Extracellular Glycosyl Hydrolases from Clostridia

WOLFGANG H. SCHWARZ,* VLADIMIR V. ZVERLOV,†
AND HUBERT BAHL‡,§

*Technical University of Munich
Institute of Microbiology, D-85350 Freising, Germany
†Russian Academy of Science, Institute of Molecular Genetics
123182 Moscow, Russia
‡University of Rostock, Institute of Biological Sciences
Department of Microbiology, D-18051 Rostock, Germany
§Author for correspondence. E-mail: hubert.bahl@biologie.uni-rostock.de

I. Introduction 215
II. Modular Structure of the Enzymes 217
III. Function of Noncatalytic Modules 218
   A. Substrate Binding 219
   B. SLH Module 223
   C. Fibronectin Type III Module 224
IV. Characterization of Enzyme Systems 225
   A. Starch Degradation 228
   B. Cellulose Degradation 229
   C. Xylan and Hemicellulose Degradation 247
V. Concluding Remarks 251

References 252

I. Introduction

Anaerobic bacteria are key players in the fate of rotting biomass. They play a major role in the digestion of biomass by herbivores and insects (such as termites), possibly even as endosymbionts of flagellates common in the intestinal tract of plant-feeding animals, such as the rumen of cattle. The hosts help by mechanical degradation (chewing) and by providing a favorable environment. A part of the natural rotting process of biomass in soil and compost heaps is also performed by the anaerobic bacteria when the easily degradable constituents (e.g., soluble sugars and proteins) of the biomass are already used up. Among the anaerobic bacteria are specialists for the degradation of the insoluble components of biomass that are most difficult to degrade: crystalline starch, hemicellulose, and cellulose.

In nature, polysaccharide-degrading bacteria thrive in symbiotic relationships with secondary microorganisms (Ljungdahl and Eriksson,
The enzymes secreted by the primary cellulose degraders break the substrate down into celldextrins, cellobiose, and glucose, only a part of which is assimilated by the polymer-degrading strains themselves. The rest is utilized by the secondary microbial flora, as are the fermentation products of the anaerobic cellulose degraders: hydrogen, carbon dioxide, alcohols, and short-chain fatty acids. Thus polysaccharide degradation is just the first step in a food chain within a complex ecosystem. Approximately $1.8 \times 10^{12}$ metric tons of biomass (dry weight) exist on the continents, which are continuously recycled by enzymatic processes. Polysaccharides from plant material form a major part of the biomass: They are the most important factors in the carbon cycle in nature that regulates the CO$_2$ content of the atmosphere. An estimated 40 GT per year alone of cellulose are produced by land plants—about the same amount is degraded. The natural rotting process is catalyzed by hydrolytic enzymes produced from ubiquitous microorganisms. The energy contained in the resulting sugars drives the build-up of micro- and macrobiotic biomass. But the energy gradient from polysaccharide to CO$_2$ can also be exploited for industrial purposes without increasing the CO$_2$ content in the atmosphere: biomass, through burning or enzymatic hydrolysis, is a CO$_2$-neutral source of environmentally friendly energy for the future.

Anaerobic bacteria, among them primarily the clostridia, are an excellent source for hydrolytic enzymes able to hydrolyze polysaccharides in biomass to fermentable sugars. An example of special interest is the utilization of the hydrolytic extracellular enzymes of the solventogenic bacterium *Clostridium acetobutylicum* for the fermentation of starch to the organic solvents butanol and acetone (Dürrre, 1998; Gapes, 2000). Although, because of economic reasons, the industrial process at present is not utilized in the Western world, it is still a very attractive alternative to the mineral oil–based production of energy and bulk chemicals, since it runs with renewable substrates, enabling sustainable energy production. Consequently, research on the bacterial solvent production process is going on; for example, a number of new strains degrading a wide range of polysaccharides have been isolated (Dürrre, 1998; Montoya et al., 2001). Meanwhile, determination of the genomic sequence of *C. acetobutylicum* made a thorough analysis of its genes possible, and a complete cluster of genes for the expression of a cellulosome was detected (Nölling et al., 2001). Unfortunately, only few of these genes are expressed, and hydrolysis of cellulosic substrates could not be achieved (Sabathe et al., 2002). Nevertheless, this opens the possibility that related strains may exist that express the whole operon and would then be able to produce solvents directly from cellulose. In
addition to genetic engineering of producer organisms and new fermentation and product separation technologies, this will help to make the bacterial solvent production economically feasible in the near future.

So far none of the strains used for industrial production of acetone and butanol have been able to degrade cellulose as a cheap and available substrate to fermentable sugars. However, only a few of the industrial strains have survived the shutdown of the production facilities. The rest of the valuable strains are permanently lost and cannot be tested. The search for new solventogenic strains capable of efficient lignocellulose hydrolysis is therefore going on, and research on the clostridial extracellular enzymes is an increasingly urgent necessity. This chapter will focus on the most prevalent polysaccharides present in biomass: starch, cellulose, and hemicellulose—the latter two of which are especially difficult substrates to degrade. The unique strategies of the clostridia to cope with these substrate problems are discussed.

II. Modular Structure of the Enzymes

In contrast to the enzymes isolated from eukaryotic organisms (mostly fungi) and aerobic bacteria, many extracellular enzymes of the anaerobic bacteria have a modular structure—that is, they consist not only of a catalytic module but of a complex arrangement of different modules: one or even more than one catalytic module(s) and in addition, noncatalytic modules. In Fig. 1, a schematic modular structure of a hypothetical clostridial glycosyl hydrolase is depicted. The catalytic module can be accompanied by one, several, or all of the following modules: carbohydrate binding (CBM), immunoglobulin (Ig)-like, dockerin (Doc), fibronectin type III (Fn3), and S-layer homology (SLH). These modules constitute independent folding units that often are covalently connected by flexible linkers such as the so-called PTS boxes (irregular stretches of hydroxy amino acids). As a consequence of the presence of several modules, these enzymes are often quite large, consisting of more than 1,000 amino acids with a molecular mass above

![Fig. 1. Schematic representation of the modular structure of a hypothetical clostridial extracellular enzyme. CBM: carbohydrate-binding module; Doc: dockerin module; Fn3: fibronectin type III module; GH: catalytic module of glycosyl hydrolase family; Ig: immunoglobulin-like module; SLH: surface-layer homology module. Numbers indicate the relative position of the modules.](image)
100 kDa. The order of modules in a given enzyme does not follow strict rules. Noncatalytic modules may appear on the N- or the C-terminal end of the catalytic units. SLH or Doc modules are in most cases located near the C-terminus of the enzymes. In general, the noncatalytic modules may support or even modulate the catalytic activity. Some are stuffer proteins between a catalytic unit and a functionally important noncatalytic module; some are closely connected with a catalytic module and stabilize it against thermal denaturation. Binding modules are known for the substrate (CBM) or for the host cell (SLH).

The catalytic modules of these enzymes hydrolyze the glycosidic bond between two or more carbohydrates or between a carbohydrate and a noncarbohydrate moiety. Based on amino acid sequence similarities, a classification of glycoside hydrolases into families has been proposed (Henrissat, 1991): The updated list (October 2003) contains 91 families, GH1 to GH91 (Coutinho and Henrissat, 1999a). The catalytic mechanism is a general acid catalysis that requires two critical amino acid residues: a proton donor and nucleophile/base (Davies and Henrissat, 1995). The hydrolysis results in either retention or inversion of the configuration at the anomeric C-atom.

The roles of the Doc, Fn3, and SLH modules are summarized in the following section, whereas the function of the Ig module has not yet been successfully addressed.

The most complex enzymes are those of the extremely thermophilic cellulose degraders Cellulosiruptor cellulolyticus and the closely related Anaerocellum thermophilum, which contain a so-called multifunctional enzyme system; these are not included in this review (Bayer et al., 2000). They also belong to the order Clostridiales. Many of their cellulases and hemicellulases are composed of more than one catalytic module, connected with binding modules and stuffing peptides. Functionally related and mutually synergistic catalytic components are combined in one polypeptide chain to enhance the effectiveness of enzymatic action. This seems to be an independent evolutionary way towards an enzyme complex that combines all necessary functions in a close spatial arrangement but with more flexibility in structure and composition. The most advanced of these complexes is the cellulosome, which is described now.

III. Function of Noncatalytic Modules

The functionally most important and best-characterized noncatalytic module in the extracellular enzymes of the clostridia is the CBM. In recent years the SLH module was in the focus of functional analysis,
whereas only limited knowledge exists on the function of the Fn3 module. Therefore these three modules will be described in this part of the chapter. The function of the Doc module can be found within the description of the cellulosome.

A. Substrate Binding

The interaction of enzymes with polymeric substrates is severely slowed by the limited diffusion of the enzyme as well as the substrate. This difficulty is greatly overcome by the introduction of binding modules. These are protein modules of about 35 to less than 180 amino acid residues that target the enzyme in a noncatalytic way to suitable areas of the large substrate, a single polysaccharide molecule thread as in soluble or in amorphic parts of insoluble substrates, or a bundle of insoluble substrate molecules as in crystalline cellulose (Linder and Teeri, 1997). This increases the enzyme concentration on the substrate surface and improves substrate interaction (Bolam et al., 1998). The carbohydrate binding modules (CBM) are categorized into families according to sequence homology and the consequent three-dimensional fold (Coutinho and Henrissat, 1999b). A list of the presently known CBMs with links to nucleotide and amino acid sequences and a short compilation of general information on each family is given at the CAZY server (Coutinho and Henrissat, 1999a).

Some CBMs have a flat strip of aromatic amino acid residues for binding to the surface of an array of parallel substrate molecules as in crystals; others bind single substrate molecules in a pocket-like structure (reviewed in Bayer et al., 1998). The anchoring is mediated by polar residues such as asparagine or glutamine (Tormo et al., 1996). Some families can be separated into slightly different subfamilies that have the same global fold but differ in their binding abilities. An example is family CBM3, where subfamily CBM3a modules bind tightly to crystalline cellulose, whereas CBM3b modules seem to be more variable, and CBM3c modules modulate the enzymatic activity by feeding a single substrate molecule with a predefined directionality through the active site pocket of the catalytic module, leading to processive cleavage (Sakon et al., 1997). An enzyme with endoglucanase activity is consequently transformed into an exo-glucanase (or a processive endo-glucanase) by the activity of a binding module. This emphasizes the important role of the noncatalytic modules for the enzyme activity, especially for the hydrolysis at different sites on crystalline or amorphic cellulose (Carrard et al., 2000).
It has been proposed that some CBMs may be degenerated and function as thermostabilizing modules, such as the CBM3c in *C. stercorarium* cellulase Cel9Z or *C. thermocellum* cellulase Ce19I (see below). Although it is not clear if such CBMs have lost their binding capacity, they are functionally attached to the catalytic module. This seems to stabilize the structure of the catalytic core and in some cases increases the thermostability up to 30 °C (Riedel et al., 1998a). The loss of activity at high temperature on deletion of the CBM may be so drastic that the function of the module was in some cases interpreted as essential for activity through binding. Similar results have been reported for the CBM22 modules that, for example, exert a thermostabilizing effect on xylanase XynA from *Thermotoga maritima* but at the same time bind to xylan and -1,3-1,4-glucan (Meissner et al., 2000). The authors argue that thermostabilization is a side effect of the close association of the enzyme with its substrate binding module.

Despite a very tight binding to the substrate through a CBM, the enzymes seem to diffuse laterally along the substrate molecule (Jervis et al., 1997). This was shown with fungal enzymes in impressive pictures of the crystal surface through atomic force microscopy (Lee et al., 2000). Some dynamic experiments have been performed with the family CBM3a, which is important for the effectiveness of the cellulosome of *C. thermocellum*. In *in vitro* experiments, this CBM binds the scaffolding irreversibly to crystalline cellulose and allows the cellulytic cellulosome components to be effective without a CBM and probably with a greater freedom of movement for activity around the binding site.

The binding ability can be investigated primarily by two methods (Tomme et al., 1996). In equilibrium assays, the binding protein is mixed with an insoluble substrate in a suitable buffer; after equilibration, the decrease of protein in the cleared supernatant is estimated. Alternatively, a soluble binding substrate is mixed into the gel matrix of a native polyacrylamide gel; the binding protein is subjected to electrophoresis with these gels, and retardation in comparison with other (nonretarded control) proteins is determined. However, the *in vivo* function of the CBM can be detected only by constructing deletion mutants of enzymes and thorough characterization of the mode of the enzymatic activity, especially the processivity in the case of the cellulases.

The binding specificity is not necessarily identical with the substrate specificity of the associated catalytic module. Especially with the xylanases it is common to find cellulose binding modules. The xylanases present in the cellulosome do not have their own CBMs but
rely on the cellulose binding capacity of the scaffoldin-linked CBM. This may have to do with the function of these enzymes that serve as supporters in biomass degradation, where xylan is associated with cellulose (see below). In contrast, a CBM6 module was ascribed a role in hemicellulose degradation by binding to insoluble xylan and several soluble polysaccharides (Sun et al., 1998).

CBMs are not randomly connected with catalytic modules. A limited number of them play a role in the extracellular enzymes in clostridia, and patterns of module structures emerge that seem to be evolutionarily successful. Examples are the “themes” A to D of the GH9 cellulases depicted by Bayer et al. (2000), where the GH9 cellulases form a repeatedly observed pattern with noncatalytic modules:

Theme A: GH9
Theme B: GH9–CBM3c
Theme C: Ig–GH9
Theme D: CBM4–Ig–GH9

Others are given in the next list, where the preferential binding activity of the CBM families is also indicated.

In connection with cellulases:

- CBM3a (with scaffoldin)—crystalline cellulose
- CBM3b (often as GH9-CBM3c-CBM3b)—crystalline cellulose
- CBM3c (often as GH9-CBM3c)—as thermostabilizing module and possibly for substrate feeding (binding amorphous cellulose)
- CBM4 (often as CBM4-Ig-GH9)—amorphous cellulose or soluble oligosaccharides
- CBM6—amorphous cellulose, β-1,3-glucan, or xylan
- CBM9 (e.g., GH5-CBM9)—cellulose

In connection with xylanases:

- CBM22—thermostabilizing module, or xylan or β-1,3-1,4-glucan
- CBM4 (also in β-glucanase Ct-Lic16A)—amorphous cellulose or soluble oligosaccharides

With esterases/glycosidases:

- CBM6 (with esterase CE1, PL6, PL1, PL11, GH30, GH43, GH39)—amorphous cellulose, β-1,3-glucan or xylan

In connection with amylases:

- CBM 20—binding to starch
- CBM 26—binding to cyclodextrins
In connection with other enzymes:

CBM2 (PL, GH74)—crystalline cellulose, chitin or xylan
CBM13 (GH43)—as threefold repeat (xylan binding?)

The best-characterized carbohydrate binding module in starch degrading enzymes belongs to the CBM20 family. This module is also known as starch-binding domain: It consists of approximately 100 residues and has granular starch-binding activity. CBM26 has been found in enzymes with different enzymatic activities (e.g., α-amylases, β-amylases, glucoamylases, and especially cyclodextrin glucanotransferases). The presence in the latter enzyme is consistent with the fact that CBM26 strongly interacts with cyclodextrins. In clostridial enzymes this module has been found so far only in a few cases (Table I). Other CBMs with affinity to starch have been identified in not further characterized gene products of *C. acetobutylicum* (Table I).

<table>
<thead>
<tr>
<th>Module</th>
<th>Species</th>
<th>Enzyme</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBM20</td>
<td><em>Thermoanaerobacterium</em></td>
<td>β-amylase</td>
<td>Kitamoto <em>et al.</em>, 1988</td>
</tr>
<tr>
<td></td>
<td><em>thermosulfurigenes</em></td>
<td>amylase-pullulanase</td>
<td>Matuschek <em>et al.</em>, 1994</td>
</tr>
<tr>
<td></td>
<td><em>EM1</em></td>
<td>cyclodextrin glucanotransferase</td>
<td>Wind <em>et al.</em>, 1995</td>
</tr>
<tr>
<td></td>
<td><em>Thermoanaerobacter ethanolicus</em></td>
<td>amylase-pullulanase</td>
<td>Mathupala <em>et al.</em>, 1990</td>
</tr>
<tr>
<td></td>
<td><em>saccharolyticum</em></td>
<td>amylase-pullulanase</td>
<td>Ramesh <em>et al.</em>, 1994</td>
</tr>
<tr>
<td></td>
<td><em>thermohydrosulfuricus</em></td>
<td>amylase-pullulanase</td>
<td>Melasniemi <em>et al.</em>, 1990</td>
</tr>
<tr>
<td>CBM21</td>
<td><em>Clostridium acetobutylicum</em></td>
<td>CAP 0129</td>
<td>Nölling <em>et al.</em>, 2001</td>
</tr>
<tr>
<td>CBM25</td>
<td><em>Clostridium acetobutylicum</em></td>
<td>α-amylase</td>
<td>Nölling <em>et al.</em>, 2001</td>
</tr>
<tr>
<td>CBM26</td>
<td><em>Clostridium acetobutylicum</em></td>
<td>CAC 2252</td>
<td>Nölling <em>et al.</em>, 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAC 2891</td>
<td></td>
</tr>
</tbody>
</table>

Several thermophilic *Clostridium* species have been reclassified as members of the genera *Thermoanaerobacterium* and *Thermoanaerobacter* (Lee *et al.*, 1993). Data taken from Coutinho and Henrissat (1999b).
The function of SLH modules (for surface-layer homology) in the extracellular enzymes of clostridia is not immediately obvious. SLH modules are found in proteins from phylogenetically unrelated bacteria (e.g., Gram-positive and Gram-negative bacteria) and are present in three types of proteins: surface-layer (S-layer) proteins, extracellular enzymes/proteins, and outer membrane proteins (Engelhardt and Peters, 1998). In most cases the SLH module is present in three copies of about 50–60 residues each. A single module is predicted to have the following secondary structure pattern: $\alpha$-helix ($H_I$)—$\beta$-sheet ($S$)—loop ($L_I$)—$\alpha$-helix ($H_{II}$)—loop ($L_{II}$) (Fig. 2). The overall similarity of SLH modules in proteins from different organisms is low, but they contain at least two highly conserved motifs, a FxDV motif at the N-terminus and an TRAE motif at the beginning of the second $\alpha$-helix. Our data indicate that at least the TRAE motif contributes to the function of SLH modules (unpublished results).

In S-layer and outer membrane proteins these modules are generally located at the N-terminus and in enzymes at the C-terminus. Their role as cell-wall targeting modules was initially proposed mostly on the basis of sequence comparison (Fujino et al., 1993; Lupas et al., 1994; Matuschek et al., 1994). Now, because of several in vitro and in vivo studies, there is strong evidence that the SLH modules indeed serve as an anchor to the cell wall for the different protein types (Brechtl et al., 1999; Lemaire et al., 1995; Mesnage et al., 1999; Olabarria et al., 1996; Ries et al., 1997). Although it was initially thought that SLH modules bind to peptidoglycan, it is now clear that the adhesion component in the cell wall is not the peptidoglycan itself but a polymer covalently linked to it (Brechtl and Bahl, 1999; Ilk et al., 1999; Mesnage et al., 1999; Ries et al., 1997; Sára et al., 1996). Complete structural analysis has indicated that these cell-wall associated polymers are teichuronic acids (Ilk et al., 1999). Furthermore, it has been found that they are pyruvylated and that a strong correlation between the existence of SLH modules and genes involved in the addition of pyruvate to the wall-associated polymer exists in different bacteria (Mesnage et al., 2000).

![Figure 2](image.png)

**FIG. 2.** Consensus sequence and highly conserved motifs of SLH modules. H: $\alpha$-helix; S: $\beta$-sheet; L: loop.
Thus the SLH-mediated anchoring mechanism is one of several, but highly conserved strategy bacteria have developed to display proteins on their surface.

In clostridia, SLH modules have been found in S-layer proteins and in several hydrolases (e.g., cellulases, xylanases, amylase-pullulanases) (Fuchs et al., 2003; Matuschek et al., 1996). Figure 3 illustrates how SLH modules mediate the attachment of S-layer proteins, single enzymes, or multienzyme complexes to the cell wall.

C. FIBRONECTIN TYPE III MODULE

The Fn3 module is one of three types of internal repeats found in the plasma protein fibronectin. Many animal proteins contain the Fn3 module, including extracellular, intracellular, and membrane-spanning proteins, and adhesion molecules. Surprisingly, Fn3-like modules are also found in bacterial glycosyl hydrolases (Little et al., 1994), including cellulases, pullulanases, and polygalacturonases from clostridia (Matuschek et al., 1996; Zverlov et al., 1998b). In
Thermoanaerobacterium thermosulfurigenes EM1 Fn3 modules are present in the amylase-pullulanase (AmyB) and the polygalacturonate hydrolase (PglA) (Matuschek et al., 1996). Interestingly, the Fn3 modules of PglA and the exopolygalacturonate hydrolase of Erwinia chrysanthemi (He and Collmer, 1990) share a higher degree of similarity (38% identical residues) than the module of PglA and the two modules of AmyB (27% and 29% identical residues, respectively). On the other hand, the Fn3 modules of AmyB are 64% identical to the corresponding modules in the amylase-pullulanase of Thermoanaerobacter thermosulfuricus E101-69 (Melasniemi et al., 1990). Therefore the Fn3 modules appear to be clustered by protein type and not by organism.

Very little information is available on the function of Fn3 modules in extracellular enzymes of bacteria. It has been postulated that they serve as spacers or linkers allowing optimal interaction between the catalytic and substrate-binding modules (Little et al., 1994). In agreement with this suggestion, Watanabe et al. (1994) reported that deletion of the Fn3 module(s) located between the catalytic and substrate-binding modules of a chitinase from Bacillus circulans did not affect binding to chitin but decreased hydrolytic activity of the enzyme to colloidal chitin.

Recently the first evidence of a function for Fn3 in a clostridial enzyme during hydrolysis of a polysaccharide was presented. It was shown that the two Fn3 modules of the multi-modular cellobiohydrolase CbhA of Clostridium thermocellum are able to change the surface of cellulose that had been loosened up and crenellated. That promoted hydrolysis by the catalytic domain (Kataeva et al., 2002).

IV. Characterization of Enzyme Systems

Polysaccharides are difficult substrates for enzymes. They are usually larger than the enzyme itself, and quite often they are not soluble (i.e., they are not hydrated or occur in tight aggregates or even in crystalline form). Moreover, many natural polysaccharides such as hemicelluloses are extremely heterogeneous and contain many different sugar moieties with different linkage types, or they are derivatized. Others may contain only one type of sugar moiety, which, however, as in starch, are linked in different ways (1->4 or 1->6). Others are chemically homogeneous (such as cellulose: β-1,4-glicosidic linkages only) but are at least partially crystalline and topologically diverse.

Long substrate molecules can be shortened by a statistical hydrolytic cut in one of the many linkages between the building blocks of the substrates. This is the so-called endo-mode of action. The sites for such
enzymatic attacks may be extremely limited because of occlusion by other molecules, by the viscosity of the polymer solution, or by the tight assembly of many molecules (e.g., by crystal formation). Only freely accessible, hydrated parts of a long molecule can be recognized by enzymes, and it could be postulated that even within the same polymer strand different enzyme types are needed to hydrolyze different topologies.

The substrate size poses the problem of diffusion: It is not the substrate that tumbles around until it finds an enzyme pocket, but the enzyme must find its substrate. The diffusion of such large molecules is slow, and hence it is an advantage to stick to the substrate and degrade it successively once a large substrate molecule is found (see the previous section on binding modules). The sequential, processive action is executed by the exo-mode enzymes, which recognize either a reducing or a nonreducing end of the substrate molecule and feed it through the activity pocket, chopping off a monomer (e.g., $\alpha$-glucosidase), a dimer (e.g., $\beta$-amylase, cellobiohydrolase), or a multimer (several exo-glucanases). A synergism between endo-glucanases that produce the open ends, and exo-glycanases and processive enzymes that widen the gap, has been observed (i.e., the sum of the single activities is smaller, as if both types of enzymes act in combination simultaneously). Another synergism exists between the exo-glycanases active from the reducing and the nonreducing end: they can be thought to act from one open cut in a long molecule into both directions, opening a hole (e.g., in the surface of a cellulose crystal).

Polysaccharides are degraded by extracellular enzymes (sometimes also called “depolymerases”). The resulting monosaccharides or oligosaccharides are either taken up by the cell or degraded extracellularly by secreted glycosidases. Examples are discussed below.

Commonly, the potential of a bacterial strain to produce extracellular enzymes is evaluated by assaying the cell free culture supernatant for enzymatic activities. This was done, for example, in an effort to compare the thermophilic polysaccharide hydrolyzing bacteria $\textit{C. stercorarium}$ and $\textit{C. thermocellum}$ for enzymes in the culture supernatant (Table II). The cellulolytic activities of these two species were comparable, but $\textit{C. stercorarium}$ culture supernatant had a higher soluble activity for mixed-linkage glucan and xylan, and especially for glucoside, xyloside, and arabinoside. This can be taken as an indication for a higher activity on hemicellulose in $\textit{C. stercorarium}$. The higher cellobiosidase activity of $\textit{C. thermocellum}$ corroborates its specialization for cellulose as substrate, since the major components of cellulases are cellobiohydrolases that are active on the cellobioside.
However, the results of such assays have to be interpreted with great care:

1. A great portion of the exo-enzymes in *C. thermocellum* are located on the cell surface, such as the cellulosome complex and some single enzymes such as the β-1,3-glucanase Lic16A (*Fuchs et al.*, 2003); the same is true for one xylanase of *C. stercorarium* (see below); their activity escapes the assay.

2. Many enzymes are not specific (e.g., β-glucosidases are also active on cellobiosides and xylosides).

3. Xylanases of GH10 have high activity on mixed-linkage glucans.

Thus, information from the activity of culture supernatants on model substrates alone is not sufficient to estimate the hydrolytic potential of a given bacterium.

Nevertheless, *C. stercorarium* was identified as a thermophilic bacterium with a more general activity on polysaccharides present in biomass, whereas *C. thermocellum* is known as a specialist for the degradation of cellulose, despite the presence of a number of other enzymatic activities for β-1,3-glucan, xylan, mannan, pectin, chitin, and probably other polysaccharides.

Valuable information on a complete catabolic pathway of a bacterium comes from fermentation experiments; if a given polymeric substrate can be depolymerized into oligosaccharides and further to monomeric sugars, transported into the cell and metabolized, the complete set of at

---

**TABLE II**

*COMPARISON BETWEEN TWO THERMOHYDROLYTIC, SACCHAROLYTIC CLOSTRIDIA*

<table>
<thead>
<tr>
<th>Substrate</th>
<th><em>C. thermocellum</em> (mU/ml)<em>a</em></th>
<th><em>C. stercorarium</em> (mU/ml)<em>a</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Microcrist. cellulose</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Phosphoric acid swollen cellulose</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Carboxymethyl cellulose (CMC)</td>
<td>140</td>
<td>120</td>
</tr>
<tr>
<td>1,3-1,4-β-glucan (lichenan)</td>
<td>6,500</td>
<td>12,000</td>
</tr>
<tr>
<td>Arabino-xylan</td>
<td>3,000</td>
<td>20,000</td>
</tr>
<tr>
<td>pNP-β-glucopyranoside</td>
<td>1,3</td>
<td>7</td>
</tr>
<tr>
<td>pNP-β-celllobioside</td>
<td>12</td>
<td>1,7</td>
</tr>
<tr>
<td>pNP-β-xylopyranoside</td>
<td>0,3</td>
<td>2</td>
</tr>
<tr>
<td>pNP-α-arabinofuranoside</td>
<td>1,3</td>
<td>21</td>
</tr>
</tbody>
</table>

*aActivity in cell free culture fluid (grown on cellobiose).*
least the hydrolytic enzymes must be present. The presence of the genes alone as shown by gene cloning or genome sequencing is not sufficient; expression and secretion of the proteins must also occur. Furthermore, the presence of a hydrolytic gene (e.g., for an endo-xylanase) is not sufficient to prove xylan degrading activity, even if the enzyme is found extracellularly. To hydrolyze xylan, cellulose, raw starch, and a number of other natural polysaccharides, a network of enzymes is needed, a so-called enzyme system. Examples of such enzyme systems are discussed below.

It is obvious from the synergism explained above and from the immobility of most polymeric substrates that a high local concentration of all enzyme components necessary for the substrate degradation is needed for efficient hydrolysis. Two strategies are possible:

1. To increase the concentration of all enzyme components in the medium, or
2. To combine the necessary components in a complex and to add a binding module for the substrate to hold the complex on the substrate once it has found it.

Both possibilities are realized by clostridia as is explained with the cellulase enzyme systems below. An intermediate possibility is found with the species *Caldicellulosiruptor* and *Anaerocellum*, where large proteins are secreted containing several catalytic and noncatalytic modules in one polypeptide chain. The cellulosome of the clostridia, a protein complex on a scaffolding protein, seems to be the more elegant and flexible solution.

### A. STARCH DEGRADATION

Starch is an abundant polymer in plant biomass and consists of two components: amylose, a linear polymer of α-1,4-linked glucose residues, and amylopectin, a branched polymer in which amylose chains are connected via α-1,6-linkages. The relative distribution of amylose and amylopectin in a starch molecule and the degree of branching depends on the source of starch. Complete degradation of starch is achieved by endo-acting (α-amylase) and exo-acting (β-amylase and glucoamylase) enzymes. Enzymes that hydrolyze the α-1,6-linkages are named pullulanase or debranching enzymes. Maltose and short oligosaccharides produced during primary hydrolysis are converted to glucose by α-glucosidase. Starch is a good substrate for most of the saccharolytic clostridia, and all types of starch hydrolases have been found among them (Mitchell, 2001). Nevertheless, the knowledge
about starch-degrading enzymes from clostridia is limited. In contrast to the many *Clostridium* species able to degrade starch, there are relatively few entries in the database on hydrolases of GH13 (α-amylase, pullulanase, cyclodextrin glucanotransferase), GH14 (β-amylase), or GH15 (glucoamylase) (*Coutinho and Henrissat, 1999b*). Furthermore, although starch was one of the preferred substrates for the industrial acetone-butanol fermentation by *C. acetobutylicum*, very little is known about its starch-degrading enzyme system (*Gerischer and Dürrre, 1988; Paquet et al., 1991*). Sequencing of the genome has identified a few genes related to starch degradation in this important organism (*Table I; Nölling et al., 2001*).

Thermophilic species and their thermostable enzymes have attracted particular interest (*Antranikian, 1990*). In some of these organisms, which later were reclassified as members of the genera *Thermoanaerobacter* and *Thermoanaerobacterium*, a novel type of pullulanase, which hydrolyzes α-1,4- and α-1,6-linkages, was identified (*Spreinat and Antranikian, 1990*). Other enzymes that hydrolyze both types of linkages are the glucoamylase and the α-glucosidase of *C. thermosaccharolyticum* and of *C. beijerinckii* (*Albasheri and Mitchell, 1995; Ganghofner et al., 1998; Specka et al., 1991*). In addition, synergistic action of pullulanase and α-amylase (cyclodextrin glucanotransferase) has been observed (*Spreinat and Antranikian, 1992*). In Fig. 4, the action of enzymes from clostridia and related bacteria on a starch molecule is illustrated.

**B. CELLULOSE DEGRADATION**

Cellulose is a completely insoluble, partially nonhydrated, and crystalline substrate that poses special difficulties for enzymatic hydrolysis. Although it is a chemically homogeneous, unbranched polymer of β-1,4-linked glucopyranose residues, it is structurally heterogeneous. Only a very small fraction of the substrate molecules on the surface or in amorphic regions of the crystal are susceptible to immediate enzyme attack. The current understanding of enzymatic cellulose hydrolysis is as follows: an endo-glucanase binds with its attached binding module (CBM) to the surface of a substrate bundle, opens the cellulose molecule at one of a few accessible sites and consequently produces a new reducing and a nonreducing end. The endo-glucanase stays bound near this site and may open other available cellulose chains in reachable distance. Processive glucanases (exo-glucanases) find the open ends and walk successively along the cellulose thread either from the nonreducing or the reducing end. They produce
Fig. 4. Enzymes from clostridia and related bacteria involved in the degradation of starch. Glucose units in the starch molecule with a reducing end are drawn in black; those with a nonreducing end are in grey. Data taken from Coutinho and Henrissat (1999a).
cellobiose (cellobiohydrolases) or cellotetraose (processive endo-glucanases), depending on the enzyme type (Reverbel-Leroy et al., 1997), widen the gap, and expose another layer of cellulose chains on the surface of the crystal, which may in turn be attacked by endo-glucanases. The cellulodextrins produced are transported into the bacterial cells, where they are hydrolyzed by β-glucosidases to glucose or, energetically more favorable, cleaved phosphorolytically by phosphorylases to glucose-1-phosphate.

All cellulases cleave a β-1,4-glucosidic bond by a hydrolytic reaction (hence “β-1,4-glucanases”). It is the same type of chemical reaction that takes place, but the mode of attack differs: cellulases may be endo- or exo-glucanases (cellobiohydrolases or processive endo-glucanases); exo-glucanases may be active from the reducing or the nonreducing end of the molecule. Only a combination of enzymes with a different mode of action works synergistically and degrades the crystalline substrate effectively (Barr et al., 1996). Some enzymes hydrolyze the glucosidic bond by an inverting mechanism, others by a retaining mechanism (Davies and Henrissat, 1995). However, this difference in the hydrolytic mechanism does not seem to play a role in action modes; neither is the basic fold of an enzyme important, which is reflected in the glycosyl hydrolase family (GH family) to which a catalytic module is assigned (Coutinho and Henrissat, 1999a,b).

Some of the GH families contain exo- as well as endo-glucanases (e.g., GH5, GH9). The endo- or exo-mode of a given enzyme is determined by the depth and the accessibility of the active site pocket; the processivity seems to be a function of attached substrate-binding modules and their orientation towards the hydrolytic center of the catalytic module (Barr et al., 1996; Bayer et al., 2000). At least in some enzymes the direction of the processivity could be explained by the way the substrate is bound and released and not by the gross structure of the protein backbone (Parsiegla et al., 2000).

An especially high synergism between bacterial cellulases has been described between enzymes of GH48 and GH9 (Riedel et al., 1997), which are present in all cellulase enzyme systems known so far in bacteria. The synergism is higher with higher enzyme concentration, (i.e., the vicinity of proteins of both types is a crucial factor). By placing two enzymes close to each other, the cellulolysis can be optimized. Many clostridia reach this goal by packing enzymes of suitable types together in a cell- or substrate-bound huge enzyme complex, the cellulosome. One species secretes the enzymes separately as a “soluble” enzyme system. Examples for both types are given next.

Interestingly, cellulosomes have been found exclusively in clostridia and some closely related Lachnospiraceae, such as Butyrivibrio and
<table>
<thead>
<tr>
<th>Taxonom. group</th>
<th>Genus</th>
<th>Species</th>
<th>Temp.</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clostridiaceae</td>
<td>Clostridium</td>
<td>acetobutylicum</td>
<td>M</td>
<td>Sewage</td>
<td>Sabathe et al., 2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cellulovorans</td>
<td>M</td>
<td>Wood fermenter</td>
<td>Tamaru et al., 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>celllobioparum</td>
<td>M</td>
<td>Rumen</td>
<td>Lamed et al., 1987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cellulolyticum</td>
<td>M</td>
<td>Compost</td>
<td>Bélaich et al., 1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>josui</td>
<td>M</td>
<td>Compost</td>
<td>Kakiuchi et al., 1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>papyrosolvans</td>
<td>M</td>
<td>Paper mill</td>
<td>Pohlschröder et al., 1995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>thermocellum</td>
<td>T</td>
<td>Sewage, soil</td>
<td>Lamed et al., 1987</td>
</tr>
<tr>
<td></td>
<td>Acetivibrio&lt;sup&gt;a&lt;/sup&gt;</td>
<td>celulolyticus</td>
<td>M</td>
<td>Sewage</td>
<td>Ding et al., 1999</td>
</tr>
<tr>
<td></td>
<td>Bacteroides&lt;sup&gt;a&lt;/sup&gt;</td>
<td>cellulosolvans</td>
<td>M</td>
<td>Sewage</td>
<td>Lamed et al., 1991</td>
</tr>
<tr>
<td>Lachnospiraceae</td>
<td>Butyrivibrio</td>
<td>fibrisolvans</td>
<td>M</td>
<td>Rumen</td>
<td>Berger et al., 1990</td>
</tr>
<tr>
<td></td>
<td>Ruminococcus</td>
<td>albus</td>
<td>M</td>
<td>Rumen</td>
<td>Ohara et al., 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>flavfaciens</td>
<td>M</td>
<td>Rumen</td>
<td>Aurilia et al., 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>succinogenes</td>
<td>M</td>
<td>Rumen</td>
<td>Fields et al., 2000</td>
</tr>
</tbody>
</table>

<sup>a</sup>Unambiguous assignment to the Clostridiaceae due to the 16S rDNA sequence in phylogenetic tree construction by ARB.

Temperature: M, mesophilic; T, thermophilic.
Ruminococcus strains (Table III). Other species have been sporadically reported but were so far not proven by genetic data.

1. Clostridium stercorarium: A Soluble Cellulase System

Only two enzymes that hydrolyze cellulose could be isolated from the culture supernatant of \textit{C. stercorarium} NCIB 11764, the type strain (Bronnenmeier \textit{et al.}, 1990, 1991). The search for other components in culture supernatants was not successful (Bronnenmeier, personal communication). Extensive screening of genomic libraries in cosmid and bacteriophage Lambda vectors also revealed not more than two genes involved in the production of cellulases: \textit{celZ} and \textit{celY} (Schwarz \textit{et al.}, 1989). These genes coded for the cellulases Cel9Z, an endo-glucanase with some exo-glucanase activity, and Cel48Y, an exoglucanase. The gene products were identical to the previously isolated Avicelases I and II, respectively (Bronnenmeier \textit{et al.}, 1997; Jauris \textit{et al.}, 1990). Both enzymes are modular proteins consisting of an N-terminal catalytic module, a CBM3c module, and a binding module of family 3b with high affinity for crystalline cellulose (CBM3b) (Fig. 5). In both enzymes the presence of a cellulose binding module enhances the local concentration of the enzymes on the substrate surface and is necessary for the activity on the solid substrate.

The arrangement CBM3c-CBM3b occurs in both enzymes. In addition, both modules in inverse order are located downstream in Cel9Z, resulting in the order CBMc-b-b-c (Jauris \textit{et al.}, 1990). The CBM3c module adjacent to the catalytic module of Cel9Z is essential for the enzymatic activity at elevated temperature and has no experimentally

![Fig. 5. Structure of the cellulase cluster in the \textit{Clostridium stercorarium} genome. The order and approximate size (numbers) of the genes \textit{celY} and \textit{celZ}, the direction of transcription (arrows), and the module architecture of the cellulases is indicated.](image)
detectable binding activity (Riedel et al., 1998a). It may function, however, in feeding the emerging cellulose molecule into the active site pocket and thus determining the orientation of the processive activity of the enzyme, as described previously.

No indication for a cellulosomal structure could be found for the cellulolytic activities of C. stercorarium: (1) both isolated enzymes were “free” exo-enzymes and could be purified as single proteins from the culture supernatant; (2) despite the isolation of dozens of glucanase clones, no gene other than celY and celZ was obtained (unpublished observation); and (3) none of the genes for extracellular enzymes cloned so far from C. stercorarium contained the dockerin module, which is typical for all hitherto identified cellulosome components (see below). Nevertheless, the two cellulases, Cel48Y and Cel9Z, constitute a functionally complete enzyme system in which both components are essential for the hydrolysis of crystalline cellulose.

The combination of exo- and endo-glucanases is typical for the soluble enzyme systems of the cellulolytic bacteria. Moreover, GH9 and GH48 enzymes are the major components in all bacterial cellulase systems. Cel48Y and Cel9Z show a distinct synergistic interaction in the degradation of microcrystalline cellulose, which is dependent on the ratio of the two enzymes and on the type of the cellulosic substrate (Riedel et al., 1997). The synergism depends on the simultaneous presence of both enzymes and is not expressed by sequential addition of the two activities.

To investigate this synergism further, a bifunctional hybrid, Cel48Y-Cel9Z, was constructed with the structure GH48-CBM3c-CBM3b-GH9-CMB3c (Riedel et al., 1998b). The large fusion protein (170 kDa) was expressed in E. coli and purified. It exhibited endo- as well as exo-glucanase activity, and it retained the thermostability of the parent enzymes. But its cellulolytic activity was threefold to fourfold higher than the sum of the individual enzyme activities, underscoring the effect of packing two catalytic activities physically together.

A natural hybrid enzyme, CelA, with a similar structure was identified in the extremely thermophilic, nonclostridial bacterium Anaerocellum thermophilum (Zverlov et al., 1998a). It also consists of GH9 and GH48 modules connected to CBM3 modules. It is able to hydrolyze microcrystalline cellulose. Both catalytic modules showed sequence identities of about 70% to the C. stercorarium cellulases Cel48Y and Cel9Z, respectively, and were active if expressed separately as recombinant proteins.

C. stercorarium so far is the only Clostridium species shown to have a soluble cellulase system comparable to that of other bacterial genera,
especially the actinomycetes, and does not possess a cellulosome. Other cellulolytic clostridia have not been investigated in such detail (Schwarz, 2003). Whether *C. stercorarium*, similar to *C. acetobutylicum*, lost a cellulosome during evolution and became a hemicellulose specialist or if it simply did not acquire the cellulosomal genes can only be determined by genomic sequencing. The reduction or development to the simplest known cellulase systems with only two components is astonishing. This is a singular observation among the bacteria and could be an encouraging model for an industrial cellulase preparation. However, cellulose hydrolysis in *C. stercorarium* is comparatively slow and incomplete, although it allows growth of the cells on pure cellulose as sole carbon source.

*C. stercorarium* feeds the cello-oligosaccharides into its catabolism by phosphorylation through the cellobiose phosphorylase CepA and cellodextrin phosphorylase CepB, rather than by the energy-wasting hydrolytic β-glucosidase action (Reichenbecher et al., 1997). These enzymes can be isolated from the cell extract and seem to be located intracellularly.

2. Clostridium thermocellum Cellulosomes

The cell-free culture supernatant of *C. thermocellum* contains a cocktail of enzymes with high cellulolytic activity on crystalline cellulose (Fig. 6). Dependent on the growth phase of the bacterial culture, the majority of this activity is cell bound. Whether enriched

---

**Fig. 6.** (Left) Hydrolysis of crystalline cellulose by *C. thermocellum*. Cellulose powder (MN300) in a thin layer of agar-medium on top of an agar plate is completely degraded around *C. thermocellum* colonies; the dark background is shining through. (Right) A strip of Whatman No. 1 filter paper is decomposed by a growing culture (left to right: uninoculated, culture after 1 and 2 days).
from the cell surface or from the culture fluid, there are more than 25 different extracellular enzymes visible in a denaturing electrophoresis gel. Many of these proteins are hydrolytically active on cellulose. However, the complete cellulolytic activity could not be reconstituted from single, isolated components (Bhat et al., 1994; Morag et al., 1996; Wu et al., 1987). Nevertheless, a huge multienzyme complex was isolated, which in intact form was highly active on crystalline cellulose. It was called a cellulosome and could be isolated either from the culture supernatant or from the cells (Lamed et al., 1983). A purification method called “affinity digestion” uses the adsorption of the cellulosomes to cellulose fibers, washing the cellulose-cellulosome complex, followed by the complete hydrolysis of the cellulose. Cellulosomes are then purified by gel filtration chromatography (Morag et al., 1992).

This purification scheme makes use of the basic characteristics of the cellulosomes: They not only bind to the substrate but also to the cell surface and thus form a bridge that holds the cell on its much larger substrate. This is energetically favorable because a sufficient enzyme concentration can easily be reached on the cell surface without producing a high amount of extracellular protein. Furthermore, the enzymatic components necessary for optimum synergism stay in close proximity, and the products of hydrolysis are present in high concentration near the cell surface, ready for uptake by the enzyme-producing cells. They are not “wasted” for other competing bacteria (Lynd et al., 2002). The large protein complexes on the outer surface of the bacterial cells could be made visible by electron microscopy, but only after fixation with cationized ferritin (Bayer and Lamed, 1986; Madkour and Mayer, 2003; Mayer et al., 1987). The attachment of the cells to the substrate via the cellulosomes was also observed (Bayer et al., 2000).

Isolated, purified cellulosomes of C. thermocellum vary in size depending on the strain (from 2.0 to 6.5 Mda) and may even aggregate to large supercomplexes, called poly-cellulosomes (up to 100 Mda). Approximately 25 genes for cellulosomal genes have been isolated from genomic libraries by random screening for hydrolytic activity (for review see Schwarz, 2001). There is no proof yet for all gene products that they are actually present in the cellulosome. Moreover, a number of components surprisingly have a hydrolytic activity that apparently has nothing to do with cellulose degradation. These components degrade other polysaccharides in biomass—such as mixed-linkage β-glucan, pectin, xylan, mannan or chitin—which in natural substrates wrap the cellulose crystals (Blum et al., 2000; Kurokawa et al., 2001; Spinnler et al., 1986; Zverlov et al., 1994, 2002a). This hydrolytic activity is in contrast to the lack of fermentation ability of
C. thermocellum for pentoses: evidence is accumulating that the cello-
dextrins derived from cellulose (not glucose or cellobiose) are the best
substrate providing the most energy (Lynd et al., 2002). Hemicellulases
seem to have an accessory function in providing access to the preferred
substrate. However, it was shown recently that C. thermocellum also
ferments $\beta$-1,3-glucan (Fuchs et al., 2003).

Gene cloning, together with immunological investigations, provided
clues for the presence of noncatalytic proteins in the cellulosomes that
are involved in structure-forming or other functions. Most important
was the discovery of a scaffolding protein (CipA, “cellulosome integrat-
ing protein”), the so-called scaffoldin, which has nine docking sites
called cohesins (Fujino et al., 1992; Gerngross et al., 1993). The binding
partner on the catalytic cellulosome components is a conserved twofold
repeat of 24 amino acid residues, the dockerin (type I) (Tokatlidis et al.,
1991). The dockerin sequences of the different cellulosomal genes are
well conserved. Slight differences together with the differences in the
sequence of the cohesins may lead to preferences of specific celluloso-
mal components to specific sites. This assumption is not in contrast to
experimental results showing that a single cellulosomal component can
bind to different cohesins (Fierobe et al., 2001). X-ray analysis of the
cohesin structure revealed a flat binding area exposing surface residues
for relatively unspecific interaction with the dockerins (Lytle et al.,
2000; Mechaly et al., 2000; Shimon et al., 1997).

The scaffoldin CipA brings together nine catalytic components in
close proximity and thus may stimulate the synergism between the
enzymes. However, there are still other modules besides the cohesins
in the scaffoldin: a cellulose binding module (CBM3a) on the C-termi-
nus for tight binding to a crystalline cellulose surface and a dockerin
module on the N-terminus. This dockerin type II is not as closely
related to the dockerins of the catalytic components as they are to each
other. Type I and II dockerins bind to their complementary cohesin
types but are not cross-reacting (Fierobe et al., 2001). The binding
partner for the CipA dockerin was found in other noncatalytic extra-
cellular proteins, SdbA and OlpB (and probably others), which carry
S-layer homologous modules, anchoring the proteins in the bacterial
cell wall, and cohesins of type II (Leibowitz and Béguin, 1996; Lemaire
et al., 1995). The role of other outer layer proteins, which were identi-
fied as reading frames in the genome of C. thermocellum, has still to
be elucidated.

To date, only one of the dockerin(I)-bearing components of the cel-
lusosome was obviously noncatalytic: CseP with homology to CotH, a
spore coat forming structural protein in Bacillus subtilis (Table IV;
<table>
<thead>
<tr>
<th>Reading frame/function</th>
<th>Structure$^a$</th>
<th>Ref.$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structural component</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. CipA (c) scaffoldin, Cthe1933-1930</td>
<td>2(Coh1)-CBM3a-7(Coh1)-X2-Doc2</td>
<td>(Fujino et al., 1992; Zverlov et al., in prep.)</td>
</tr>
<tr>
<td><strong>GH2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Cthe1580</td>
<td>GH2-CBM6-Doc1</td>
<td></td>
</tr>
<tr>
<td><strong>GH5</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. CelO cellobiohydrolase, Cthe1674</td>
<td>CBM3b-GH5-Doc1</td>
<td>Zverlov et al., 2002b</td>
</tr>
<tr>
<td>4. Cthe1575</td>
<td>GH5-CBM6-Fn3-Doc1</td>
<td></td>
</tr>
<tr>
<td>5. CelB endoglucanase, Cthe0374</td>
<td>GH5-Doc1</td>
<td>Grepinet and Béguin, 1986</td>
</tr>
<tr>
<td>6. CelG endoglucanase, Cthe0885</td>
<td>GH5-Doc1</td>
<td>(Lemaire and Béguin, 1993)</td>
</tr>
<tr>
<td>7. Cthe0444</td>
<td>GH5-Doc1</td>
<td></td>
</tr>
<tr>
<td><strong>GH8</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. CelA endoglucanase, Cthe0722</td>
<td>GH8-Doc1</td>
<td>(Béguin et al., 1985; Zverlov et al., in prep.)</td>
</tr>
<tr>
<td><strong>GH9</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. CbhA cellobiohydrolase, Cthe1222</td>
<td>CBM4-Ig-GH9-2(Fn3)-CBD3b-Doc1</td>
<td>(Zverlov et al., 1998b)</td>
</tr>
<tr>
<td>10. CelK cellobiohydrolase, Cthe2598</td>
<td>CBM4-Ig-GH9-Doc1</td>
<td>(Zverlov et al., 1999)</td>
</tr>
<tr>
<td>11. CelD endoglucanase, Cthe0968</td>
<td>Ig-GH9-Doc1</td>
<td>Joliff et al., 1986</td>
</tr>
<tr>
<td>12. Cthe1953</td>
<td>GH9-CBM3c-CBM3b-Doc1</td>
<td></td>
</tr>
<tr>
<td>13. Cthe0850</td>
<td>GH9-CBM3c-CBM3b-Doc1</td>
<td></td>
</tr>
<tr>
<td>14. CelN endoglucanase, Cthe1222</td>
<td>GH9-CBM3c-Doc1</td>
<td>(Zverlov et al., in prep.)</td>
</tr>
<tr>
<td>15. CelR endoglucanase, Cthe1837</td>
<td>GH9-CBM3c-Doc1</td>
<td>(Zverlov et al., in prep.)</td>
</tr>
<tr>
<td>16. CelQ endoglucanase, Cthe0300</td>
<td>GH9-CBM3c-Doc1</td>
<td>(Arai et al., 2001)</td>
</tr>
<tr>
<td>17. CelF endoglucanase, Cthe0382</td>
<td>GH9-CBM3c-Doc1</td>
<td>Navarro et al., 1991</td>
</tr>
</tbody>
</table>

(continued)
### Xylanases

21. XynD xylanase, Cthe0688
   - CBM22-GH10-Doc1  
   - + (Zverlov et al., in prep.)

22. XynC xylanase, Cthe0626
   - CBM22-GH10-Doc1  
   - + (Hayashi et al., 1997)

23. XynA, XynU xylanase, Cthe1161
   - GH11-CBM4-Doc1-NodB  
   - + (Hayashi et al., 1999)

24. XynB, XynV xylanase
   - GH11-CBM4-Doc1  
   - + (Hayashi et al., 1997)

### Other hemicellulases

25. LicB lichenase
   - GH16-Doc1  
   - + (Zverlov et al., 1994a,b)

26. ChiA chitinase
   - GH18-Doc1  
   - + (Zverlov et al., 2002a)

27. ManA mannanase, Cthe0533
   - CBM-GH26-Doc1  
   - + (Halstead et al., 1999)

28. Cthe2142
   - GH26-Doc1

29. Cthe1127
   - GH30-CBM6-Doc1

30. Cthe2333
   - GH53-Doc1

31. Cthe0269
   - GH81-Doc1

### Putative glycosidases

32. Cthe1665
   - GH39-2(CBM6)-Doc1

33. Cthe1579
   - GH43-CBM6-Doc1

34. Cthe0268
   - GH43-CBM13-Doc1

35. Cthe0484
   - GH43-2(CBM6)-Doc1

### GH48

36. CelS exoglucanase, Cthe0939
   - GH48-Doc1  
   - + (Wang et al., 1993)

### Xyloglucanase

37. XghA xyloglucanhydrolase, Cthe2335
   - GH74-CBM2-Doc1  
   - + (Zverlov et al., in prep.)

---

*(continued)*
<table>
<thead>
<tr>
<th>Reading frame/function</th>
<th>Structure</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Putative carbohydrate esterases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38. Cthe0066</td>
<td>Fn3-CE12-Doc1-CBM6-CE12</td>
<td></td>
</tr>
<tr>
<td>39. Cthe1577</td>
<td>CE1-CBM6-Doc1</td>
<td></td>
</tr>
<tr>
<td><strong>Putative pectinases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40. Cthe2008</td>
<td>GH28-Doc1</td>
<td></td>
</tr>
<tr>
<td>41. Cthe2236</td>
<td>PL1-Doc1-CBM6</td>
<td></td>
</tr>
<tr>
<td>42. Cthe1810</td>
<td>&lt;-Doc1-CBM6-PL9</td>
<td></td>
</tr>
<tr>
<td>43. Cthe2234</td>
<td>PL10-UN-Doc1</td>
<td></td>
</tr>
<tr>
<td>44. Cthe0702</td>
<td>Doc1-CBM6-PL11</td>
<td></td>
</tr>
<tr>
<td><strong>Multifunctional components</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45. CelJ cellulase, Cthe0301</td>
<td>X-Ig-GH9-GH44-Doc1-X</td>
<td>+ (Ahsan et al., 1996)</td>
</tr>
<tr>
<td>46. CelH endoglucanase, Cthe0837</td>
<td>GH26-GH5-CBD9-Doc1</td>
<td>Yaguee et al., 1990</td>
</tr>
<tr>
<td>47. Cthe1667</td>
<td>GH30-GH54-GH43-Doc1</td>
<td></td>
</tr>
<tr>
<td>48. Cthe1211</td>
<td>GH54-Doc1-GH43</td>
<td></td>
</tr>
<tr>
<td>49. Cthe1666</td>
<td>GH54-GH43-Doc1</td>
<td></td>
</tr>
<tr>
<td>50. XynZ xylanase, Cthe1691</td>
<td>CE1-CBM6-Doc1-GH10</td>
<td>+ (Grepinet et al., 1988; Zverlov et al., in prep.)</td>
</tr>
<tr>
<td>51. XynY xylanase, Cthe2036</td>
<td>CBM22-GH10-CBM22-Doc1-CE1</td>
<td>Fontes et al., 1995</td>
</tr>
<tr>
<td>52. CelE endoglucanase, Cthe0940, Cthe2702, Cthe2514</td>
<td>GH5-Doc1-CE2</td>
<td>+ (Hazlewood et al., 1990)</td>
</tr>
<tr>
<td><strong>Putative protease inhibitors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53. Cthe1412</td>
<td>Fn3-Doc1-serpin</td>
<td></td>
</tr>
<tr>
<td>54. Cthe1413</td>
<td>Doc1-serpin</td>
<td></td>
</tr>
<tr>
<td><strong>Components with unknown function</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55. Cthe0694</td>
<td>2(UN)-UN-UN(CelP550–870)-Doc1</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
Zverlov et al., 2003). Its suspected role in the three-dimensional stabilization of the cellulosome or in the multimerization of different cellulosomal particles has still to be demonstrated. Other hypothetical non-catalytic proteins in the cellulosome were identified among the reading frames obvious from the genomic sequence. Two of them carry protease inhibitor modules (#53 and #54 in Table IV); others carry modules with unknown function (#55–62). These proteins, however, have not been shown to be expressed; their function can only be speculated.

The CBM3a module locates the whole cellulosome firmly on the surface of the crystalline substrate. The large size of the structure and the spacer modules (X modules) may give single components some flexibility to successively attack a number of sites around that location. In a cellulosome it is not necessary that each single catalytic component has its own CBM as with soluble enzyme systems. Indeed, only a few components have CBMs, many of them with a different binding
specificity compared to the CipA-CBM3. The role of the CipA-CBM3a was shown experimentally by fusing the recombinant Cel5E endo-glucanase with the CipA-CBM ([GH5-DD]CelE-CBM3CipA). The hydrolytic activity of this hybrid protein was strongly enhanced only on crystalline cellulose but not on amorphous cellulose. No stimulation was found if Cel5E and CBMCipA were added separately to the crystalline substrate (Ciruela et al., 1998). This is an indication that a structural arrangement like that given in the cellulosome is necessary for the synergistic effect.

Only a few genes of cellulosomal components were identified by targeted immunological methods (e.g., cipA) or by partial protein sequencing. And although a number of laboratories have isolated clones active against polysaccharide substrates from genomic libraries, the list of cloned components is surely incomplete (albeit highly repetitive) (Guglielmi and Béguin, 1998). Cloned genes whose presence in the cellulosome has been demonstrated are shown in Table IV. Other genes may be inactive (not expressed) or expressed only under certain circumstances (e.g., by induction through specific substrates or substances). On the other hand, some genes may be hard to screen for because of a lack of a suitable substrate or its unusual activity profile, or may be hard to clone because their DNA is toxic to or not well expressed in the cloning host or is easily degraded. Using a pure cloning approach, we will never get all genes and we will not know if a cloned gene is expressed and incorporated in the cellulosome.

Only with the emerging genomic sequence of the type strain ATCC 27405 did a rational approach toward a complete list of the cellulosomal components become available. The purified cellulosomes have been denatured and the components separated by proteomic methods. The protein spots were identified by comparing the MALDI-TOF data with the translated genomic sequence, which, however, is still fragmentary and full of alignment errors (Zverlov et al., in preparation). The major components of the cellulosome (as it is produced from cellobiose grown cells) were identified by this method. Surprisingly (but not unexpectedly), the first results showed that three of the major proteins had not been cloned before. The advantage of this protein-directed approach is that only proteins show up that are translated, secreted, and correctly incorporated in the huge extracellular complex. At the same time, the distribution pattern of major and minor components can be addressed, together with an analysis of the possible shift in composition that is due to changes in the substrate.

Another approach possible with the genomic sequence is the compilation of a list of the proteins that contain a dockerin or a cohesin
module and thus must have something to do with the cellulosome. A provisional list is shown in Table IV. This list is still incomplete, and in some cases the DNA sequence shows an unverified composition of modules that is due to the unfinished character of the genomic sequence. Conspicuous is the large number of hemicellulase (xylanase, β-1,3/1,4- and β-1,6-glucanase, chitinase, mannanase, galactanase, xyloglucanase), glycosidase (β-xyllosidase, α-arabinofuranosidase) and esterase genes that did not show up in the genomic library screening, possibly because they have not been screened for. A complete set of genes potentially involved in pectin degradation showed up (pectate lyase, polygalacturonase, esterase) (Zverlov et al., in preparation). The reading frames with homology to protease inhibitors may have to do with the suspected proteolysis of the cellulosomes in outgrowing cultures of \( C. \) thermocellum.

Interesting is the occurrence of genes containing more than one catalytic module—for example, combining -xyllosidase and α-arabinofuranosidase modules (GH43/54) or xylanase and arabinoxylan-esterase modules (GH10/CE1), which are functionally related. This enhances the capacity of the cellulosome with its only nine cohesin sites on the scaffoldin and reminds the module arrangement of the extracellular enzymes of the \( Caldicellulosiruptor\)-\( Anaerocellum \) group. The number of modules with unknown or unrelated sequences is relatively small compared with the average situation in reading frames obtained from genomic sequencing. This reflects the high intensity of research on glycosyl hydrolases and the high coverage of cloned genes. Further investigations of the composition of cellulosomes have to reveal the presence of the encoded proteins in the cellulosome of \( C. \) thermocellum.

In addition to the cellulosomal proteins, membrane binding proteins are essential for the attachment of the cellulosome (e.g., SdbA, OlpA, OlpC) or single proteins to the cell wall. These may also play a role in the assembly of the huge protein complex. Many of these proteins may be detected by subproteomic approaches (e.g., with isolated bacterial cell walls).

3. Does Clostridium thermocellum Have a Soluble Cellulase System Too?

In addition to the cellulosome, other soluble, noncellulosomal enzymes have been cloned from \( C. \) thermocellum (Table V). Cel5C and Lic16A seem to have a role in the hydrolysis of soluble β-1,3(4)-glucans and may be restricted to substrates connected with the cellulose surface to which they bind via their CBMs. Lic16A is, in addition, able to degrade β-1,3-glucan (Fuchs et al., 2003). CelM was reported to have
endo-glucanase activity (Kobayashi et al., 1993); however, its role remains doubtful because the sequence is not homologous to any other endo-glucanase but to aminopeptidases.

The soluble enzyme component Cel9I has a high degree of similarity to Cel9Z of C. stercorarium (in structure and sequence homology), which is also a soluble cellulase. Both enzymes are processive endo-glucanases (Riedel et al., 1998a; Zverlov et al., 2003). In C. stercorarium, Cel9Z together with Cel48Y is able to degrade crystalline cellulose in synergistic cooperation (Riedel et al., 1998b). Indeed, database screening of the unfinished C. thermocellum genome revealed a reading frame homologous to Cel48Y that also is not connected to a dockerin module (Table V). Although the expression of both genes has not been investigated so far, C. thermocellum seems to possess a second, soluble cellulase system besides the cellulosome. C. thermocellum is so far the only bacterium having more than one gene with a GH48 module.

Once the genome sequence is complete, more reading frames may be identified that are not connected to a dockerin module and thus are not located in the cellulosome. The expression of such genes has to be verified—for example, by concentration of culture supernatants and analysis of the activities of their proteins. Intracellular cellulolysis related enzymes are, for example, the glucosidases BglA, BglB, and the cellobiose and cellodextrin phosphorylases CdP and CbP. The latter enzymes enhance the energetical efficiency of cellulose hydrolysis for the cell considerably (Lynd et al., 2002).

**TABLE V**

**Noncellulosomal Extracellular Proteins of C. thermocellum**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Structure</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lic16A, Cthe0535</td>
<td>SLH(1–3)-GH16-CBM4a(1–4)</td>
<td>Fuchs et al., 2003</td>
</tr>
<tr>
<td>Cel5C, Cthe0537</td>
<td>GH5</td>
<td>Schwarz et al., 1988</td>
</tr>
<tr>
<td>Cel9I, Cthe1219</td>
<td>GH9-CBM3c-CBM3b</td>
<td>Hazlewood et al., 1993; Zverlov et al., 2003</td>
</tr>
<tr>
<td>CelM, Cthe0317</td>
<td>homologous to peptidases</td>
<td>Kobayashi et al., 1993</td>
</tr>
<tr>
<td>XynX</td>
<td>CBM(TSM)-GH10-CBM9-CBM9</td>
<td>Kim et al., 2000</td>
</tr>
<tr>
<td>Cthe0247</td>
<td>GH48-CBM3c-CBM3b (hom. to Cel48Y/C. stercorarium)</td>
<td>This chapter</td>
</tr>
</tbody>
</table>

Abbreviations as in Table IV. TSM, thermostabilizing module.
4. Cellulosomes of Mesophilic Clostridia

A number of mesophilic clostridia have been found to produce cellulosomes for the hydrolysis of crystalline cellulose. Cellulolytic genes were cloned from *Acetivibrio cellulolyticus* (belonging to the genus *Clostridium* by 16S rRNA sequencing), *C. acetobutylicum*, *C. cellulolyticum*, *C. cellulovorans*, and *C. josui*. A recent review nicely summarizes the data on mesophilic cellulosomes (Doi *et al.*, 2003).

In contrast to the scattered gene location in the genome of *C. thermo- cellum*, the cellulosomal genes of the mesophilic clostridia are basically arranged as large gene clusters with nine (e.g., in *C. cellulovorans*) or even more genes as in *C. cellulolyticum* (Bélaich *et al.*, 2002; Tamaru *et al.*, 2000). The arrangement of the genes in the clusters is surprisingly similar: it starts with the scaffoldin gene that is followed immediately by the only GH48 enzyme gene, the major exo-glucanase component (compiled in Schwarz, 2001). These two components seem to be indispensable for the cellulolytic function of a cellulosome and probably are expressed with the highest efficiency.

It is interesting to note that in the scaffoldins of the cellulosomes of *C. cellulovorans*, *C. cellulolyticum*, *C. josui*, and *C. acetobutylicum* (but not of *C. thermo- cellum*) hydrophobic modules were observed (up to five copies: formerly called X-modules). This module is not present in any catalytic cellulosomal components. The 3-D structure of the X2 module from CipC of *C. cellulolyticum* was determined to be an immunoglobulin-like fold with a remarkable conformational stability. It may have a function as a structural linker and a solubility enhancer between other modules (Mosbah *et al.*, 2000).

The majority of the enzyme genes contain GH9 and GH5 catalytic modules, both of which contain endo- and exo-glucanases. Besides the catalytic module, noncatalytic domains play an important role in defining the function of an enzyme. The architecture of the cellulosomal components is summarized in Table VI. Given the similarity in the operon structures, it is not surprising that the module composition of single components is also quite similar. For almost any module architecture type present in one bacterium, a similar component is found in almost any other cellulosome. This list is not complete, as many components may not yet have been identified and cloned. In addition to the gene clusters, cellulosomal components are also encoded by unlinked monocistronic genes.

It is rather improbable that a complicated enzyme complex like the cellulosome would be invented independently by different bacteria. Heterologous gene transfer of gene clusters followed by slight
<table>
<thead>
<tr>
<th>Structure</th>
<th>Possible function</th>
<th><em>C. acetobutylicum</em></th>
<th><em>C. cellulolyticum</em></th>
<th><em>C. cellulovorans</em></th>
<th><em>C. josui</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>GH48-DD</td>
<td>Exoglucanase</td>
<td>CelF</td>
<td>CelF</td>
<td>ExgS</td>
<td>CelD</td>
</tr>
<tr>
<td>GH5/DD</td>
<td>Endoglucanase/mannanase</td>
<td>CelA, ManA</td>
<td>ManK, CelD, CelA</td>
<td>ManA, EngB</td>
<td></td>
</tr>
<tr>
<td>(SLH)3-GH5-X-DD</td>
<td>Endoglucanase</td>
<td>CAC3469</td>
<td></td>
<td>EngE</td>
<td></td>
</tr>
<tr>
<td>GH8-DD</td>
<td>Endoglucanase</td>
<td>CelC</td>
<td></td>
<td></td>
<td>CelB</td>
</tr>
<tr>
<td>GH9-DD</td>
<td>Endoglucanase</td>
<td>CelL</td>
<td>CelM</td>
<td>EngL</td>
<td></td>
</tr>
<tr>
<td>GH9-CBM3c-DD</td>
<td>Endoglucanase</td>
<td>CelH, CelG</td>
<td>CelG, CelH, CelJ</td>
<td>EngH</td>
<td>CelE</td>
</tr>
<tr>
<td>CBM2-GH9-DD</td>
<td>Endoglucanase</td>
<td></td>
<td></td>
<td>EngY</td>
<td></td>
</tr>
<tr>
<td>CBM4-Ig-GH9-DD</td>
<td>Endoglucanase</td>
<td>CelE</td>
<td>CelE</td>
<td>EngK, EngM</td>
<td></td>
</tr>
<tr>
<td>GH11-DD-CE4</td>
<td>Endoxylanase</td>
<td>CelE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GH27-DD</td>
<td>α-Galactosidase</td>
<td></td>
<td></td>
<td></td>
<td>XynA</td>
</tr>
<tr>
<td>GH44-DD</td>
<td>Endoglucanase</td>
<td>EngA</td>
<td></td>
<td></td>
<td>AgaA</td>
</tr>
<tr>
<td>GH74-DD</td>
<td>“Sialidase”</td>
<td></td>
<td></td>
<td>CAC0919</td>
<td></td>
</tr>
<tr>
<td>X-CBM2-PL9-DD</td>
<td>Pectate lyase</td>
<td></td>
<td></td>
<td></td>
<td>PelA</td>
</tr>
</tbody>
</table>

The components are sorted by structure types. The modules are: GH, glycosyl hydrolase family; DD, dockerin module; SLH, S-layer homologous module; CBM, carbohydrate binding module; Ig, immunoglobulin-like fold; X, hydrophobic module; PL, pectate lyase module; CE, feruloyl esterase module.
rearrangements and adaptation of new components is a probable explanation. This is corroborated by the aforementioned high similarity of the arrangement of the functionally related genes. Moreover, at the 3’ end of the C. cellulovorans gene cluster a transposase gene was observed as a possible source of the gene transfer (Tamaru et al., 2000).

Although C. acetobutylicum does not degrade cellulose or produce an extracellular cellulosome, some components of the cellulosome were detected in culture supernatants. The presence of cellulosomal gene clusters became obvious during the annotation of the genomic sequence (Nölling et al., 2001; Sabathe et al., 2002). For a bacterium like C. acetobutylicum, which can grow on starch and different oligosaccharides and monosaccharides, one might expect it to be an unnecessary burden to produce such a large and costly enzyme complex as the cellulosome. It might have acquired the genes in its evolution a long time ago but has managed to shut down its production. Single enzymatic components are helpful for the breakdown of biomass in its natural habitat and are produced at a low rate, but the complete set of enzymes is no longer produced and the genes underwent permanent changes. In the genome we still can see the remnants of this very recent evolutionary process.

In an attempt to create a solventogenic bacterium that could use lignocellulosic biomass as substrate for the production of organic solvents on an industrial scale, a number of frame shifts and deletions in the cellulosomal genes were corrected and promoter sequences optimized (Sabathe et al., 2002). However, the bacterium still could not hydrolyze cellulose, and no enzymatically active large protein complex was obtained, although single components were active as recombinant enzymes expressed in other bacteria, and a mini-cellulosome was secreted by the Clostridium (Sabathe and Soucaille, 2003). The search for a cellulose-degrading, solvent-producing bacterium goes on (Montoya et al., 2001). Such a strain might be found among the mesophilic, cellulytic clostridia that are widespread in natural environments.

C. XYLAN AND HEMICELLULOSE DEGRADATION

The major components of the noncellulosic polysaccharides in plant biomass are xylans, at least in angiosperms, where they account for 20–30% of the dry weight of woody tissue (Aspinall, 1980). Xylan has a heterogeneous composition, consisting of a homopolymeric, linear backbone of α-1,4-linked D-xylopyranosyl residues, which are substituted with α-1,3-linked L-arabinofuranosyl, α-1,2-linked 4-O-methylglucuronic acid residues, or other sugar residues. Substitutions with acetic, p-coumaric, or ferulic acid are common.
The enzymatic hydrolysis of arabinoxylan and glucuronoxylan requires the activity of backbone and side-chain cleaving activities. Involved are endo-$\beta$-1,4-xylanases (1,4-$\beta$-d-xylan xylanohydrolase, EC 3.2.1.8), possibly exo-$\beta$-1,4-xylanases (1,4-$\beta$-d-xylan xylohydrolase), and $\beta$-d-xylosidases (1,4-$\beta$-d-xyloside xylohydrolase, EC 3.2.1.37); the side groups and substituents are removed by $\alpha$-l-arabinofuranosidases (EC 3.2.1.55), $\alpha$-d-glucuronosidases (EC 3.2.1), and esterases such as feruloyl esterase (Donaghy et al., 2000).

The best-investigated xylan hydrolyzing enzyme system within the clostridia is that of *C. stercorarium*. Hydrolytic proteins from the culture supernatant of the type strain NCIMB 27405 have been separated and characterized; some selected proteins are also from another strain, the Japanese isolate F-9 (Berenger et al., 1985; Bronnenmeier et al., 1990; Hayashi et al., 1997, 1999). An endo-xylanase and a celloxylanase—both as multiple protein species caused by partial degradation—a $\beta$-xylosidase, an $\alpha$-arabinofuranosidase, and a feruloyl esterase were purified. They correspond to the proteins Xyn10A, Xyn11C, Bx13B, and Arf51B, respectively (Adelsberger et al., submitted). The gene of the extracellular feruloyl esterase has not yet been identified (Donaghy et al., 2000). Ali et al. (1999) showed the presence of Xyn11C in the culture supernatant and most probably on the cell surface of strain F-9 by immunological methods. Another xylanase, Xyn11B, and an $\alpha$-glucuronidase could not be identified in culture supernatants but were isolated from cell extracts and seem to be located intracellularly (Bronnenmeier et al., 1995; Sakka et al., 1996).

The complete set of enzymes sufficient to hydrolyze arabinoxylan could be cloned from the genome of *C. stercorarium*. The extracellular xylanases are surprisingly complex enzymes consisting of four and five modules (Table VII). It is conspicuous that only the intracellular xylanase Xyn11B in *C. stercorarium* does not have any additional, noncatalytic modules, as extracellular enzymes often possess. The cellulose binding modules CBM6 of Xyn11A and either CBM22 or CBM9 of Xyn10C show strong binding activity to crystalline cellulose (Ali et al., 2001; Bronnenmeier et al., 1996; Sun et al., 1998). Cellulose binding could fix the xylanases firmly to the surface of the plant cell wall and thus keep them in the vicinity of their substrate xylan. The xylanases that have these binding modules might help the cellulolytic bacteria to polish the cellulose fibers from the attached xylan, which then enables the degradation of the cellulose crystals. The cellodextrins derived from cellulose are of a much higher metabolic value for the bacterium than xylan and probably are the preferred substrate. Xylan is nevertheless a possible fermentation substrate for *C. stercorarium*.
The xylanase Xyn10C seems to be able to form an anchor between the bacterial cell and the cellulosic surface. Ali et al. (2001) have shown that Xyn11C is one of the major surface layer proteins. Its high affinity to cellulose brought about co-precipitation of bacterial cells washed of surface layer proteins, but only if the SLH modules were present. This indicates that Xyn10C on one hand binds to the surface layer with virtue of its SLH modules and on the other to cellulose with its CBMs. The binding of the cell to its substrate has a great advantage for the bacterium: a lower amount of exo-enzymes to be produced for saturation of the substrate and a higher concentration of hydrolysis products near the cell surface combined with diminished loss by diffusion.

Xyn11A, Xyn10C, Arf51B, and Bx13B are major components in the C. stercorarium culture supernatant. These enzymes, and not the products of other genes also isolated from the genome, are sufficient to hydrolyze arabinoxylan effectively and completely to its monomers (Adelsberger et al., submitted). Arf51B releases /α/-arabinofuranosyl residues from the intact arabinoxylan, as well as from the oligosaccharides produced by the action of the endo-xylanases Xyn11A and/or Xyn10C. The resulting xylo-oligosaccharides are degraded to xylose by Bx13B. The point of attack of the two endo-xylanases seems to be different: the produced oligosaccharides differ as well in size as in

### Table VII

**Extracellular Enzymes of C. stercorarium Involved in the Xylanolytic Pathway**

<table>
<thead>
<tr>
<th>Designation</th>
<th>NCIMB Acc. No.</th>
<th>Design. F-9</th>
<th>Synonym</th>
<th>Acc. No. F-9</th>
<th>Module architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xyn11A</td>
<td>AJ508403</td>
<td>Xyn11A</td>
<td>XynA</td>
<td>D13325</td>
<td>GH11-CBM6-CBM6-(CBM6)</td>
</tr>
<tr>
<td>Xyn10B</td>
<td>AJ508407</td>
<td>Xyn10C</td>
<td>XynB</td>
<td>D12504</td>
<td>GH10</td>
</tr>
<tr>
<td>Xyn10C</td>
<td>AJ508408</td>
<td>Xyn10B</td>
<td>XynC</td>
<td>AB024743</td>
<td>CBM22-CBM22-GH10-CBM9-SLH-SLH</td>
</tr>
<tr>
<td>Arf51B</td>
<td>AF00264</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>GH51</td>
</tr>
<tr>
<td>Bx13B</td>
<td>AJ508405</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>GH3</td>
</tr>
<tr>
<td>Bg13Z</td>
<td>Z94045</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>GH3</td>
</tr>
<tr>
<td>Ram78A</td>
<td>AJ238748</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>GH78</td>
</tr>
</tbody>
</table>

The accession number of the sequence in the GeneBank is given, as well as the synonyms for the homologous sequences from strain F-9.

The xylanase Xyn10C seems to be able to form an anchor between the bacterial cell and the cellulosic surface. Ali et al. (2001) have shown that Xyn11C is one of the major surface layer proteins. Its high affinity to cellulose brought about co-precipitation of bacterial cells washed of surface layer proteins, but only if the SLH modules were present. This indicates that Xyn10C on one hand binds to the surface layer with virtue of its SLH modules and on the other to cellulose with its CBMs. The binding of the cell to its substrate has a great advantage for the bacterium: a lower amount of exo-enzymes to be produced for saturation of the substrate and a higher concentration of hydrolysis products near the cell surface combined with diminished loss by diffusion.

Xyn11A, Xyn10C, Arf51B, and Bx13B are major components in the C. stercorarium culture supernatant. These enzymes, and not the products of other genes also isolated from the genome, are sufficient to hydrolyze arabinoxylan effectively and completely to its monomers (Adelsberger et al., submitted). Arf51B releases /α/-arabinofuranosyl residues from the intact arabinoxylan, as well as from the oligosaccharides produced by the action of the endo-xylanases Xyn11A and/or Xyn10C. The resulting xylo-oligosaccharides are degraded to xylose by Bx13B. The point of attack of the two endo-xylanases seems to be different: the produced oligosaccharides differ as well in size as in
the digestibility by β-xylosidase Bx13B. The structure of single oligosaccharides could be resolved by sequential digestion with Bx13B and Arf51B. The complete degradation of arabinoxylan is schematically shown in Fig. 7.

In analogy to arabinoxylan, it could be supposed that glucuronoxylan was hydrolyzed by extracellular enzymes. However, the only glucuronosidase activity detected in *C. stercorarium* cultures was located intracellularly. Consequently, it has to be surmised that glucuronoxylan is degraded by the extracellular xylanases to soluble, charged oligosaccharides that are—in contrast to the noncharged arabinoxyloligosaccharides—transported through the cell barrier and further degraded there.

With *in vitro* assays using model substrates it has not yet been possible to clarify whether the two major extracellular xylanases Xyn11A and Xyn10C act synergistically on natural xylan. The presence
of different binding modules suggests a role in the digestion of different topologies in naturally occurring plant cell walls.

It should be mentioned that the extracellular cellulose of C. thermocellum contains a number of xylanase components (see Table IV). Five genes have been cloned and biochemically characterized, another two potential genes were identified in the genomic sequence. The breakdown may be enhanced by the activity of esterases: one esterase connected with Xyn11A (the NodB module), two with Xyn11Y and Xyn11Z (Blum et al., 2000), and two others identified in the genomic sequence (#38 and 39 in Table IV). The oligosaccharides are broken down by $\alpha$-arabinofuranosidases and $\beta$-xylosidases, seven genes of which are present in the genome, but none has been characterized yet. In addition, potential genes for two mannanas (manA already characterized), an arabinogalactanase and a xyloglucanase are present, as well as genes for a pectate lyase, a polygalacturonase, and a rhamnogalacturonan lyase for the breakdown of pectin. It is not known if these genes code for active enzymes and if they are expressed. The presence of a chitinase in the cellulose has already been shown; the gene chiA was identified (Zverlov et al., 2002a).

V. Concluding Remarks

Clostridia are very important organisms for modern biotechnology. In the past, they were used for the production of acetone and butanol, flax retting, and indigo dyeing. In this chapter we have highlighted the features of some of the extracellular enzymes produced by these bacteria to degrade/hydrolyze biopolymers such as starch or cellulose. Some of theses enzymes or enzyme systems are unique among microorganisms. The enormous potential of clostridia as producers of industrially important enzymes is obvious. In the last decade, significant progress has been achieved in the understanding of the structure-function relationships of the clostridial type of enzymes and their modules. The technique of enzyme modification by adding modules or the increase in enzyme activity by complex formation will be a great stimulus for modern enzymology. In addition, genetic tools for clostridia have been developed for custom engineering of new production strains. Thus it seems to be possible now to engineer an enzyme with optimal features for a given purpose or even to create a special Clostridium species that is able to convert cheap, renewable biomass into desired valuable products. Thus it is not utopian to believe that, for example, C. acetobutylicum one day will transform cellulosic wastes directly to solvents.
We thank W. L. Staudenbauer for encouragement and many stimulating discussions and Karin Bronnenmeier for providing unpublished results. Experimental work carried out in the authors’ laboratories was supported by the Deutsche Forschungsgemeinschaft, the former Bundesministerium für Forschung und Technologie, and the Fonds der Chemischen Industrie. This study was also supported by a grant of the AvHumboldt Foundation to V.V.Z.

REFERENCES


of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose. Microbiology 148, 247–255.


Zverlov et al., in preparation. Genome sequence available at URL http://genome.ornl.gov/microbial/cthe/