymes and subcomponents

that characterizes integration of the enzymes into the complex is contrasted by the apparent single mode of cohesin-dockerin binding displayed by the attachment of the scaffoldin subunit to the cell surface [32].

Conventional sequencing of genes identified in various bacteria, together with associated bioinformatics and biochemical analyses, has provided novel information regarding the components and architecture of cellulosome systems from different anaerobic cellulolytic bacteria [33]. Following the initial sequencing and description of the cellulosomal scaffoldins in *Clostridium thermocellum* and ***Clostridium cellulovorans*** [34,35], other clostridial scaffoldins have been sequenced [36,37]. These, in turn, were followed by the sequencing of various other scaffoldins from non-clostridial species [38–43]. A list of the diverse nature of known scaffoldins is provided in **Table 2**.

Recent genome sequencing projects of several true cellulosome-producing bacteria, including *Clostridium thermocellum* [202], *Clostridium acetobutylicum* [203], *Clostridium cellulolyticum* [204], *Clostridium papyrosolvens* [205], *Ruminococcus albus* [44] and *Ruminococcus flavefaciens* [45], have further enriched our knowledge of the bacterial cellulosomes and the information achieved has allowed us to grasp their sheer complexity. In addition to the intricate scaffoldins, these genomes encode large numbers of dockerin-containing proteins (from several dozen in some cellulosome-producing bacteria to over 200 in *R. flavefaciens*, ranging in size from ~40 to ~170 kD) – much larger than what a single scaffoldin can accommodate. The cellulosomal enzymes include numerous cellulases from divergent glycoside hydrolase families, but the majority of the enzymes are noncellulolytic and specialize in the degradation of pectin and hemicelluloses. Each genome also reveals the presence of dockerin modules attached to proteins that are not carbohydrate-acting enzymes (including putative proteases and protease inhibitors) as well as dockerin-containing proteins of currently unknown function.

Polysaccharide-cleaving enzymes (glycoside hydrolases and polysaccharide lyases) and accessory carbohydrate esterases are classified in a large number of families based on amino acid sequence similarities [9], accessible online via the carbohydrate-active enzyme (CAZy) database [201]. Preliminary examination of the occurrence of members of these families in cellulosomes shows that in each of the fully sequenced (public) genomes mentioned above [46,47], there is only one cellulosomal GH48 cellobiohydrolase, in sharp contrast to the occurrence of much larger numbers of GH9 endoglucanases. In *C. thermocellum*, the dockerin-containing GH48 enzyme is believed to be a major and decisive component of the cellulosome, when grown on microcrystalline cellulose [48,49]; growth of the bacterium on cellobiose results

Table 3. Selected bacterial multifunctional enzymes.

Modular structure‡	Mode (f,c)§	Accession code	Source
Cellulase–cellulase			
GH9-CBM3-CBM3-CBM3-GH48	*f*	CAB06786	*Anaerocellum thermophilum*
GH5-CBM10-GH6	*f*	ABS72374	*Teredinibacter turnerae T7902*
GH6-CBM3-GH12-CBM2	*f*	ABK52388	*Acidothermus cellulolyticus 11B*
Cellulase–hemicellulase			
GH10-CBM3-CBM3-GH48	*f*	ACM60945	*Anaerocellum thermophilum DSM 6725*
CBM30-Ig-GH9-GH44-Doc-CBM44	*c*	BAA12070	*Clostridium thermocellum*
Hemicellulase–hemicellulase			
CBM22-CBM22-GH10-CBM3-CBM3-CBM3-GH43-CBM6	*f*	AAB95326	*Caldicellulosiruptor sp. Rt69B.1*
GH5-CBM3-CBM3-GH44	*f*	ABP66691.	*Caldicellulosiruptor saccharolyticus*
GH43-CBM6-CBM2-CBM22-GH10	*f*	ABD82867	*Saccharophagus degradans 2–40*
GH44-Fn3-GH26-CBM3	*f*	ABC88431	*Paenibacillus polymyxa*
GH30-GH54-GH43-Doc	*c*	ZP_00510825	*Clostridium thermocellum*
GH11-CBM22-GH10-Doc-GH11	*c*	ORF00468	*Ruminococcus flavefaciens*
UNK-CBM22-GH10-CBM22-Doc-GH43-CBM6	*c*	ORF03865	*Ruminococcus flavefaciens*
Hemicellulase–carbohydrate esterase			
CE1-CBM6-Doc-GH10	*c*	ABN53181	*Clostridium thermocellum*
GH5-Doc-CE2	*c*	AAA23224	*Clostridium thermocellum*
GH11-CBM6-Doc-CE4	*c*	AAC04579	*Clostridium thermocellum*
GH11-CBM22-GH10-Doc-CBM22-CE4	*c*	ORF01222	*Ruminococcus flavefaciens*
GH11-CBM22-GH10-Doc-GH11-CE4	*c*	ORF03896	*Ruminococcus flavefaciens*
GH43-CBM6-CBM22-Doc-CE1	*c*	ORF00764	*Ruminococcus flavefaciens*
GH53-CE3-Doc	*c*	ORF01739	*Ruminococcus flavefaciens*

‡Family numbers follow the GH, CE and CBM abbreviations
f: Free (dockerin-containing) enzyme; c: Cellulosomal (dockerin-containing) enzyme.
CBM: Cellulose-binding module; CE: Carbohydrate esterase; Doc: Dockerin; Fn3: Fibronectin 3 domain; GH: Glycoside hydrolase; Ig: Immunoglobulin-like domain; UNK: Unknown domain/module.

in reduced amounts of this component in the cellulosome [26,50]. Moreover, this bacterium also produces a CBM-containing noncellulosomal GH48 [51], the only microorganism known to produce two of these intriguing exoglucanases. The GH9 enzymes include several different modular themes and consequent alterations in activity patterns [52]. The versatility in the repertoire of the cellulosomal GH9 enzymes may be a necessary and advantageous adaptation of cellulosomes [53,54]. In addition to the latter cellulosomal enzymes, additional components, both enzymatic and structural, seem to be affected by the carbon source on which the bacterium is grown [43,48,50,55–58]. More recently, genome sequencing combined with metabolic profiling have further revealed in much greater detail the striking dependence of growth substrate on cellulosome content [26,45,59].

The concept of producing artificial multienzyme complexes based on recombinant DNA technology was initiated long ago [60]. After a lengthy 'learning process', prototype 'designer cellulosomes' were indeed reported [61–63], whereby chimeric cohesin-containing scaffoldins and complementary dockerin-containing enzymes could be produced. Enhanced synergistic activities of exoglucanases and endoglucanases, selectively incorporated into discrete designer cellulosomes, as well as cellulase–hemicellulase complexes, served to demonstrate unambiguously both the CBM-mediated targeting effect and the enzyme-proximity effect for the first time.

In subsequent studies, mixed fungal and bacterial enzymes were shown to be appropriate in the cellulosome mode [64], as well as the incorporation of distinctly foreign enzymes (e.g., noncellulosomal enzymes, including GH6 cellulases) into designer cellulosomes [64–67]. Intriguingly, a GH6 endoglucanase was recently shown to perform well in designer cellulosomes, whereas a GH6 exoglucanase exhibited measurable but markedly reduced activities in the cellulosome mode [68]. A particularly ingenious report [69] demonstrated extraordinary flexibility in the possible 'geometries' of designer cellulosomes, indicating the utility of their modular components (e.g., cohesins, dockerins and CBMs) as building blocks for future synthetic cellulosomes.

It is hoped that this synthetic biology approach may someday alleviate the problem of limited production capacity inherent in the anaerobic setting, since the rather few cellulosome-producing organisms thus far identified all appear to be strict anaerobes. In this context, designer cellulosomal components can be produced independently in a prolific aerobic host cell system. Alternatively, the genes encoding these components can be introduced directly into host bacteria [70–72], fungi [73] or yeasts [74,75], thereby providing novel cellulolytic microbes for efficient degradation of cellulosic biomass.

Multifunctional enzyme systems

Enzymes that are composed of two or more catalytic modules related to the degradation of plant cell walls are termed 'multifunctional enzymes'. The significant feature of their modular architectures is that they are composed of more than one catalytic module and distinct CBM(s) and these enzymes are usually of very high molecular weight **(Figure 3)**. The presence of two different enzymes in the same polypeptide chain would seem to indicate that the forced proximity of the designated catalytic modules necessitates concerted action on a given portion of the lignocellulosic substrate. Naturally occurring multifunctional enzymes exist in both free enzyme systems and cellulosomal systems. Some multifunctional enzymes and their modular architectures are listed in **Table 3**. Based on their primary catalytic modules, multifunctional enzymes can be classified into four types as described below.

Cellulase–cellulase systems

This type of multifunctional enzyme may include two or more cellulases, such as the catalytic GH5, GH6, GH9, GH48 and other ancillary modules or domains, such as CBMs, X modules, fibronectin domains and Ig-like domains, in their molecular architecture **(Table 3)**. These very large enzymes should, in theory, be capable of hydrolyzing microcrystalline cellulose, since some of them contain both an endoglucanase and an exoglucanase module in the same polypeptide chain, along with one or more CBMs. Such enzymes would presumably act with enhanced synergy, since their proximity to one another would allow the newly formed free chain ends in the midst of the cellulose chain, created by the endoglucanases, to be exposed to processive action by the exoglucanase. Interestingly, a multifunctional cellulase, CelA, of the hypothermophile, *Anaerocellum thermophilum* [76] (now renamed *Caldicellulosiruptor bescii*), carries three consecutive CBM3s, which raises the question of why so many? One of these is a CBM3c fused to the neighboring GH9, which would seemingly modify its activity from a simple to a processive endoglucanase (as described in an earlier section). The presence of the other two, presumably conventional cellulose-binding CBM3s, may be linked to the very high temperatures (75–80°C) of the natural environment in which the bacterium thrives; the CBM3 dyad may be necessary to ensure tight binding to the substrate under such extreme conditions. Interestingly, this bacterium is reportedly capable of efficiently degrading various types of lignocellulosic biomass without a pretreatment step [77], although additional research will be required to corroborate this claim.

Hemicellulase–hemicellulase systems

These multifunctional enzymes are comprised primarily of two or more modules of hemicellulases as well as CBMs related to the binding to various hemicelluloses. These hemicellulases include GH10, GH26, GH43 and GH54, whereas the CBMs consist primarily of hemicellulose-binding CBMs, such as CBM6, CBM22 and CBM30 **(Table 3)**. CBMs related to the binding of hemicelluloses usually bind to single polysaccharide chains and may serve to direct one or both of the associated enzymes to relevant portions of the substrate, either before or during the process of degradation, as newly exposed sites on the substrate become accessible. The structures of their binding sites are extended clefts of various depths in which aromatic residues, appropriately oriented, recognize and bind to ligands [78,79]. It is interesting that distinctive cellulose-binding CBMs have also been found in these multifunctional enzymes. For example, GH26 usually exhibits mannanase activity, GH10 is one of the major xylanase families and GH43 and GH44 have xylosidase and xyloglucan activity, respectively. Nevertheless, multifunctional enzymes, which contain these catalytic modules also include CBM3s that are thus far associated almost exclusively with the property of cellulose binding. Again, multifunctional hemicellulase–hemicellulase enzymes of the hyperthemophile, *Caldicellulosiruptor saccharolyticus*, a close relative of *A. thermophilum* [76], carry a CBM3 dyad, presumably owing to the very high temperatures at which the bacterium thrives.

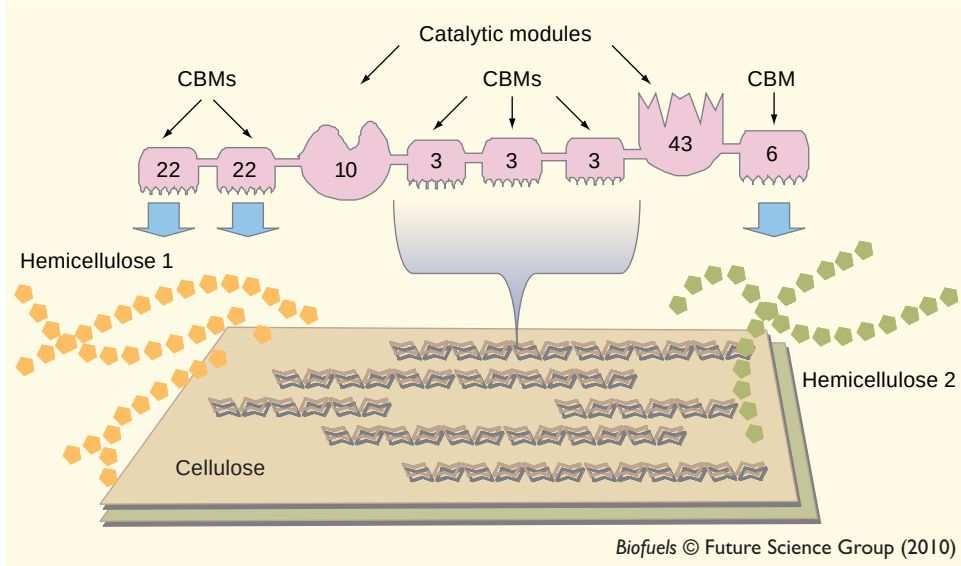

Figure 3. Interaction of multifunctional enzyme systems with the plant cell wall. The enzyme shown schematically in the figure is based on xylanase C from the hyperthermophile, *Caldicellulosiruptor sp.* Rt69B.1 [126]. In this case, the enzyme is attached to the crystalline cellulosic substrate by the strong interaction of adjacent family-3 CBMs and the catalytic modules are then directed secondarily to two different hemicellulosic portions of the plant cell wall by different types of hemicellulose-specific CBMs (from families 6 and 22, respectively, as depicted in the figure).
CBM: Cellulose binding module.

Cellulase–hemicellulase systems

Mixtures of cellulase catalytic modules (GH9 and GH48) and hemicellulase catalytic modules (GH10 and GH44) have also been found as multifunctional enzymes. CBMs with the ability to bind to cellulose (CBM3) and hemicellulose (CBM30), as well as other modules, such as Fn3 and Ig-like, were also identified in these enzymes **(Table 3)**. One would be tempted to speculate that such mixed types of multienzyme systems would infer that the enzymes act at areas of the substrate interface where cellulose and certain hemicelluloses meet within the plant cell wall.

Hemicellulase–carbohydrate esterase systems

These complex enzymes consist of hemicellulase catalytic modules (GH5, GH10, GH11, GH43 and GH53) and carbohydrate esterase modules (CE1, CE2, CE3 and CE4), as well as selected CBMs (CBM3, 6 and 22) **(Table 3)**. Members of CE1 through C4 are known to exhibit acetyl xylan esterase activity (CE2 and CE3 thus far show such activity exclusively). The utility of coupling a xylanase with a xylan esterase is clear, since the two would simultaneously sever the acetyl substituent and hydrolyze the main-chain glycosidic bond, thus effecting a particularly efficient cleavage at this common type of site on the xylan polymer. Even more remarkable is the combination of a xylanase and CE1 feruloyl esterase on the same polypeptide chain, as in one of the *C. thermocellum* cellulosomal enzymes. In this case, effective cleavage of the xylan main chain together with severing of the xylan–lignin junction would be especially beneficial to the parent bacterium.

Biochemical significance

The biochemical characteristics of some multifunctional glycoside hydrolases have been examined in detail. Some of these modules displayed enzymatic activities on various substrates, including Avicel, xylan, mannan, lichenin, chitin and carboxymethylcellulose [80,81].

Table 4. Cell-surface plant cell wall-degrading enzymes from different bacteria.

Enzyme	Enzyme activity	Mode of cell attachment	Bacterium	Ref. or source
XynX	Xylanase	SLH	Clostridium thermocellum	M67438.1, GI:144775
LicA	Lichenase	SLH	C. thermocellum	X89732.2, GI:20152505
AapT	α-amylase-pullulanase	SLH	Bacillus sp.	D28467.1, GI:460686
AlkA	Cellulase	SLH	Bacillus sp.	M27420.1, GI:142664
XynB	Xylanase	SLH	Caldicellulosiruptor sp.	AF036923.1, GI:2760904
Man26A	Mannanase	SLH	Cellulomonas fimi	AF126471.1, GI:5359709
EgA	Endoglucanase	SLH	Clostridium josui	D85526.2, GI:109715768
ManA	Mannanase	SLH	Anaerocellum thermphilum	AF126471.1, GI:5359709
PglA	Polygalacturonase	SLH	Thermoanaerobacterium thermosulfurigenes	U50951.1, GI:1542972
GH10	Xylanase	Ssm	Ruminococcus flavefaciens	ORF01899
CsxA	Exo-β-D-glucosaminidase	CBM35	Amycolatopsis orientalis	[125]
Cel48A	Cellobiohydrolase	CBM37	Ruminococcus albus	[97]
Cel9B	Endoglucanase	CBM37	R. albus	[97]
Cel9C	Endoglucanase	CBM37	R. albus	[98]
Cel5G	Endoglucanase	CBM37	R. albus	[98]
Xyn11C	Xylanase	CBM37	R. albus	[98]

CBM35, CBM37, CBMs from families 35 and 37, respectively.
CBM: Cellulose binding module; SLH: S-layer homology module; Ssm: Sortase-signal motif (LPXTG).

Remarkably, as mentioned earlier, an especially large cellulase gene, coding CelA with the modular architecture GH9-CBM3-CBM3-CBM3-GH48, was isolated from the extremely thermophilic, cellulolytic bacterium *A. thermophilum*. The GH9 module at its N-terminus exhibited endoglucanase activity; however, removal of the C-terminal GH48 module led to a significant reduction of the activity of the truncated CelA towards Avicel. The enhanced Avicelase activity of the intact CelA presumably resulted from intramolecular synergism between the GH9 endoglucanase and the GH48 exoglucanase [82]. This work provided direct experimental evidence to support the concept that intramolecular synergy of various catalytic modules can be obtained in a single multifunctional enzyme.

The formation of multifunctional enzymes could be regarded as the naturally occurring fusion of various catalytic and other modules to perform their respective functions in the hydrolysis of plant cell walls. Similar to the designer cellulosomes described above, we can adopt this 'natural' strategy to design and construct artificial multifunctional enzymes (chimeras) based on the structures and functions of the modular components, according to the requirement of a particular substrate configuration, in order to attain optimal levels of degradation. Compared with the mixture of individual enzymes, the chimeras possess the following advantages:

- Efficient degradation due to intramolecular synergy;
- Reduced requirement for protein purification;
- Simplified protein immobilization;
- Easier optimization of physical characterization, including pH and tolerance to higher temperature.

Like designer cellulosomes, some multifunctional enzyme chimeras have been created successfully and have shown promising functions [83–85]. In this context, a fusion protein of a xylanase (Xyln) and arabinofuranosidase (Ara) was constructed via a flexible linker peptide. Compared with the Xyln–Ara free enzyme mixture, the Xyln–Ara chimera yielded 30% higher activity on wheat arabinoxylan. This result supports the concept of intramolecular synergy of these modules and demonstrates the feasibility of generating effective multifunctional enzymes for the improvement of xylan bioconversion [86]. Therefore, the creation of artificial, multifunctional, lignocellulosic hydrolases is a realistic and practical approach for the improvement of biomass conversion.

A key consideration in artificial construction of multifunctional enzymes is the preservation or improvement of the protein and enzymatic characteristics of the individual components. It is a challenge to create a chimera that possesses optimal enzyme functions for one or more parental enzymes. The following factors may need to be considered for this purpose:

- The order of the connected enzymatic modules;
- The composition of the CBMs;
- The types and lengths of intermodular linker peptides.

Cell-anchored enzyme systems

Free enzyme systems of aerobic microorganisms are characterized by enzymes, which are usually secreted directly into the extracellular milieu in relatively large quantities and with essentially no residence time on

the microbial cell surface. In contrast, the cellulosome resides on the cell surface and is released into the extracellular medium only at later stages in the growth cycle. Consequently, the cellulosome is essentially a cell-surface entity. Indeed, as described above, the presence of a tenacious, cellulose-binding CBM on the scaffoldin subunit also serves to mediate binding of the entire bacterial cell to its cellulosic substrates prior to their enzymatic hydrolysis. In fact, the presence of this CBM and the fact that the cellulosome is attached to the cell are the major reasons that led to the initial discovery of the cellulosome [87–89].

Both cellulosomal and noncellulosomal enzymes can be attached to the cell surface, using several possible mechanisms **(Figure 4)**. In *C. thermocellum,* the cellulosome is attached to the bacterial cell surface via a special class of anchoring proteins **(Figure 4B)**, which can be defined as scaffoldins by virtue of a second, divergent type of cohesin that they possess in various copies (1 to 7 in *C. thermocellum*) at their N-terminus. At their C-terminus, these anchoring scaffoldins contain an S-layer homology (SLH) module that mediates attachment to the bacterial cell surface **(Table 2)** [90–95].

Interestingly, at least two *C. thermocellum* enzymes, a xylanase and a lichenase, also contain a C-terminal SLH module, indicating their individual attachment to the cell surface **(Table 4 & Figure 4A)**. Likewise, SLH modules are components of glycoside hydrolases from several other plant cell wall-degrading bacteria. In *R. flavefaciens*, however, the cellulosome is attached covalently to the cell surface via the ScaE scaffoldin [43], which contains a sortase signal motif at its C-terminus **(Table 2 & Figure 4C)**. Genome sequencing of this bacterium has revealed several other putative structural proteins that include a similar sortase signal motif, indicating that the host proteins are potential cell surface components. At least one of these proteins includes a module whose sequence is consistent with a GH10 xylanase.

More recently, an alternative mechanism of attachment of plant cell wall-degrading enzymes to the surfaces of their parent bacterial cells has been documented. In this context, numerous CBMs were discovered recently in the rumen bacterium, *R. albus*. This novel type of CBM, eventually classified as family 37, is found exclusively in this bacterium. Indeed, *R. albus* remains an intriguing case. In the draft genome sequence of *R. albus* strain 8 (~90% coverage), approximately 40 proteins contain at least one CBM37 module and half of these proteins are classified as putative carbohydrate-acting enzymes (glycoside hydrolases, pectate lyases and carbohydrate esterases) [44]. Likewise, a similar number of dockerin-containing proteins have been detected. However, to date, no cohesin module has been sequenced or unambiguously identified in *R. albus*. This would either indicate that there may be cohesin sequences in the residual (~10%) unsequenced portion of the genome or that there are, in fact, no cohesins in this strain and hence no scaffoldins. If the latter case is true, then why would the bacterium produce dockerin-containing proteins? The answer may lie in the nature of the rumen environment in which this bacterium inhabits. We have learned recently that the related rumen bacterium, *R. flavefaciens*, populates the bovine rumen in multiple strains, which may each produce different types of cellulosomal components [96]. If *R. albus* also follows this pattern, then the possibility exists that this particular strain (strain 8) may produce and secrete various dockerin-bearing proteins, while another (or other) strain(s) may produce complementary cohesin-containing scaffoldins. Future genome sequencing of other *R. albus* strains should provide insights into whether this bacterium can be considered a true cellulosome-producing microorganism.

The CBM37s were initially discovered in *R. albus* on the basis of adhesion-defective strains that lacked specific surface proteins [97]. Surprisingly, these proteins were subsequently identified as the critical GH48 enzyme and a GH9 enzyme, both of which were noncellulosomal but carried a module of unknown function on their C-terminus. Subsequent studies revealed that the latter type of module was a CBM [98], which was then classified in a new CBM family (family 37). The CBM37s were collectively shown to exhibit a broad specificity pattern, which indicated a mechanism for binding the parent enzymes to cellulosic substrates. The anchoring of the CBM37-bearing enzymes to the bacterial cell surface was later demonstrated [99] – presumably to cell envelope polysaccharides **(Figure 4D)**, thus indicating a multiplicity of roles (combined substrate and cell-surface binding) for this fascinating type of CBM. It seems that the cell-attachment role is not confined to the CBM37s, since a similar role has also been demonstrated for at least one member of another CBM family (family 35) [100]. Moreover, the status of a conserved type of module from anaerobic fungi, long being considered a 'dockerin' [101–105], has recently been demonstrated to bind to saccharide components of a surface-attached GH3 β-glucosidase [106]. On the one hand, the attachment of a multiplicity of cellulolytic enzymes onto a β-glucosidase is very logical, since the concerted breakdown of cellulose into soluble cellodextrins in the presence of such an enzyme would presumably counteract product inhibition of the cellulases and promote efficient hydrolysis of cellulosic substrates. On the other hand, categorizing such an enzyme complex as a cellulosome is probably invalid.

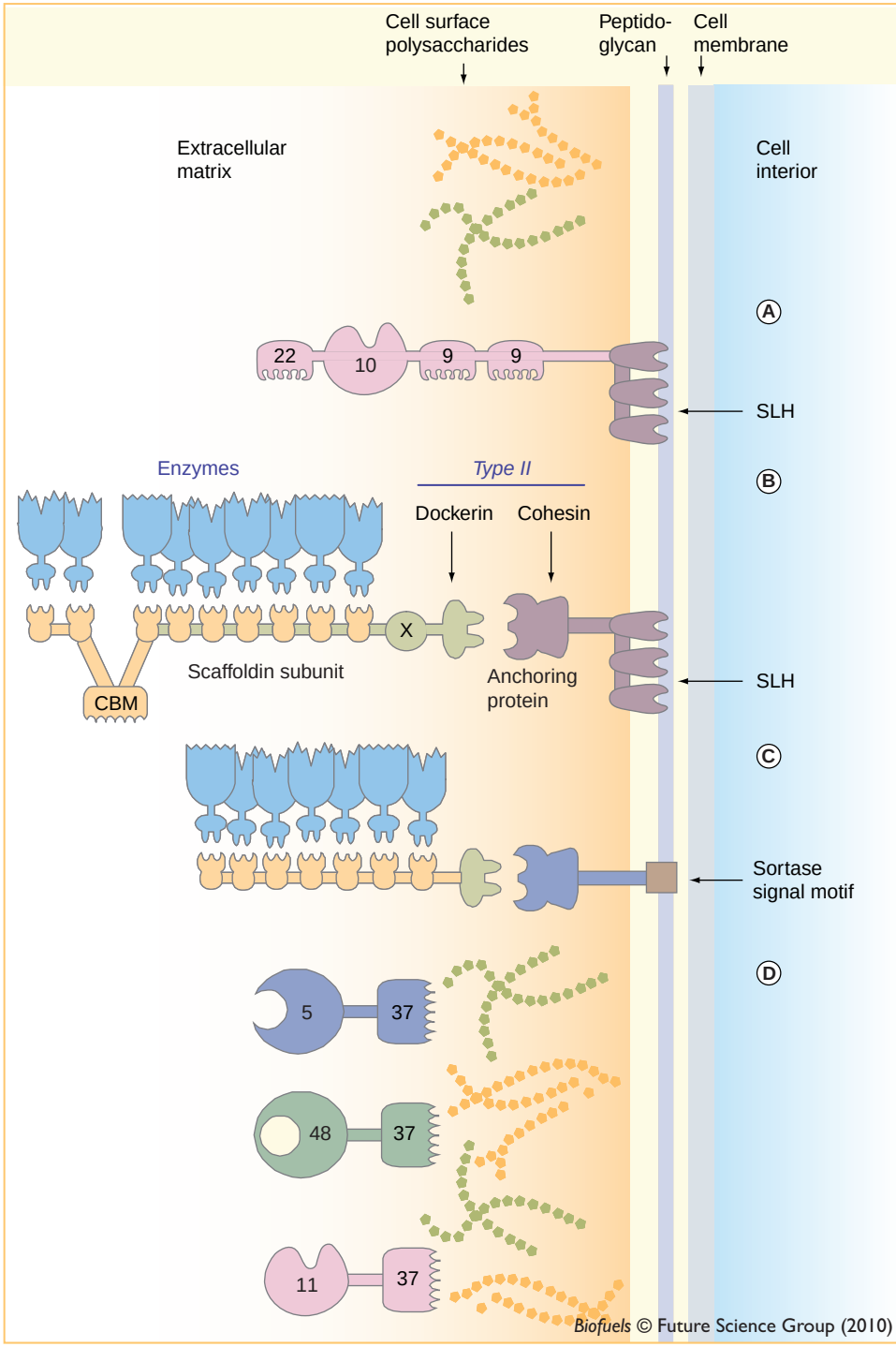

Figure 4. Modes of attachment of glycoside hydrolases and cellulosomes to the bacterial cell surface. (A) Some single enzymes bear SLH modules, which can interact noncovalently either with the peptidoglycan layer[93,95] or with secondary cell wall polymers [94] and the enzymes are thereby attached individually to the cell surface. The cell-surface enzyme shown schematically in the figure is based on *Clostridium thermocellum* Xyn10X that also contains family 9 and 22 CBMs. **(B)** Several cellulosomes are known to be attached to the cell wall via a special type of cohesin-containing anchoring scaffoldin that also carries an SLH module [93,95]. The type-II anchoring cohesins interact with a complementary type-II dockerin of the enzyme-bearing scaffoldin subunit. **(C)** The *Ruminococcus flavefaciens* cellulosome is attached to the cell surface via covalent linkage to the peptidoglycan layer via an enzyme-mediated process. At least one enzyme from the same bacterium is also known to contain a similar sortase-signal motif, which may thus represent a more common mode of enzyme attachment to the bacterial cell surface. **(D)** An alternative mode of enzyme attachment to the cell surface involves special types of CBM, such as family-37 CBMs from *Ruminococcus albus,* that mediate individual noncovalent attachment of different enzymes to cell-surface polysaccharides. CBM: Cellulose-binding module; SLH: S-layer homology.

If the β-glucosidase is a glycoprotein and the alleged dockerins recognize its saccharide moieties and not a *bona fide* cohesin, then the said dockerin would be a CBM and not a dockerin *per se*. Consequently, the fungal 'cellulosomes' cannot be considered authentic cellulosomes but would represent a completely novel type of CBM-mediated multienzyme complex and unique type of paradigm for biomass degradation.

Mixed, exotic & undiscovered paradigms

It is clear from the above sections that the division into distinct paradigms is not necessarily strict. Microbes capable of using lignocellulosic substrates and perhaps the bacteria in particular, must employ particularly 'intelligent' strategies to survive in often extreme environments. To this end, the cellulolytic microbes exploit to their advantage every possible

weapon and scheme they possess in their molecular and cellular arsenal.

As we have observed, the anaerobic thermophilic cellulolytic bacterium, *C. thermocellum*, is highly reputed for its intricate cellulosome system – the first to have been described. However, the bacterium also produces several free cellulases, both with and without an ancillary substrate-targeting CBM. Likewise, its mesophilic relative, *C. cellulovorans*, also produces both a well-characterized cellulosome and noncellulosomal enzymes, which are believed to work together in a synergistic manner. In addition, *C. thermocellum* can produce several cell-anchored hemicellulases, attached to the cell surface via an SLH module. Finally, many of the *C. thermocellum* and *R. flavefaciens* cellulosomal enzymes are multifunctional, some having two or three different catalytic modules as well as several ancillary modules in addition to the dockerin – up to a total of seven different modular components in at least one case. Therefore, some cellulolytic bacteria and *C. thermocellum* in particular, can apparently exploit mixed strategies in order to decompose cellulosic substrates in an efficient manner.

The above-described strategies for microbial attack on lignocellulosic materials are not the only ones that can be used in nature. As we proceed in our study of microbial cellulolytic systems, it becomes clear that some bacteria and fungi fail to conform to our current dogma. For example, a completely different strategy and newly emerging paradigm is evident in the xylanolytic system of *Geobacillus stearothermophilus* T-6, in which the relevant genes are arranged on a large (~40 kb) chromosome cluster. When grown on xylan, the bacterium secretes a lone extracellular endoxylanase that cleaves the main chain of the substrate, thereby producing short, branched xylodextrin units of two or more sugars. These unusual branched xylosaccharides are then taken up into the cell via highly specialized ABC sugar transporter systems, which are lacking in other competing bacteria [107]. Once inside the cell, the branched xylosaccharides are subjected to further degradation by a cadre of intracellular enzymes, which include a GH10 xylanase and associated enzymes from GH39, GH43, GH51, GH52, GH67 and CE4. The imported carbohydrates are thus converted to monosaccharides, which are then assimilated by the usual metabolic pathways. In this manner, the bacterium competes with other microbes successfully for its fair share of the xylan in its ecosystem. The connection between the specificities) of the hydrolytic enzymes and characteristics of their products, to those of the sugar-binding lipoprotein components of the transporter systems, appears to be a more general phenomenon in the lignocellulolytic bacteria. In *C. thermocellum*, for example, the sugar specificities of the different binding lipoproteins are consistent with the observed substrate preference, whereby the larger cellodextrins are assimilated faster than cellobiose [108].

In truth, we may only be at the beginning of our renewed quest to understand the different lignocellulosic-degrading paradigms in nature. It seems that none of the above-described strategies are consistent with the highly efficient cellulose-degrading systems of some other bacteria, such as *Cytophaga hutchinsonii* and *Fibrobacter succinogenes* [109], both of which have been subjects of recent genome sequencing projects. Both of these bacteria appear to lack known processive cellulases, most of their cellulases appear to lack CBMs and none of them bear dockerins. Further analyses of the different cellulolytic bacteria and fungi and intensive programs for characterization of their enzymes will be necessary for discovery of additional mechanisms of microbial degradation of lignocellulosic substrates. Future genome sequencing programs that focus on plant cell wall degrading microorganisms will undoubtedly reveal new and exotic paradigms.

Key terms

Hybrid hydrolysis and fermentation: Process that permits staging of the saccharification and fermentation steps, often designed to allow a high temperature enzyme treatment followed by a lower temperature fermentation step

Simultaneous saccharification and fermentation: Process that combines the hydrolysis of cellulose and other plant cell wall polysaccharides with the fermentation of the sugars released

Consolidated bioprocessing: Combines cellulase production, substrate hydrolysis and fermentation into a single step, by employing a single microorganism capable of both expressing cellulolytic enzymes and converting soluble sugars to ethanol

Biomass conversion schemes

Today, biomass conversion is based on the fermentation of biomass sugars to biofuels, especially ethanol. The primary unit operations for processing lignocellulose biomass include biomass pretreatment, hydrolysis of cellulose and hemicelluloses to monosaccharides and fermentation of these sugars to liquid fuels products [110]. The following general strategies have been used for the conversion of biomass to ethanol by fermentation: **hybrid hydrolysis and fermentation** (HHF), **simultaneous saccharification and fermentation** (SSF), simultaneous saccharification and co-fermentation (SSCF) and **consolidated bioprocessing** (CBP). HHF allows saccharification to proceed at different temperatures or for longer times, for example, than fermentation. SSCF was introduced more recently and accommodates new and primarily engineered fermentative stains able to convert all biomass sugars to ethanol with equal efficiency. HHF, SSF and SSCF currently require extensive pretreatment strategies of the cellulosic feedstock and the addition of exogenous cellulolytic enzymes. The capital cost of pretreatment and utilization of cellulase enzymes are still the main barriers to enabling cost-effective bioethanol production.

Potential single-step conversion of biomass to biofuels, afforded by CBP strategies [111–113], would thus provide an ideal approach for solving the biomass conversion cost problem. Emerging evidence today is beginning to demonstrate the feasibility of CBP at the industrial scale [114].

Why are better cellulases needed?

Regardless of the biomass conversion scheme eventually employed in industrial biofuels plants, the catalytic efficiency of the glycoside hydrolases utilized, also known as specific activity, must be near theoretical values if the process is to be cost effective and robust. This consideration is especially true of the cellulases. As long as deconstruction remains a biochemical-based process, the cost of cellulose-degrading enzymes will remain a major process cost [110,115,116]. The reason for this outcome is based in the very nature of heterogeneous catalysis. Enzymes that act upon insoluble polymeric substrates (e.g., cellulose, chitin and protein) typically display lower catalytic rates (k_{cat}) than enzymes acting upon soluble substrates [117,118]. This situation results in very high protein-based loadings for cellulases relative to other glycoside hydrolases. For example, 25 mg of a *Trichoderma reesei* cellulase preparation (Spezyme or GC220) converts 80% of a 1-g sample of Avicel to glucose in 5 days at 50°C. Whether the cellulases are free (noncomplexed) or complexed (cellulosomal), the impact on process cost is important. Free cellulases are expected to be produced 'off site' at special enzyme production facilities and will be priced based on performance, whereas cellulosomal enzymes or other enzymes produced during CBP have a carbon-demand cost, which taxes both metabolic flux and product formation. Regarding sales margins of free enzymes, profits from sales of these enzymes can be enhanced if the specific performance of the enzymes is increased while keeping the cost of production the same; consequently, pricing will be based on performance. Optimizing the performance of the molecular catalysts will permit other unit operations in the plant to undergo excursions of inefficiency, a very likely scenario when one considers the uncertainty in feedstock type and quality that the processing plant will encounter.

Future perspective

To date, only glimpses of the picture regarding the molecular mechanisms of plant cell wall-degrading enzymes have been accomplished. It is expected that new enzymes and new paradigms will be found, based on aggressive genome sequencing and analysis of cell wall-degrading microorganisms. Metagenomic sequencing will also help in this regard, especially when teamed with new strategies for understanding the roles of participating microbes in masses of decaying biomass, including those strains that are today 'uncultureable' [119,120]. Although many hydrolytic enzymes have been found, all indications point to a vast reservoir of yet-to-be-discovered analogs in the biosphere. Building on this concept and even more important than discovery of another ten examples of GH10, for example, would be the discovery of entirely new paradigms of biological cell wall hydrolysis or molecular fragmentation. Of course, we need to know more about the structure–function relationships of glycoside hydrolases in general and in the case of cellulases these studies are further hampered by a lack of good structural data for cellulose. We can also expect continued utilization of diverse disciplines in science to be brought to the problem of cell wall hydrolysis. One outstanding example of this is the recent linkage between computational science and classical biochemistry in solving cellulase mechanistic problems [121,122]. The advent of more accessible supercomputers running thousands of processors, improved (more scalable) codes for molecular dynamics and quantum mechanics/molecular mechanics and better force fields for biological problems (CMAP for proteins [123] and TeamSugar for carbohydrates [124]), makes the acquisition of biologically relevant simulations of cellulose hydrolysis more promising. When teamed with mutational biochemistry and high-throughput robotics enzyme assays, the aggressive time line for improving existing enzymes and even developing biomimetic counterparts seems imminently possible.

It has been 60 years since the Army Natick Laboratories first isolated *T. reesei* from the rotting cotton military accruements sent back from the war in the south Pacific and it remains remarkable that the precise molecular mechanisms of action of the enzymes produced by this and related lignocellulose degrading fungi are as yet unknown. In many ways, the diversity and probable extensive bandwidth of plant cell wall-degrading enzymatic and microbial systems harbored by the biosphere remains a mystery. However, this mystery we must solve. Terrestrial plant ecosystems produce carbonaceous compounds from CO_2 in a highly oxygenated form, the polysaccharides, and store these polymers in both metabolically accessible (storage) and inaccessible (structural) forms. In many ecosystems, dead plant matter takes years if not centuries to fully degrade; an indicator that the biosphere can tolerate removal of some fraction of carbon with no ill effects. Lignocellulose, therefore, is the best feedstock for a sustainable and reliable source of liquid fuels. We just need a fully cost effective and robust conversion scheme to produce fermentable sugars from biomass, such as corn stover and other crop residues, hard and soft woods, construction waste and perhaps even municipal solid waste. A

glimpse at the recent literature in this field confirms that progress towards gaining a deep understanding of these processes is building momentum. With continued focus on biomass conversion science, we should enjoy new sources of biofuels for centuries to come.

Financial & competing interests disclosure

This work was supported by the US Department of Energy Office of Science, Office of the Biological and Environmental Research, through the BioEnergy Science Center (BESC), a DOE Bioenergy Research Center, by grants from the Israel Science Foundation (grant nos 966/09 and 159/07) and from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel. Edward A Bayer is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry at the Weizmann Institute of Science. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Lignocellulosic biomass

- Most abundant source of carbon and energy on Earth.
- Attractive alternative to current petroleum-based fossil fuels.
- Resistant to chemical and enzymatic conversion: high cost of conversion the major deterrent for commercialization.
- Lack of knowledge of molecular architecture of the plant cell wall polymers.

Free enzyme systems

- Carbohydrate-active enzymes include cellulases, hemicellulases (e.g., xylanases, mannanases and arabinofuranases), pectate lyases and carbohydrate esterases that together deconstruct plant cell wall polysaccharides in an efficient manner.
- Cellulases and hemicellulases are classified as glycoside hydrolases: multimodular enzymes from 115 different families, which contain a catalytic module that cleaves the glycoside bond and (usually) a carbohydrate-binding module (CBM) that targets the enzyme to the (poly)saccharide substrate.
- Cellulases include endo- and exo-acting enzymes, together with β-glucosidases, which work synergistically to hydrolyze the particularly resistant crystalline cellulose fibrils.

Cellulosomes

- Noncatalytic scaffoldins contain a CBM for substrate targeting and multiple cohesin modules for integrating the enzymatic subunits into a multicomponent complex.
- Cellulosomal enzymes contain a dockerin module that binds strongly to the scaffoldin-borne cohesins.
- Synergistic action is achieved by concentrating the enzymes together at defined sites on the lignocellulosic substrate.
- Artificial 'designer cellulosomes' may provide a future solution for efficient conversion of cellulosic biomass to soluble sugars.

Multifunctional enzyme systems

- Composed of two or more catalytic modules on the same polypeptide chain.
- May be cellulosomal (dockerin-containing) or noncellulosomal (CBM-containing).
- Some cellulolytic hyperthermophiles contain numerous multifunctional enzymes with multiple CBMs.
- This 'natural' strategy may serve as a concept for artificial chimeric multifunctional enzymes.

Cell-anchored enzyme systems

- May be cellulosomal (scaffoldin-attached) or noncellulosomal (direct attachment of enzymes to the microbial cell surface).
- Cell-surface attachment may be mediated by noncovalent or covalent interaction.

Mixed, exotic & undiscovered paradigms

- Lignocellulolytic microbes that have evolved employ a diverse set of strategies for degradation of plant cell wall biomass.
- It is logical to expect that the biosphere harbors new paradigms for biomass deconstruction not yet discovered and the lessons learned from their analysis will greatly benefit biomass conversion science.

Biomass conversion schemes

- In order to make biomass conversion schemes cost effective, sustainable and attractive to industry, the deconstruction of plant cell walls must be fully understood at the molecular level.
- New biomass-processing schemes, based on past experience and future innovation, will optimize biocatalyst performance, both enzymatic and microbial.

Why are better cellulases needed?

- Capture of fermentable sugars from plants, either crop residues or plants grown for energy purposes, is essential to maintain our thriving world energy economy.
- Dramatic progress in the basic sciences of plant recalcitrance can be traced to new programs supporting multidisciplinary research with long-term goals.
- Ultimately, mankind will have the knowledge to manipulate the biosphere with the skills needed to safely withdraw plant biomass and convert it into much needed fuels and chemicals.

Bibliography

Papers of special note have been highlighted as:

■ of interest

■■ of considerable interest

1 Wei H, Xu Q, Taylor LE, Baker JO, Tucker MP, Ding SY. Natural paradigms of plant cell wall degradation. *Curr. Opin. Biotechnol.* 20, 330–338 (2009).

2 Perlack RD, Wright LL, Turhollow AF, Graham RL, Stokes BJ, Erbach DC. Biomass as feedstock for a bioenergy and bioproduct industry: the technical feasibility of a billion-ton annual supply. In: *A joint study sponsored by the US Department of Energy and the US Department of Agriculture*. Oak Ridge National Laboratory, TN, USA Oak Ridge (2005).

3 Rose JKC. *The Plant Cell Wall (Volume 8)*. Blackwell Publishing, Garsington, UK (2003).

4 Ding SY, Himmel ME. The maize primary cell wall microfibril: a new model derived from direct visualization. *J. Agric. Food Chem.* 54, 597–606 (2006).

5 Sugiyama J, Vuong R, Chanzy H. Electron diffraction study of the two crystalline phases occurring in native cellulose from an algal cell wall. *Macromolecules* 24, 4168–4175 (1991).

6 Atalla RH, VanderHart DL. Native cellulose: a composite of two distinct crystalline forms. *Science* 223, 283–285 (1984).

7 Nishiyama Y SJ, Chanzy H, Langan P. Crystal structure and hydrogen bonding system in cellulose Iα from synchrotron X-ray and neutron fiber diffraction. *J. Am. Chem. Soc.* 124, 14300–14306 (2003).

8 Boraston AB. The interaction of carbohydrate-binding modules with insoluble non-crystalline cellulose is enthalpically driven. *Biochem. J.* (2004).

9 Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate–Active Enzymes database (CAZy): an expert resource for glycogenomics. *Nucl. Acids Res.* 37, D233–238 (2009).

10 Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853–859 (1995).

11 Rouvinen J, Bergfors T, Teeri T, Knowles JKC, Jones TA. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* 279, 380–386 (1990).

12 Divne C, Ståhlberg J, Reinikainen T *et al*. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* 265, 524–528 (1994).

13 Chanzy H, Henrissat B. Unidirectional degradation of *Valonia* cellulose microcrystals subjected to cellulase action. *FEBS Lett.* 184, 285–288 (1985).

14 Imai T, Boisset C, Samejima M, Igarashi K, Sugiyama J. Unidirectional processive action of cellobiohydrolase Cel7A on Valonia cellulose microcrystals. *FEBS Lett.* 432, 113–116. (1998).

15 Teeri TT. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends Biotechnol.* 15, 160–167 (1997).

16 Li Y, Irwin DC, Wilson DB. Processivity, substrate binding, and mechanism of cellulose hydrolysis by *Thermobifida fusca* Cel9A. *Appl. Environ. Microbiol.* 73, 3165–3172 (2007).

17 Tenkanen M, Poutanen K. Significance of esterases in the degradation of xylans. In: *Xylans and Xylanases*. Visser J, Beldman G, Kusters-van Someren MA, Voragen AGJ (Eds). Elsevier, UK, 203–212 (1992).

18 Thomas SR, Laymon RA, Chou Y-C *et al*. Initial approaches to artificial cellulase systems for conversion of biomass to ethanol. In: *Bioconversion of Lignocellulosics, ACS Book Series 618*. Saddler JN, Penner M (Eds). American Chemical Society, USA (1995).

19 Puls J, Tenkanen M, Korte HE, Poutanen K. Products of hydrolysis of beechwood acetyl-4-O-methylglucuronoxylan by a xylanase and an acetyl xylan esterase. *Enzyme Microb. Technol.* 13, 483–486 (1991).

20 Greve LC, Labavitch JM, Hungate RE. α-L-arabinofuranosidase from *Ruminococcus albus* 8: purification and possible role in hydrolysis of alfalfa cell wall. *Appl. Environ. Microbiol.* 47, 1135–1140 (1984).

21 Johnson KG, Silva MC, MacKenzie CR, Schneider H, Fontana JD. Microbial degradation of hemicellulosic materials. *Appl. Biochem. Biotechnol.* 20/21, 245–258 (1989).

22 Bayer EA, Belaich J-P, Shoham Y, Lamed R. The cellulosomes: multi-enzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* 58, 521–554 (2004).

23 Demain AL, Newcomb M, Wu JH. Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.* 69, 124–154 (2005).

24 Doi RH, Kosugi A. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat. Rev. Microbiol.* 2, 541–551 (2004).

25 Schwarz WH, Zverlov VV, Bahl H. Extracellular glycosyl hydrolases from Clostridia. *Adv. Appl. Microbiol.* 56, 215–261 (2004).

26 Raman B, Pan C, Hurst GB *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).

27 Hammel M, Fierobe H-P, Czjzek M *et al*. Structural basis of cellulosome efficiency explored by small angle X-ray scattering. *J. Biol. Chem.* 280, 38562–38568 (2005).

28 Hammel M, Fierobe HP, Czjzek M, Finet S, Receveur-Brechot V. Structural insights into the mechanism of formation of cellulosomes probed by small angle X-ray scattering. *J. Biol. Chem.* 279, 55985–55994 (2004).

29 Carvalho AL, Dias FM, Prates JA *et al*. Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex. *Proc. Natl Acad. Sci. USA* 100, 13809–13814 (2003).

30 Carvalho AL, Dias FMV, Nagy T *et al*. Evidence for a dual binding mode of dockerin modules to cohesins. *Proc. Natl Acad. Sci. USA* 104, 3089–3094 (2007).

■■ **The dual binding of the dockerin module to cohesin was demonstrated experimentally for the first time; crucial to elucidate the mechanism of cellulosome action, particularly with reference to its purported plasticity.**

31 Pinheiro BA, Proctor MR, Martinez-Fleites CC *et al*. The *Clostridium cellulolyticum* dockerin displays a dual binding mode for its cohesin partner. *J. Biol. Chem.* 283, 18422–18430 (2008).

32 Adams JJ, Pal G, Jia Z, Smith SP. 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).

33 Bayer EA, Lamed R, White BA, Flint HJ. From cellulosomes to cellulosomics. *Chem. Rec.* 8, 364–377 (2008).

34 Shoseyov O, Takagi M, Goldstein MA, Doi RH. Primary sequence analysis of *Clostridium cellulovorans* cellulose binding protein A. *Proc. Natl Acad. Sci. USA* 89, 3483–3487 (1992).

35 Gerngross UT, Romaniec MPM, Kobayashi T, Huskisson NS, Demain AL. 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).

36 Pagès S, Belaich A, Fierobe H-P, Tardif C, Gaudin C, Belaich J-P. 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).

37 Kakiuchi M, Isui A, Suzuki K *et al*. Cloning and DNA sequencing of the genes encoding *Clostridium josui* scaffolding protein CipA and cellulase CelD and identification of their gene products as major components of the cellulosome. *J. Bacteriol.* 180, 4303–4308 (1998).

38 Ding S-Y, Bayer EA, 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).

39 Ding S-Y, Rincon MT, Lamed R *et al*. Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *J. Bacteriol.* 183, 1945–1953 (2001).

40 Ding S-Y, Lamed R, Bayer EA, Himmel ME. The bacterial scaffoldin: Structure, function and potential applications in nanoscience. In: *Genetic Engineering: Principles and Methods*. Setlow JK (Eds). Kluwer Academic Publishers, Germany, 209–226 (2003).

41 Xu Q, Gao W, Ding S-Y *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).

42 Xu Q, Bayer EA, Goldman M, Kenig R, Shoham Y, Lamed R. Architecture of the *Bacteroides cellulosolvens* cellulosome: description of a cell-surface anchoring scaffoldin and a family-48 cellulase. *J. Bacteriol.* 186, 968–977 (2004).

43 Rincon MT, Cepeljnik T, Martin JC *et al*. Unconventional mode of attachment of the *Ruminococcus flavefaciens* cellulosome to the cell surface. *J. Bacteriol.* 187, 7569–7578 (2005).

44 Morrison M, Daugherty SC, Nelson WC, Davidsen T, Nelson KE. The FibRumBa database: a resource for biologists with interests in gastrointestinal microbial ecology, plant biomass degradation, and anaerobic microbiology. *Microb. Ecol.* PMID 19609599, (2009).

45 Berg Miller ME, Antonopoulos DA, Rincon MT *et al*. Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PLoS ONE* 4, e6650 (2009).

▪ **Genomic analysis of a key rumen cellulolytic bacterium, producing the most sophisticated cellulosome yet found in nature.**

46 Flint HJ, Bayer EA. Plant cell wall breakdown by anaerobic bacteria from the mammalian digestive tract. *Ann. NY Acad. Sci.* 1125, 280–288 (2008).

47 Flint HJ, Bayer EA, Lamed R, White BA. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Rev. Microbiol.* 6, 121–131 (2008).

48 Bayer EA, Setter E, Lamed R. Organization and distribution of the cellulosome in *Clostridium thermocellum*. *J. Bacteriol.* 163, 552–559 (1985).

49 Wang WK, Kruus K, Wu JHD. Cloning and DNA sequence of the gene coding for *Clostridium thermocellum* cellulase S_S (CelS), a major cellulosome component. *J. Bacteriol.* 175, 1293–1302 (1993).

50 Dror TW, Morag E, Rolider A, Bayer EA, Lamed R, Shoham Y. Regulation of the cellulosomal *cel*S (*cel*48A) gene of *Clostridium thermocellum* is growth-rate dependent. *J. Bacteriol.* 185, 3042–3048 (2003).

51 Berger E, Zhang D, Zverlov VV, Schwarz WH. Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. *FEMS Microbiol. Lett.* 268, 194–201 (2007).

52 Jindou S, Xu Q, Kenig R, Shoham Y, Bayer EA, Lamed R. Novel architectural theme of family-9 glycoside hydrolases identified in cellulosomal enzymes of *Acetivibrio cellulolyticus* and *Clostridium thermocellum*. *FEMS Microbiol. Lett.* 254, 308–316 (2006).

53 McGrath CE, Wilson DB. Endocellulolytic activity of the *Clostridium thermocellum* Cel9C (formerly CbhA) catalytic domain. *Indust. Biotechnol.* 4, 99–104 (2008).

54 Arai T, Kosugi A, Chan H *et al*. Properties of cellulosomal family 9 cellulases from *Clostridium cellulovorans*. *Appl. Microbiol. Biotechnol.* 71, 654–660 (2006).

55 Dror TW, Rolider A, Bayer EA, Lamed R, Shoham Y. Regulation of expression of scaffoldin-related genes in *Clostridium thermocellum*. *J. Bacteriol.* 185, 5109–5116 (2003).

56 Dror TW, Rolider A, Bayer EA, Lamed R, Shoham Y. Regulation of major cellulosomal endoglucanases of *Clostridium thermocellum* differs from that of a prominent cellulosomal xylanase. *J. Bacteriol.* 187, 2261–2266 (2005).

57 Zverlov VV, Kellermann J, Schwarz WH. Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: Identification of the major catalytic components in the extracellular complex and detection of three new enzymes. *Proteomics* 5, 3646–3653 (2005).

58 Stevenson DM, Weimer PJ. Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or cellobiose in continuous culture. *Appl. Environ. Microbiol.* 71, 4672–4678 (2005).

59 Gold ND, Martin VJ. Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J. Bacteriol.* 189, 6787–6795 (2007).

60 Bayer EA, Morag E, Lamed R. The cellulosome – a treasure-trove for biotechnology. *Trends Biotechnol.* 12, 378–386 (1994).

61 Fierobe H-P, Mechaly A, Tardif C *et al*. Design and production of active cellulosome chimeras: selective incorporation of dockerin-containing enzymes into defined functional complexes. *J. Biol. Chem.* 276, 21257–21261 (2001).

62 Fierobe H-P, Mingardon F, Mechaly A *et al*. Action of designer cellulosomes on homogeneous versus complex substrates: Controlled incorporation of three distinct enzymes into a defined tri-functional scaffoldin. *J. Biol. Chem.* 280, 16325–16334 (2005).

63 Fierobe H-P, Bayer EA, Tardif C *et al*. Degradation of cellulose substrates by cellulosome chimeras: substrate targeting versus proximity of enzyme components. *J. Biol. Chem.* 277, 49621–49630 (2002).

64 Mingardon F, Chanal A, López-Contreras AM, Dray C, Bayer EA, Fierobe H-P. Incorporation of fungal cellulases in bacterial minicellulosomes yields viable, synergistically acting cellulolytic complexes. *Appl. Environ. Microbiol.* 73, 3822–3832 (2007).

65 Caspi J, Irwin D, Lamed R *et al*. *Thermobifida fusca* family-6 cellulases as potential designer cellulosome components. *Biocat. Biotransform.* 24, 3–12 (2006).

66 Caspi J, Irwin D, Lamed R, Fierobe H-P, Wilson DB, Bayer EA. Conversion of noncellulosomal *Thermobifida fusca* free exoglucanases into cellulosomal components: comparative impact on cellulose-degrading activity. *J. Biotechnol.* 135, 351–357 (2008).

67 Caspi J, Barak Y, Haimovitz R *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).

68 Caspi J, Barak Y, Haimovitz R *et al*. *Thermobifida fusca* exoglucanase Cel6B is incompatible with the cellulosomal mode in contrast to endoglucanase Cel6A. *Syst. Synth. Biol.* (In Press) (2010).

69 Mingardon F, Chanal A, Tardif C, Bayer EA, Fierobe H-P. Exploration of new geometries in cellulosome-like chimeras. *Appl. Environ. Microbiol.* 73, 7138–7149 (2007).

▪ **Creative study whereby extraordinary architectures are explored in the construction of designer cellulosomes of novel content.**

70 Perret S, Casalot L, Fierobe H-P *et al*. Production of heterologous and chimeric scaffoldins by *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* 186, 253–257 (2004).

71 Mingardon F, Perret S, Belaich A, Tardif C, Belaich JP, Fierobe HP. Heterologous production, assembly, and secretion of a minicellulosome by *Clostridium acetobutylicum* ATCC 824. *Appl. Biochem. Biotechnol.* 71, 1215–1222 (2005).

72 Perret S, Belaich A, Fierobe HP, Belaich JP, Tardif C. Towards designer cellulosomes in Clostridia: mannanase enrichment of the cellulosomes produced by *Clostridium cellulolyticum*. *J. Bacteriol.* 186, 6544–6552 (2004).

73 Levasseur A, Pages S, Fierobe HP *et al*. Design and production in *Aspergillus niger* of a chimeric protein associating a fungal feruloyl esterase and a clostridial dockerin domain. *Appl. Environ. Microbiol.* 70, 6984–6991 (2004).

74 Tsai SL, Oh J, Singh S, Chen R, Chen W. Functional assembly of minicellulosomes on the *Saccharomyces cerevisiae* cell surface for cellulose hydrolysis and ethanol production. *Appl. Environ. Microbiol.* 75, 6087–6093 (2009).

75 Lilly M, Fierobe HP, van Zyl WH, Volschenk H. Heterologous expression of a *Clostridium* minicellulosome in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 9(8), 1236–1249 (2009).

76 Kataeva IA, Yang SJ, Dam P *et al*. Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium *Anaerocellum thermophilum*. DSM 6725. *J. Bacteriol.* 191, 3760–3761 (2009).

77 Yang SJ, Kataeva I, Hamilton-Brehm SD *et al*. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe *Anaerocellum thermophilum* DSM 6725. *Appl. Environ. Microbiol.* 75, 4762–4769 (2009).

78 Boraston AB, Nurizzo D, Notenboom V *et al*. Differential oligosaccharide recognition by evolutionarily-related β-1,4 and β-1,3 glucan-binding modules. *J. Mol. Biol.* 319, 1143–1156 (2002).

79 Najmudin S, Guerreiro CI, Carvalho AL *et al*. Xyloglucan is recognized by carbohydrate-binding modules that interact with β-glucan chains. *J. Biol. Chem.* 281, 8815–8828 (2006).

80 Gibbs MD, Reeves RA, Farrington GK, Anderson P, Williams DP, Bergquist PL. Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1. *Curr. Microbiol.* 40, 333–340 (2000).

81 Gibbs MD, Saul DJ, Luthi E, Bergquist PL. The β-mannanase from *Caldocellum saccharolyticum* is part of a multidomain enzyme. *Appl. Environ. Microbiol.* 58, 3864–3867 (1992).

82 Zverlov V, Mahr S, Riedel K, Bronnenmeier K. Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile '*Anaerocellum thermophilum*' with separate glycosyl hydrolase family 9 and 48 catalytic domains. *Microbiology* 144, 457–465 (1998).

83 Hong SY, Lee JS, Cho KM *et al*. Assembling a novel bifunctional cellulase-xylanase from *Thermotoga maritima* by end-to-end fusion. *Biotechnol. Lett.* 28, 1857–1862 (2006).

▪ **New notion for generation of artificial multifunctional plant cell wall-degrading enzymes for efficient degradation of lignocellulosics.**

84 Khandeparker R, Numan MT. Bifunctional xylanases and their potential use in biotechnology. *J. Ind. Microbiol. Biotechnol.* 35, 635–644 (2008).

85 Fan ZM, Wagschal K, Chen W, Montross MD, Lee CC, Yuan L. Multimeric hemicellulases facilitate biomass conversion. *Appl. Environ. Microbiol.* 75, 1754–1757 (2009).

86 Fan ZM, Wagschal K, Lee CC *et al*. The construction and characterization of two xylan-degrading chimeric enzymes. *Biotechnol Bioeng* 102, 684–692 (2009).

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

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

89 Bayer EA, Lamed R. The cellulosome saga: early history. In: *Cellulosome*. Uversky V, Kataeva IA (Eds). Nova Science Publishers, Inc, UK, 11–46 (2006).

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

91 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).

92 Lemaire M, Miras I, Gounon P, Béguin P. Identification of a region responsible for binding to the cell wall within the S-layer protein of *Clostridium thermocellum*. *Microbiology* 144, 211–217 (1998).

93 Zhao G, Li H, Wamalwa B, Sakka M, Kimura T, Sakka K. Different binding specificities of S-layer homology modules from *Clostridium thermocellum* AncA, Slp1, and Slp2. *Biosci. Biotechnol. Biochem.* 70, 1636–1641 (2006).

94 Sleytr UB, Sára M, Pum D, Schuster B. Characterization and use of crystalline bacterial cell surface layers. *Prog. Surf. Sci.* 68, 231–278 (2001).

95 Zhao G, Ali E, Sakka M, Kimura T, Sakka K. Binding of S-layer homology modules from *Clostridium thermocellum* SdbA to peptidoglycans. *Appl. Microbiol. Biotechnol.* 70, 464–469 (2006).

96 Jindou S, Levy-Assaraf M, Rincon MT *et al*. Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium *Ruminococcus flavefaciens*. *FEMS Microbiol. Lett.* 285, 188–194 (2008).

97 Devillard E, Goodheart DE, Karnati SK *et al*. *Ruminococcus albus* 8 mutants defective in cellulose degradation are deficient in two processive endocellulases, Cel48A and Cel9B, both of which possess a novel modular architecture. *J. Bacteriol.* 186, 136–145 (2004).

98 Xu Q, Morrison M, Bayer EA, Atamna N, Lamed R. A novel family of carbohydrate-binding modules identified with *Ruminococcus albus* proteins. *FEBS Lett.* 566, 11–16 (2004).

99 Ezer A, Matalon E, Jindou S *et al*. Cell-surface enzyme attachment is mediated by a family-37 carbohydrate-binding module, unique to *Ruminococcus albus*. *J. Bacteriol.* 190, 8220–8222 (2008).

▪ **Principle for cell-surface association of plant cell wall-degrading enzymes to the bacterial surface.**

100 Montanier C, van Bueren AL, Dumon C *et al*. Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proc. Natl Acad. Sci. USA* 106, 3065–3070 (2009).

101 Halstead JR, Vercoe PE, Gilbert HJ, Davidson K, Hazlewood GP. A family 26 mannanase produced by *Clostridium thermocellum* as a component of the cellulosome contains a domain which is conserved in mannanases from anaerobic fungi. *Microbiology* 45, 3101–3108 (1999).

102 Blum DL, Li XL, Chen H, Ljungdahl LG. Characterization of an acetyl xylan esterase from the anaerobic fungus *Orpinomyces* sp. strain PC-2. *Appl. Environ. Microbiol.* 65, 3990–3995 (1999).

103 Liu JH, Selinger BL, Tsai CF, Cheng KJ. Characterization of a *Neocallimastix patriciarum* xylanase gene and its product. *Can. J. Microbiol.* 45, 970–974 (1999).

104 Steenbakkers PJM, Arts J, dijkerman R *et al*. The cellulosome-like complex of the anaerobic fungus *Piromyces* E2. In: *Genetics, Biochemistry and Ecology of Cellulose Degradation*. Ohmiya K, Hayashi K, Sakka K, Kobayashi Y, Karita S, Kimura T (Eds). Uni Publishers Co., IN, USA, 507–510 (1999).

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

106 Nagy T, Tunnicliffe RB, Higgins LD, Walters C, Gilbert HJ, Williamson MP. Characterization of a double dockerin from the cellulosome of the anaerobic fungus *Piromyces equi*. *J. Mol. Biol.* 373, 612–622 (2007).

107 Shulami S, Zaide G, Zolotnitsky G *et al*. A two-component system regulates the expression of an ABC transporter for xylo-oligosaccharides in *Geobacillus stearothermophilus*. *Appl. Environ. Microbiol.* 73, 874–884 (2007).

108 Nataf Y, Yaron S, Stahl F *et al*. Cellodextrin and laminarbiose ABC transporters in *Clostridium thermocellum*. *J. Bacteriol.* 191, 203–209 (2009).

■■ **Meticulous characterization of five ABC sugar transporter systems in this principal cellulosome-producing bacterium. The specificities of the sugar-binding lipoproteins associated with the transporter were shown to reflect the substrate preference of the bacterium.**

109 Wilson DB. Three microbial strategies for plant cell wall degradation. *Ann. NY Acad. Sci.* 1125, 289–297 (2008).

110 Himmel ME, Ding S-Y, Johnson DK *et al*. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* 315, 804–807; Erratum: 316, 982 (2007).

■ **Primary review article highlighting recalcitrance of lignocellulosic biomass as the major deterrent to efficient conversion to biofuels.**

111 Lynd LR, van Zyl WH, McBride JE, Laser M. Consolidated bioprocessing of cellulosic biomass: an update. *Curr. Opin. Biotechnol.* 16, 577–583 (2005).

112 van Zyl WH, Lynd LR, den Haan R, McBride JE. Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. *Adv. Biochem. Eng. Biotechnol.* 108, 205–235 (2007).

113 Xu Q, Singh A, Himmel ME. Perspectives and new directions for the production of bioethanol using consolidated bioprocessing of lignocellulose. *Curr. Opin. Biotechnol.* 20, 364–371 (2009).

114 Lynd LR, Laser MS, Bransby D *et al*. How biotech can transform biofuels. *Nature Biotechnol.* 26, 169–172 (2008).

115 Bayer EA, Shoham Y, Lamed R. Cellulosome-enhanced conversion of biomass: On the road to bioethanol. In: *Bioengergy*. Wall J, Harwood C, Demain AL (Eds). ASM Press, DC, USA, 75–96 (2008).

116 Bayer EA, Henrissat B, Lamed R. The cellulosome: A natural bacterial strategy to combat biomass recalcitrance. In: *Biomass Recalcitrance*. Himmel ME (Ed.). Blackwell, UK, 407–426. (2008).

117 Von Ossowski I, Stahlberg J, Koivula A *et al*. Engineering the exo-loop of cellobiohydrolase Cel7A. A comparison with Cel7D. *J. Mol. Biol.* 333, 817–829 (2003).

118 Eijsink VG, Vaaje-Kolstad G, Vårum KM, Horn SJ. Towards new enzymes for biofuels: lessons from chitinase research. *Trends Biotechnol.* 26, 228–235 (2008).

119 Warnecke F, Luginbuhl P, Ivanova N *et al*. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450, 560–565 (2007).

120 Brulc JM, Antonopoulos DA, Berg Miller ME *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).

121 Zhong L, Matthews JF, Hansen PI *et al*. Computational simulations of the *Trichoderma reesei* cellobiohydrolase I acting on microcrystalline cellulose Iβ: the enzyme-substrate complex. *Carbohydr. Res.* 344, 984–992 (2009).

122 Bu L, Beckham GT, Crowley MF *et al*. The energy landscape for the interaction of the family 1 carbohydrate-binding module and the cellulose surface is altered by hydrolyzed glycosidic bonds. *J. Phys. Chem. B* 113, 10994–11002 (2009).

123 Buck M, Bouguet-Bonnet S, Pastor R, MacKerell A Jr. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. *Biophys. J.* 90, L36-L38 (2006).

124 Guvench O, Greene SN, Kamath G *et al*. Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* 29, 2543–2564 (2008).

125 Côté N, Fleury A, Dumont-Blanchette E, Fukamizo T, Mitsutomi M, Brzezinski R. Two exo-β-D-glucosaminidases/exochitosanases from actinomycetes define a new subfamily within family 2 of glycoside hydrolases. *Biochem. J.* 394, 675–686 (2006).

126 Morris DD, Gibbs MD, Ford M, Thomas J, Bergquist PL. Family 10 and 11 xylanase genes from *Caldicellulosiruptor sp*. strain Rt69B.1. *Extremophiles* 3, 103–111 (1999).

▪ Websites

201 The Carbohydrate Active enZyme (CAZy) database www.cazy.org

202 The *Clostridium thermocellum* ATCC 27405 genome project http://genome.jgi-psf.org/draft_microbes/cloth/cloth.home.html

203 The *Clostridium acetobutylicum* ATCC824 genome project http://cmr.jcvi.org/cgi-bin/CMR/GenomePage.cgi?org=ntca01

204 The *Clostridium acetobutylicum* H10 genome project http://genome.jgi-psf.org/draft_microbes/cloce/cloce.home.html

205 *Clostridium papyrosolvens* DSM 2782 genome project http://genome.ornl.gov/microbial/cpap/