ResearchGate

See discussions, stats, and author profiles for this publication at:
https://www.researchgate.net/publication/226522753

Lignocellulose-Decomposing Bacteria and Their Enzyme Systems

CHAPTER · DECEMBER 2005

DOI: 10.1007/0-387-30742-7_19

CITATIONS
41

READS
219

3 AUTHORS:

Bayer A. Edward
Weizmann Institute of Science
411 PUBLICATIONS **15,515** CITATIONS
SEE PROFILE

Yuval Shoham
Technion - Israel Institute of Technology
175 PUBLICATIONS **6,641** CITATIONS
SEE PROFILE

Raphael Lamed
Tel Aviv University
286 PUBLICATIONS **10,235** CITATIONS
SEE PROFILE

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Bayer A. Edward
Retrieved on: 22 January 2016

Prokaryotes (2006) 2:578–617
DOI: 10.1007/0-387-30742-7_19

CHAPTER 1.19

Cellulose-Decomposing Bacteria and Their Enzyme Systems

EDWARD A. BAYER, YUVAL SHOHAM AND RAPHAEL LAMED

Introduction

From an anthropocentric point of view, for millennia, human culture has been intricately involved with cellulose, the major component of the plant cell wall. The development of the wood, paper and textile industries has served to incorporate cellulosic materials into the fabric of our society. Within the past century, however, cellulosic wastes, derived mainly from the same industries, have also become a major source of environmental pollution. This chapter will concentrate mainly on cellulose and the cellulolytic bacteria, in view of their importance to mankind and world ecology. Nevertheless, the true substrate of these bacteria—i.e., the complement of plant cell wall polysaccharides in general—is much more complex than cellulose alone. Likewise, the complement of enzymes—both the cellulolytic and the non-cellulolytic glycosyl hydrolases—are produced concurrently in these bacteria for the purpose of efficient synergistic degradation of the complete substrate composite as it appears in nature. Consequently, when we discuss the cellulose-decomposing bacteria and their enzyme systems, we cannot ignore the related noncellulolytic enzymes, and these will also be treated, albeit secondarily, in the present chapter.

It should also be noted that this chapter of the *The Prokaryotes* is a sequel to the previous chapter of the same title (authored by M.P. Coughlan and F. Mayer) from the second edition of this treatise (Coughlan and Mayer, 1992). The reader is cordially invited to consult the earlier chapter (to be considered as Part A) as an excellent complement to our own (Part B).

The plant cell wall consists of an intricate mixture of polysaccharides (Carpita and Gibeaut, 1993); cellulose, hemicellulose and lignin are its major constituents. These polymers are of a very robust nature. They both equip the plant with a stable structural framework and protect the plant cell from the perils of its environment. Despite its recalcitrant nature, in the guise of dead or dying plant matter, the polysaccharides of the plant cell wall provide an exceptional source of carbon and energy, and a multitude of different microorganisms has evolved which are capable of degrading plant cell wall polysaccharides.

In any given ecosystem, the polysaccharide-degrading microbes are not alone, but rely on the complementary contribution of other bacterial and/or fungal species (Bayer and Lamed, 1992; Bayer et al., 1994; Ljungdahl and Eriksson, 1985). The polymer-degrading strains play a primary and crucial role in the ecosystem by converting the plant cell wall polysaccharides to the respective simple sugars and other degradation products (Fig. 1). They are assisted by satellite microbes, which cleanse the microenvironment from the breakdown products, producing, in the final analysis methane and carbon dioxide.

In a given polysaccharide-degrading microorganism, the enzymes that catalyze the degradation may occur either in the free state and/or in discrete complexes with other similar types of enzymes. The latter are called "cellulosomes." Both the free enzymes and cellulosomal components are usually modular proteins, which contain a multiplicity of functional domains. The "free" enzymes comprise a single polypeptide chain, which contains a catalytic domain usually connected to a cellulose-binding domain or CBD. Cellulosomes are exocellular macromolecular machines, designed for efficient degradation of cellulose and associated plant cell wall polysaccharides (Bayer et al., 1998). In contrast to the free enzymes, the cellulosome complex is composed of a collection of subunits, each of which comprises a set of interacting functional modules. Thus, one type of cellulosomal module, the CBD, is selective for binding to the substrate. Another family of modules, the catalytic domains, is specialized for the hydrolysis of the cellulose chains. Yet another complementary pair of domains—the cohesins and dockerins—serves to integrate the enzymatic subunits into the complex and the complex, in turn, into the cell surface. Multiple copies of the cohesins form an integrating subunit called "scaffoldin" to which the dockerin-containing

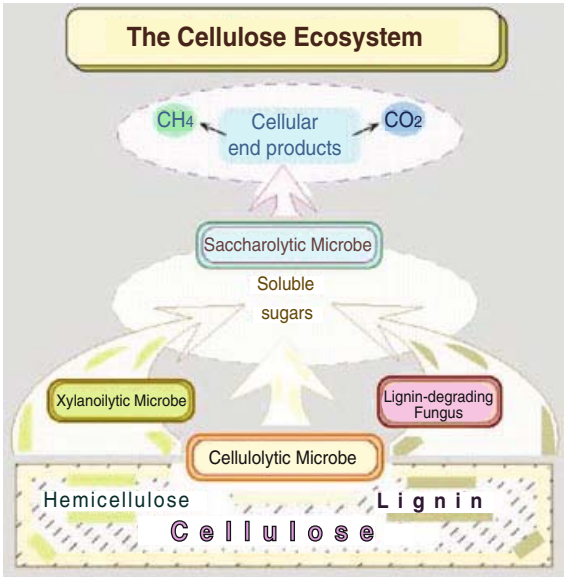

Fig. 1. Simplified schematic description of a typical ecosystem comprising degrading plant matter. Cellulolytic, xylanolytic and ligninolytic microbes combine to decompose the major polysaccharide components to soluble sugars. "Satellite" microorganisms assimilate the excess sugars and other cellular end products, which are ultimately converted to methane and carbon dioxide.

enzymes are attached. This "Lego™ike arrangement of the modular subunits generates an intricate multicomponent complex, the enzymes of which are bound en bloc to the insoluble substrate and act synergistically towards its complete digestion.

Inherent to the study of cellulases and related enzymes is their potential industrial application —particularly towards conversion of cellulosic biomass. For reviews on the potential uses of these enzymes, the reader is referred to appropriate reviews on the subject (Bhat, 2000; Himmel et al., 1999; Lynd et al., 1991).

Plant Cell Wall Polysaccharides

Plant cells produce a composite matrix of hardy and durable polysaccharides on the outer surface of the plasma membranes, called "the cell wall" (Carpita and Gibeaut, 1993). The cell wall confers a protective coating to the plant cell, providing structure, turgidity and durability, which renders the cell resistant to the outer elements, including mechanical, chemical and microbial assault. Different types of plant cell tissues exhibit different ratios of the three major types of cell wall component; on the average, the cell wall contains roughly 40% cellulose, 30% hemicellulose and 20% lignin, but the exact composition of an individual type of plant varies greatly. The first two polymers are indeed polysaccharides. On the other hand, lignin is a heterogeneous, high-molecular-weight hydrophobic polymer, which consists of nonrepeating aromatic monomers connected via phenoxy linkages (Higuchi, 1990; Lewis and Yamamoto, 1990). Unlike cellulose and hemicellulose, which are degraded aerobically or anerobically, lignin degradation requires oxygen and is limited to filamentous prokaryotes (e.g., the Actinomycetes Streptomyces viridans) and fungi (e.g., *Phanerochaete chrysosporium*, *Bejerkendera adusta* and *Pleurotus ostreatus*), which produce a complicated set of enzymes that hydrolyze the polymer. In fact, the recalcitrant lignin interferes severely with the access of enzymes to the cellulose component, and is rate limiting for anaerobic degradation of cellulose. In any case, the lignin component must be degraded or removed, before efficient degradation of cellulose can take place. Nevertheless, considering lignin is not a polysaccharide, it will not be discussed further in this chapter.

Cellulose

Cellulose is the major constituent of plant matter and thus represents the most abundant organic polymer on Earth. Cellulose is a remarkably stable homopolymer, consisting of a linear (unbranched) polymer of β-1,4-linked glucose units. Chemically, the repeating unit is simply glucose, but structurally, the repeating unit is the disaccharide cellobiose, i.e., 4-*O*-(β-D-glucopyranosyl)-D-glucopyranose, inasmuch as each glucose residue is rotated 180° relative to its neighbor (Fig. 2). The individual cellulose chains contain from about 100 to more than 10,000 glucose units, packed tightly in parallel fashion into microfibrils by extensive inter- and intrachain hydrogen bonding interactions, which account for the rigid structural stability of cellulose. The microfibrils exhibit variable amounts of crystalline and amorphous components, again depending on the degree of polymerization, the extent of hydrogen bonding and, ultimately on the source of the cellulose. The microfibrils themselves are further assembled into plant cell walls, the tunic of some sea animals, pellicles from bacterial origin, etc. Highly crystalline forms of cellulose include cotton, bacterial cellulose (from *Acetobacter xylinum*) and the cellulose from the algae, *Valonia ventricosa*, which exhibit crystallinity levels of about 45%, 75% and 95%, respectively. The following reviews are available for more information on the structure of cellulose (Atalla, 1999; Atalla and VanderHart, 1984; Chanzy, 1990; O'Sullivan, 1997).

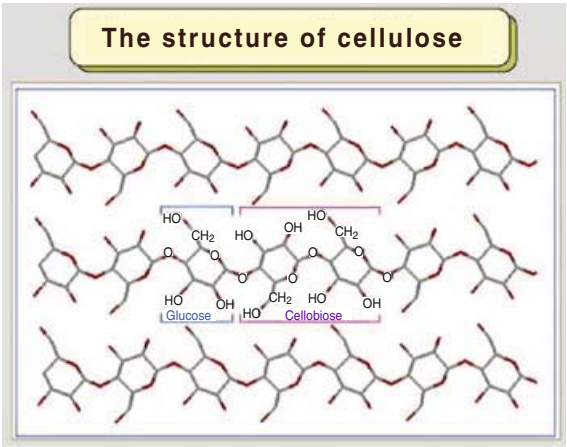

Fig. 2. Structure of cellulose. Three parallel chains that form the 0, 1, 0 face are shown, and a glucose moiety and repeating cellobiose unit are indicated. The model was built by Dr. José Tormo, based on early crystallographic data. The diagram was drawn using RasMol 2.6.

Hemicellulose

Hemicelluloses are relatively low-molecular-weight, branched heteropolysaccharides associated with both cellulose and lignin and together build the plant cell wall material (Puls and Schuseil, 1993; Timell, 1967). The main backbone of hemicellulose is usually made of one or two sugars, which determines their classification. For example, the main backbone of xylan is composed of 1,4-linked-β-D-xylopyranose units. Similarly, the backbone of galactoglucomannans is made of linear 1,4-linked β-D-glucopyranose and β-D-mannopyranose units with a-1,6-linked galactose residues. Other common hemicelluloses include arabinogalactan, lichenins (mixed 1,3-1,4-linked β-D-glucans) and glucomannan. Most hemicellulases are based on a 1,4-β-linkage and the main backbone is branched, whereas the individual sugars may be acetylated or methylated. For example, the linear xylan backbone is highly substituted with a variety of saccharide and nonsaccharide components (Fig. 3). In the

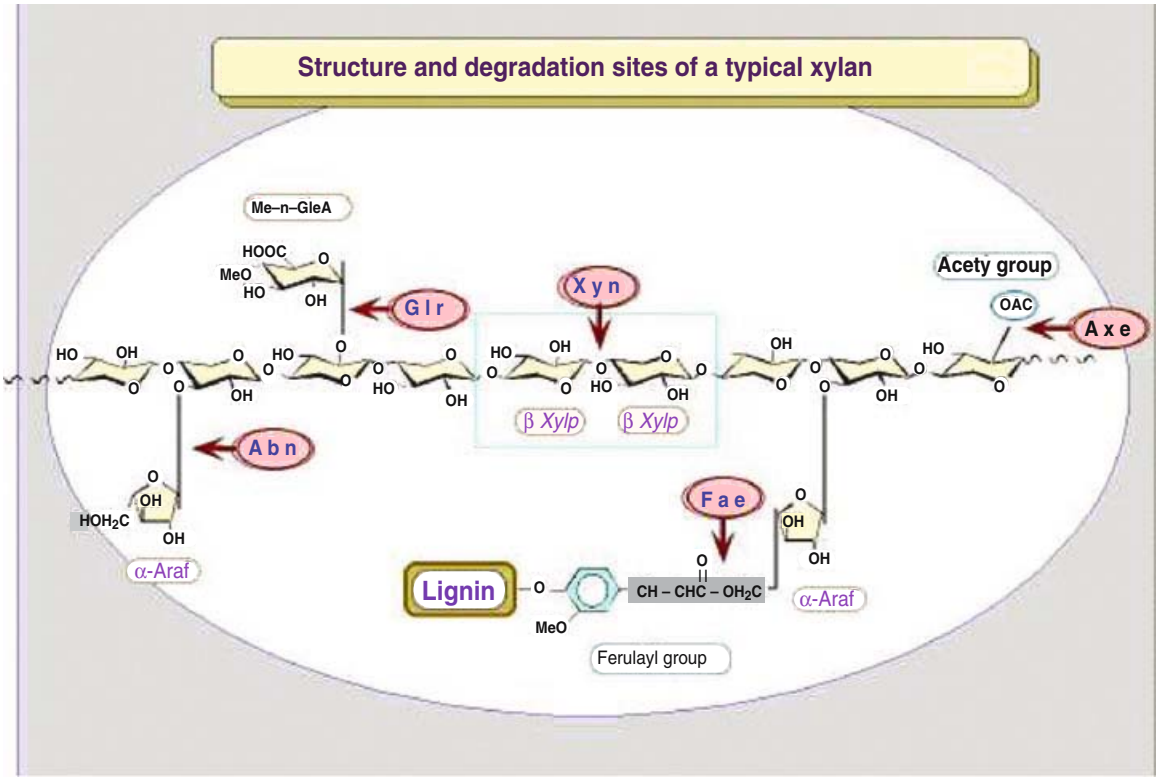

Fig. 3. Composition of a typical xylan component of hemicellulose. The xylobiose unit (β-Xylp-β-Xylp) is indicated by the blue-sided box, as are major substituents: MeaGlcA, methylglucuronic acid; αAra*f*, arabinofuranosyl; OAc, acetyl group. A presumed lignin attachment site to a feruloyl substituent of xylan is also illustrated. Sites of cleavage by selected hemicellulases and carbohydrate esterases are also shown: Xyn, xylanase; Abn, arabinofuranosidase; Glr, glucuronidase; Axe, acetyl xylan esterase; Fae, ferulic acid esterase.

plant cell wall, xylan is closely associated with other wall components. The 4-*O*-methyl-α-D-glucuronic acid residues can be ester-linked to the hydroxyl groups of lignin, providing cross-links between the cell walls and lignin (Das et al., 1984). Similarly, feruloyl substituents serve as crosslinking sites to either lignin or other xylan molecules. Thus, the chemical complexity of xylan is in direct contrast to the chemical simplicity of cellulose. Likewise, the structural diversity of the xylans is in contrast to the structural integrity of the cellulose microfibril. Consequently, unlike the crystalline-like character of cellulose, the hemicellulose component adopts a gel-like consistency, providing an amorphous matrix in which the rigid crystalline cellulose microfibrils are embedded.

Cellulose-Degrading Bacteria

The cellulolytic microbes occupy a broad range of habitats. Some are free living and rid the environment of plant polysaccharides by converting them to the simple sugars, which they assimilate. Others are linked closely with cellulolytic animals, residing in the digestive tracts of ruminants and other grazers or in the guts of wood-degrading termites and worms (Haigler and Weimer, 1991). Cellulose-based ecosystems include soils, swamps, marshes, rivers, lakes and seawater sediments, rotting grasses, leaves and wood, cotton bales, sewage sludge, silage, compost heaps, muds and decaying vegetable matter in hot and volcanic springs, acid springs, and alkaline springs (Ljungdahl and Eriksson, 1985; Stutzenberger, 1990).

The cellulolytic microorganisms include protozoa, fungi and bacteria and are ubiquitous in nature. The cellulose-decomposing bacteria include aerobic, anaerobic, mesophilic and thermophilic strains, inhabiting a great variety of environments, including the most extreme vis-à vis temperature, pressure and pH. Cellulolytic bacteria also have been found in the gut of wood-eating worms, termites and vertebrate herbivores, all of which exploit anaerobic symbionts for the digestion of wood and fodder.

In nature, many cellulolytic species exist in symbiotic relationships with secondary microorganisms (Ljungdahl and Eriksson, 1985). The primary microorganisms degrade cellulose directly to cellobiose and glucose. Only part of the breakdown products is assimilated by the polymer degrading strain(s), and the rest is utilized by the satellite microorganisms. Removal of the excess of sugars promotes further cellulose degradation by the primary species because cellobiose-induced inhibition of cellulase action and repression of cellulase synthesis are precluded.

Modern interest in cellulolytic microorganisms was spawned by the decay of cotton fabric in army tents and military clothing in the South Pacific jungles during World War II. The basic research program that resulted from this military problem led to the establishment of the United States Army Natick Laboratories (Reese, 1976). The resultant research led to the discovery that the causative agent for the costly problem was a cellulolytic fungi, *Trichoderma viride* (subsequently renamed *Trichoderma reesei*). Subsequent research, originally from the Natick Laboratories and later spreading to other research institutes and universities, led to the identification and classification of thousands of different strains of cellulolytic fungi and bacteria. Many of the major types of cellulolytic bacteria have been listed in Part A of the second edition of *The Prokaryotes* (Coughlan and Mayer, 1992). Since the latter publication, the major emphasis in the area has not concentrated on the discovery or description of new cellulolytic strains. Rather, research in the area during the past decade has centered on characterizing the enzymes and enzyme systems from selected bacteria that degrade cellulose in particular and plant cell wall polysaccharides in general.

Enzymes That Degrade Plant Cell Wall Polysaccharides

The chemical and structural intricacy of plant cell wall polysaccharides is matched by the diversity and complexity of the enzymes that degrade them. The cellulases and hemicellulases are family members of the broad group of glycosyl hydrolases, which catalyze the hydrolysis of oligosaccharides and polysaccharides in general (Gilbert and Hazlewood, 1993; Kuhad et al., 1997; Ohmiya et al., 1997; Schülein, 1997; Tomme et al., 1995a; Viikari and Teeri, 1997; Warren, 1996; Wilson and Irwin, 1999).

Historically, the type of substrate and manner in which a given enzyme interacted with its substrate were decisive in the classification of the glycosidases, as established first by the Enzyme Commission (EC) and later by the Nomenclature Committee of the International Union of Biochemistry (IUB). Enzymes were usually named and grouped according to the reactions they catalyzed. Thus, cellulases, xylanases, mannanases and chitinases were grouped a priori in different categories. Moreover, enzymes that cleave polysaccharide substrates in the middle of the chain ("endo"-acting enzymes) versus those which clip at the chain ends ("exo"-acting enzymes) were also placed in different groups.

For example, in the case of cellulases, the endoglucanases were grouped in EC 3.2.1.4, whereas the exoglucanases (i.e., cellobiohydrolases) were classified as EC 3.2.1.91.

The historical division of enzymes is inappropriate for classification of the cellulases and other glycosyl hydrolases. Like other enzymes (e.g., proteases, etc.), previous classification systems of the glycosyl hydrolases centered on the types of substrates and the bonds cleaved by a given enzyme. The problem with the glycosyl hydrolases is that the polysaccharide substrates and particularly the bonds they cleave are all quite similar, and classification of the different types of enzymes according to conventional criteria often misses the mark. Consequently, alternative approaches were pursued. The recent trend is to classify the different glycosyl hydrolases into groups based on common structural fold and mechanistic themes (Davies and Henrissat, 1995; Henrissat, 1991; Henrissat and Bairoch, 1996; Henrissat and Davies, 1997; Henrissat et al., 1998). A comprehensive website that provides a catalog of the different glycosyl hydrolase families is now available (Coutinho and Henrissat, 1999a; Coutinho and Henrissat, 1999c; [Carbohydrate-Active Enzymes server (afmb.cnrs-mrs.frl)]. The website also provides excellent introductory explanatory material, and the interested reader is encouraged to use this site extensively.

It is interesting that the distinction between endo- and exo-acting enzymes is also reflected by the architecture of the respective class of active site, even within the same family of enzymes (Fig. 4). The endoglucanases, for example, are commonly characterized by a groove or cleft, into which any part of a linear cellulose chain can fit. On the other hand, the exoglucanases bear tunnel-like active sites, which can only accept a substrate chain via its terminus. The exo-acting enzyme apparently threads the cellulose chain through the tunnel, wherein successive units (e.g., cellobiose) would be cleaved in a sequential manner. The sequential hydrolysis of a cellulose chain is a relatively new notion of growing importance, which has earned the term "processivity" (Davies and Henrissat, 1995), and processive enzymes are considered to be key components which contribute to the overall efficiency of a given cellulase system.

Though instructive, there is growing dissatisfaction with the endo/exo terminology. As our understanding of the nature of catalysis by these enzymes progresses, it has become clear that some enzymes are capable of both endo- and exo-action (Johnson et al., 1996; Morag et al.,

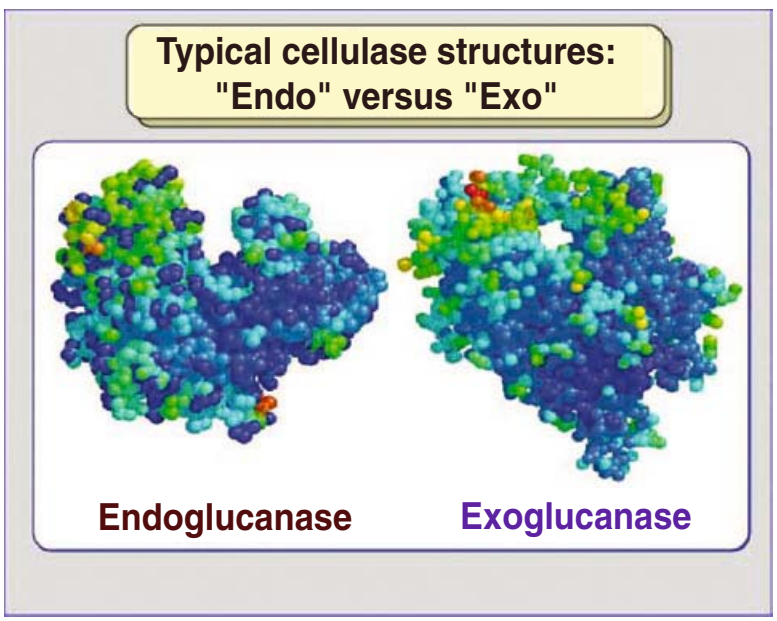

Fig. 4. Structures of a typical endoglucanase and exoglucanase. In each case, the structure is viewed from a perspective, which demonstrates the comparative architecture of the respective active site. Despite the sequence similarity of both enzymes and their classification as family-6 glycosyl hydrolases, their respective active-site architecture is different. The endoglucanase (endoglucanase E2 from the bacterium, *Thermomonospora fusca*, PDB code 1TML) is characterized by a deep cleft to accommodate the cellulose chain at any point along its length, whereas the active site of the exoglucanase (cellobiohydrolase CBHI from the cellulolytic fungus, *Trichoderma reesei*, PDB code 1CEL) bears an extended loop that forms a tunnel, through which one of the termini of a cellulose chain can be threaded. The ribbon diagrams, showing the secondary structures (α-helices and β-strands) of the two enzymes, were drawn using RasMol 2.6.

1991; Reverbel-Leroy et al., 1997; Sakon et al., 1997). Moreover, some glycosyl hydrolase families include both endo- and exoenzymes, again indicating that the mode of cleavage can be independent of sequence homology and structural fold. In this context, relatively minor changes in the lengths of relevant loops in the general proximity of the active site, may dictate the endo- or exo-mode of action without significant differences in the overall fold.

Owing to subtle but diverse chemical and structural aspects of the substrates involved, plant cell wall degrading enzymes do not follow the same rules as common enzyme standards, such as simple proteases, DNAse, RNAse and lysozyme. In fact, the cellulases and hemicellulases are usually very large enzymes, whose molecular masses often exceed those of proteases by factors of 2–5 and more. Their polypeptide chains partition into a series of functional modules and linker segments (frequently glycosylated), which together determine their overall activity characteristics and interaction with their substrates and/or with other components of the cellulolytic and hemicellulolytic system.

Cellulases

The cellulases include the large number of endo- and exoglucanases which hydrolyze β-1,4-glucosidic bonds within the chains that comprise the cellulose polymer (Béguin and Aubert, 1994; Haigler and Weimer, 1991; Tomme et al., 1995b). Thus, in principle, the degradation of cellulose requires the cleavage of a single type of bond. Nevertheless, in practice, we find that cellulolytic microorganisms produce a variety of complementary cellulases of different specificities from many different families.

It may seem somewhat surprising that the combined effect of so many different enzymes are required to degrade such a chemically simplistic substrate. This complexity reflects the difficulties an enzyme system encounters upon degrading such a highly crystalline substrate as cellulose. As described in the previous section, cellulases that degrade the cellulose chain can be either "endo-acting" or "exo-acting." Moreover, the degradation of crystalline cellulose should be viewed three-dimensionally and in situ, where the cellulose chains are packed within the microcrystal, thus generating the remarkably stable physical properties of the crystalline substrate. The enzymes have to bind to the cellulose surface, localize and isolate suitable chains, destined for degradation. It would seem logical that amorphous regions or defects in the crystalline portions of the substrate would be favorable sites for initiation of the process. The structural as opposed to chemical heterogeneity of the substrate dictates the synergistic action of a complex set of complementary enzymes towards its complete digestion.

Various models have been suggested to account for the observed synergy between and among two or more different types of cellulases. For example, an endo-acting enzyme can produce new chain ends in the internal portion of a polysaccharide backbone, and the two newly exposed chains would then be available for action of exo-acting enzymes. In addition, two different types of exoglucanases may exhibit different specificities by acting on a cellulose chain from opposite ends (i.e., the reducing versus the nonreducing end of the polymer). Likewise, an endoglucanase may be selective for only one of the two sterically distinct glucosidic bonds on the cellulosic surface. In addition, some cellulases may display high levels of activity at the beginning of the degradative process, i.e., on the highly crystalline material, whereas others would be selective for newly exposed, partially degraded chains, otherwise embedded within the crystal. Still others would show very high levels of activity after the degradative process has advanced, and cellulose chains that have been freed of the crystalline setting would then be hydrolyzed quite rapidly. A collection of various enzymes, which exhibit complementary specificities and modes of action, would account for the observed synergistic action of the complete cellulase "system" in digesting the cellulosic substrate.

In addition to endo- and exoglucanases, included in the overall group of cellulases are the β-glucosidases (EC 3.2.1.21), which hydrolyzes terminal, nonreducing β-D-glucose residues from cello-oligodextrins. In particular, this type of enzyme cleaves cellobiose—the major end product of cellulase digestion—to generate two molecules of glucose. Some β-glucosidases are specific for cellobiose whereas others show broad specificity for other β-D-glycosides, e.g., xylobiose. Often, the β-glucosidases are associated with the microbial cell surface and hydrolyze cellobiose to glucose before, during or after the transport process.

Hemicellulases

Strictly speaking, hemicellulases are not the precise subject of this chapter, since they do not directly sever the β-1,4-glucosidic bond of cellulose. Nevertheless, in nature, they are essential to the bacterial degradation of insoluble cellulose because the natural bacterial substrate—the plant cell wall—comprises an architecturally

cogent composite of cellulose and hemicellulose. In natural systems, the two types of polysaccharides cannot be easily separated, and microbial systems have to deal simultaneously with both. The xylan component is particularly of interest for several reasons: 1) xylan is a major hemicellulosic component of the plant cell wall, 2) the xylanases are well defined enzymes, closely associated with the cellulase and 3) the repeating units (both xylose and xylobiose) bear striking structural resemblance to their cellulosic counterparts (i.e., glucose and cellobiose).

In contrast to cellulose degradation, the degradation of the hemicelluloses imposes a somewhat different challenge, since this group of polysaccharides includes widely different types of sugars or non-sugar constituents with different types of bonds. Thus, the complete degradation of hemicellulose requires the action of different types of enzymes. These enzymes, the hemicellulases, can differ in the chemical bond they cleave, or, as in the case of the cellulases, they may cleave a similar type of bond but with different substrate- or product specificity (Biely, 1985; Coughlan and Hazlewood, 1993; Eriksson et al., 1990; Gilbert and Hazlewood, 1993).

Hemicellulases can be divided into two main types, those that cleave the mainchain backbone, i.e., xylanases or mannanases, and those that degrade sidechain substituents or short end products, such as arabinofuranosidase, glucuronidase, acetyl esterases and xylosidase. Like the cellulases, hemicellulases can be of the endo- or exo-types. A schematic view of the types of bonds that would be hydrolyzed by different types of hemicellulases is presented in Fig. 3.

XYLAN-DEGRADING ENZYMES. The xylanases are by far the most characterized and studied of the hemicellulases and involve the cleavage of a major mainchain backbone. Endoxylanases (1,4-β-D-xylan xylanhydrolase, EC 3.2.1.8) hydrolyze the 1,4-β-D-xylopyranosyl linkage of xylans, such as D-glucurono-D-xylans and L-arabino-D-xylan. These single-subunit enzymes from both fungi and bacteria exhibit a broad range of physiochemical properties, whereby two main classes have been described: alkaline proteins of low M_r (<30,000) and acidic proteins of high M_r. This general classification scheme correlates with their assignment into glycosyl hydrolase families 10 and 11, whereby the former represents the high M_r xylanases and the latter coincides with the low M_r enzymes. The two families also differ in their catalytic properties, such that the family 10 enzymes seem to display a greater versatility towards the substrate than that observed for those of family 11, and are thus typically able to hydrolyze highly substituted xylan more efficiently. The family 10 xylanases exhibit a $(\beta/\alpha)8$ topology whereas those from family 11 form a β-jelly roll fold. Both families show a retaining catalytic mechanism of hydrolysis.

MANNAN-DEGRADING ENZYMES. Glucomannans and galactoglucomannans are branched heteropolysaccharides found in hardwood and softwood. The degradation of these polymers again involve many hydrolytic enzymes, including endo-1,4-β-mannanase (EC 3.2.1.78), β-mannosidase (EC 3.2.1.25), β-glucosidase (EC 3.2.1.21), and α-galactosidase (EC 3.2.1.22). 1,4-β-D-Mannanases hydrolyze mainchain linkages of D-mannans and D-galacto-D-mannans. These enzymes, both of the endo- or exo-types, are produced in various microorganisms, including *Bacillus subtilis*, *Aspergillus niger* and intestinal and rumen bacteria and commonly occur in families 5 and 26.

LICHENIN-DEGRADING ENZYMES. Lichenase (1,3-1,4-β-D-glucan 4-glucanohydrolase, EC 3.2.1.73) is a mixed linkage β-glucanase, which cleaves the β-1,4 linkages adjacent to the β-1,3 bonds of the lichenin substrate. According to [{afmb.cnrs-mrs.fr/~pedro/CAZY/db.html}{modern structure-based classification, lichenases can be members of families 8, 16 or 17.

β-D-XYLOSIDASES. The 1,4-β-D-xylosidases (1,4-β-D-xylan xylohydrolase, EC 3.2.1.37) hydrolyze xylo-oligosaccharides (i.e., xylan breakdown products and mainly xylobiose) to xylose. These enzymes are either intracellular or extracellular components and are closely associated with hemicellulolytic activities. Monomeric, dimeric and tetrameric xylosidases have been found with Mrs of 26,000 to 360,000. Many of the xylosidases act on a variety of substrates. For example, *Aspergillus niger* produces an enzyme classified as a β-xylosidase that can hydrolyze β-galactosides, β-glucosides and α-arabinosides, in addition to β-xylosides.

SIDECHAIN-DEGRADING ENZYMES. α-D-Glucuronidases (EC 3.2.1.39) catalyze the cleavage of the α-1,2 glucosidic bond of 4-*O*-methyl-α-D-glucuronic acid side chain. This bond has a stabilizing effect on the neighboring xylosidic bonds of the main chain. Several α-glucuronidase genes have recently been cloned and sequenced and usually occupy family 67.

α-L-Arabinofuranosidases (α-L-arabinofuranoside arabinofuranohydrolase, EC 3.2.1.55) is another important enzyme that cleaves nonreducing terminal α-L-arabinofuranosidic linkages in arabinoxylan, L-arabinan, and other

L-arabinose containing polysaccharides. These enzymes are found either in the cell-associated or extracellular form and can be members of families 43, 51 or 62.

1,4-β-Mannosidases hydrolyze 1,4-linked β-D-mannosyl groups from the nonreducing end. These enzymes (similar to β-xylosidases) hydrolyze mainly the end products of the mannanases, i.e. mannobiose and mannotriose.

Carbohydrate Esterases

The side chain substituents of xylan are composed not only of sugars but also of acidic residues, such as acetic, ferulic (4-hydroxy-3-methoxycinnamic) or *p*-coumaric (4-hydroxycinnamic) acids. Carbohydrate esterases that cleave these residues (see Fig. 3) are found in enzyme preparations from both hemicellulolytic and cellulolytic cultures (Borneman et al., 1993). Such enzymes sometimes represent separate modules, separated by linker segments from other cellulolytic or hemicellulolytic catalytic modules in the same polypeptide chain. Like the glycosyl hydrolases, the carbohydrate esters are currently classified according to sequence homology and common structural fold.

Cellulases and Hemicellulases are Modular Enzymes

The initial contribution of biochemical methods for determining the characteristics of a given cellulase was extended immeasurably by the contribution of molecular biology and bioinformatics. By comparing the sequences of the cellulases and related enzymes, an entirely new view of these enzymes emerged.

Cellulases and hemicellulases are composed of a series of separate modules. This fact explains the very large size of some of these enzymes and gives us some insight into their complex mode of action. Each module or domain comprises a consecutive portion of the polypeptide chain and forms an independently folding, structurally and functionally distinct unit (Coutinho and Henrissat, 1999; Gilkes et al., 1991; Teeri et al., 1992). Each enzyme contains at least one catalytic module, which catalyzes the actual hydrolysis of the glycosidic bond and provides the basis for classification of the simple enzymes (i.e., those containing a single catalytic module). Other accessory or "helper" domains assist or modify the primary hydrolytic action of the enzyme, thus modulating the overall properties of the enzyme. Some of the different themes illustrating the modular compositions of the cellulases and related enzymes are illustrated in Fig. 5.

The Catalytic Modules—Families of Enzymes

The definitive component of a given enzyme is the catalytic domain. Former EC-based classification schemes according to substrate specificity are now considered somewhat obsolete because they fail to take into account the structural features of the enzymes themselves. The catalytic domains of glycosyl hydrolases are presently categorized into families according to amino acid sequence homology (Coutinho and Henrissat, 1999; Henrissat, 1991; Henrissat and Bairoch, 1996; Henrissat and Davies, 1997; Henrissat et al., 1998). For more information, see the Carbohydrate-Active Enzymes (CAZy), designed and maintained by Pedro Coutinho and Bernard Henrissat.

The enzymes of a given glycosyl hydrolase family display the same topology, and the positions of the catalytic residues are conserved with respect to the common fold. In recent years, X-ray crystallography has provided a general overview of the structural themes of the glycosyl hydrolases and their interaction with their intriguing set of substrates (Bayer et al., 1998; Davies and Henrissat, 1995; Henrissat and Davies, 1997).

The mechanism of cellulose and hemicellulose hydrolysis occurs via general acid catalysis and is accompanied by either an overall retention or an inversion of the configuration of the anomeric carbon (Davies and Henrissat, 1995; McCarter and Withers, 1994; White and Rose, 1997; Withers, 2001). In both cases, cleavage is catalyzed primarily by two active-site carboxyl groups. One of these acts as a proton donor and the other as a nucleophile or base. Retaining enzymes function via a double-displacement mechanism, by which a transient covalent enzyme-substrate intermediate is formed (Fig. 6A). In contrast, inverting enzymes employ a single-step mechanism as shown schematically in Fig. 6B. The distance between the acid catalyst and the base represents the major structural difference between the two mechanisms. In retaining enzymes, the distance between the two catalytic residues is about 5.5 Å, whereas in inverting enzymes the distance is about 10 Å. In the inverting enzymes, additional space is provided for a water molecule, involved directly in the hydrolysis, and the resultant product exhibits a stereochemistry opposite to that of the substrate. In all cases, the mechanism of hydrolysis is conserved within a given glycosyl hydrolase

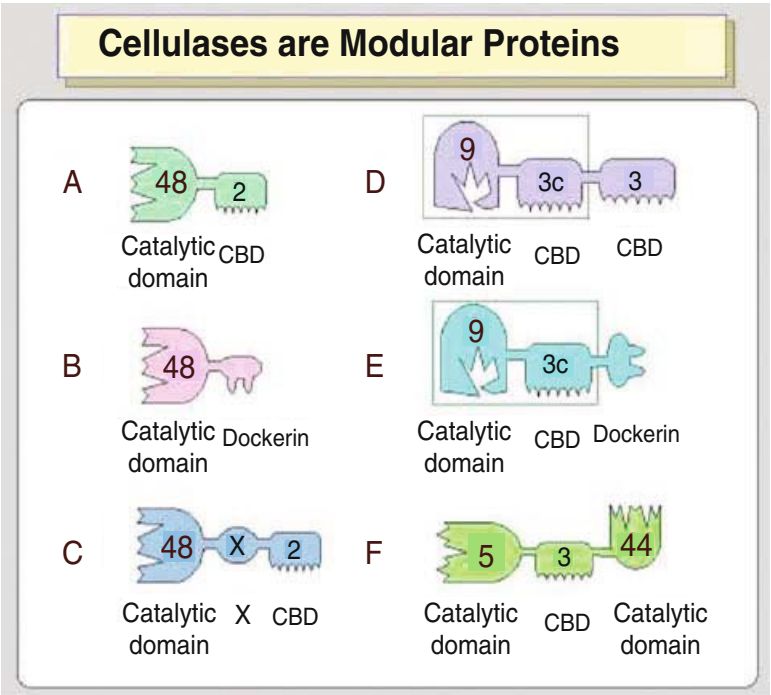

Fig. 5. Scheme illustrating the diversity of the modular architecture of cellulases and other glycosyl hydrolases. The different modules are grouped into families according to conserved sequences as shown here symbolically. A. One of the most common types of cellulases consists of a catalytic module or domain, flanked by a cellulose-binding domain (CBD) at its *N*- or *C*-terminus. This particular enzyme shown in "A" comprises a catalytic domain from family 48 and a family-2 CBD. B. Cellulosomal enzymes are characterized by a "dockerin domain" attached to a catalytic domain. In this case, the same type of enzyme as in "A," carrying a family-48 catalytic module, harbors a dockerin domain instead of a CBD. C. Many cellulases contain "X domains," i.e., domains of unknown (as yet undefined) function. D. Some enzymes have more than one CBD or other type of carbohydrate-binding module (CBM). Often, one CBD, such as the family-3 CBD shown here, serves to bind the cellulase strongly to the flat surface of the insoluble substrate, whereas the other one (the family-3c CBD) acts in concert with the catalytic module by binding transiently to a single cellulose or hemicellulose chain. E. Some cellulosomal cellulases have a CBD or CBM together with a dockerin in the same polypeptide chain. F. Some cellulases have more than one type of catalytic module, such as the family-5 and family-44 modules shown here, and the two probably work in concerted fashion to degrade the substrate efficiently.

family (Coutinho and Henrissat, 1999; Davies and Henrissat, 1995; Henrissat and Davies, 1997).

Cellulose-Binding Domains Versus Carbohydrate-Binding Modules

In addition to the catalytic module, free cellulases and hemicellulases usually contain at least one cellulose-binding domain (CBD) as an integral part of the polypeptide chain (Linder and Teeri, 1997; Tomme et al., 1995a). The CBD serves predominantly as a targeting agent to direct and attach the catalytic domain to the insoluble crystalline substrate. Like the catalytic domains, the CBDs are categorized into a series of families according to sequence homology and consequent structural fold.

In some cases, the term "CBD" is deceptive because not all of the CBDs bind to crystalline cellulose. Some families (or subfamilies or family members) bind either preferentially or additionally to other insoluble polysaccharides, e.g., xylan or chitin. For example, the family-5 CBD and some of the members of the family-3 CBDs bind to chitin as well as cellulose (Brun et al., 1997; Morag et al., 1995). Moreover, the family-2 CBDs can be divided into two subfamilies, one of which indeed binds preferentially to insoluble cellulose, but the other binds to xylan (Boraston et al., 1999). The molecular basis for this was proposed to reflect the fact that in the first subfamily, 3 surface-exposed tryptophans contribute to cellulose binding (Simpson et al., 1999b; Williamson et al., 1999). However, in the case of the xylan-binding members, one of these tryptophans is missing, whereas the other two assume a different conformation, thereby allowing them to stack against the hydrophobic surfaces of two xylose rings of a xylan substrate. Other types of CBD prefer less crystalline substrates (e.g., acid-swollen cellulose), single

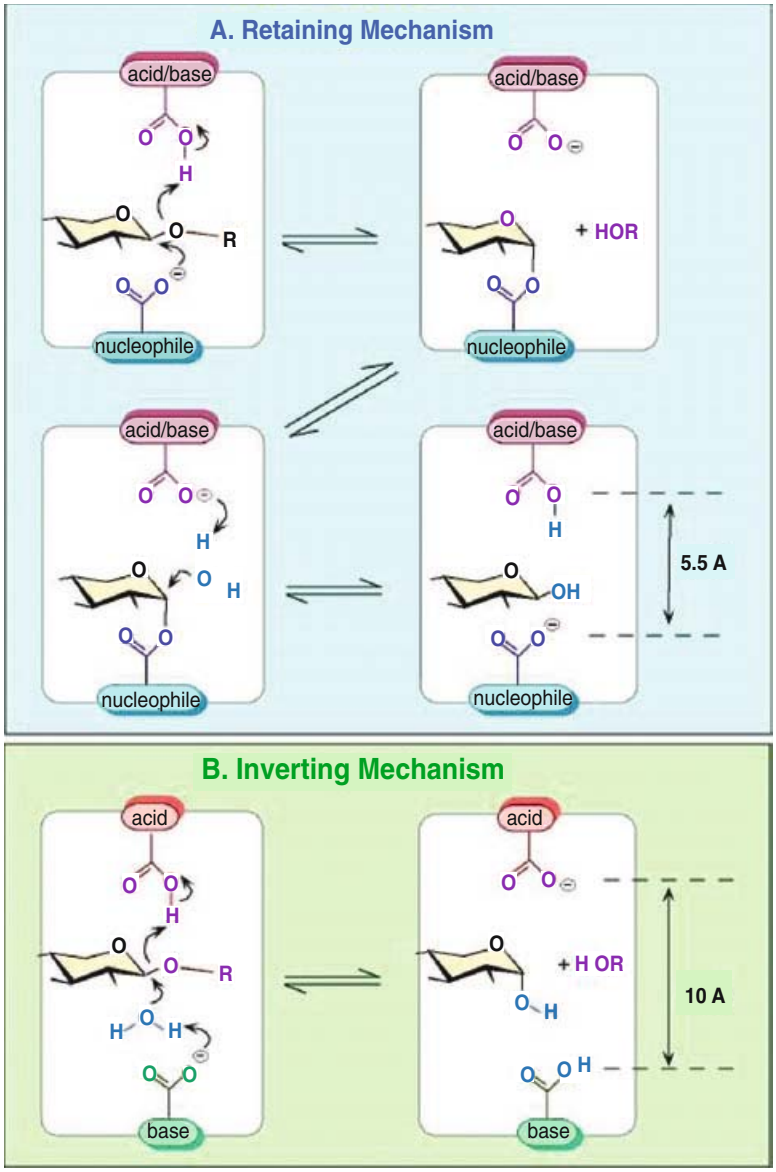

Fig. 6. The two major catalytic mechanisms of glycosidic bond hydrolysis. A. The retaining mechanism involves initial protonation of the glycosidic oxygen via the acid/base catalyst with concomitant formation of a glycosyl-enzyme intermediate through the nucleophile. Hydrolysis of the intermediate is then accomplished via attack by a water molecule, resulting in a product that exhibits the same stereochemistry as that of the substrate. B. The inverting mechanism involves the single-step protonation of the glycosidic oxygen via the acid/base catalyst and concomitant attack of a water molecule, activated by the nucleophile. The resultant product exhibits a stereochemistry opposite to that of the substrate. The type of mechanism is conserved within a given glycosyl hydrolase family and dictated by the active-site architecture and atomic distance between the acid/base and nucleophilic residues (aspartic and/or glutamic acids).

cellulose chains and/or soluble oligosaccharides, e.g, laminarin (1,3-β-glucan) and barley 1,3/1,4-β-glucan (Tomme et al., 1996; Zverlov et al., 2001). Still others exhibit alternative accessory function(s), a topic to be described below in more detail. Moreover, the CBDs responsible for the primary binding event may further disrupt hydrogen bonding interactions between adjacent cellulose chains of the microfibril (Din et al., 1994), thereby increasing their accessibility to subsequent attack by the hydrolytic domain.

Consequently, the concept of CBD has been broadened and redefined as "CBM" i.e., carbohydrate-binding module (Boraston et al., 1999; Coutinho and Henrissat, 1999). To date (March 2001), 26 different CBM families have been described. The structures of CBDs from a number of families and subfamilies have been determined, and an understanding of their structures has provided interesting information regarding the mode of binding to cellulose. Those that bind to crystalline substrates, appear to do so via a similar type of mechanism. One of the surfaces of such CBDs is characteristically flat and appears to complement the flat surface of crystalline cellulose. A series of aromatic amino acid residues on this flat surface form a planar strip (Mattinen et al., 1997; Simpson and Barras, 1999a; Tormo et al., 1996) that stack opposite the glucose rings of a single cellulose chain. In addition, to the planar aromatic strip, several polar amino acid residues on the same surface appear to anchor the CBD to two adja-

cent cellulose chains. The binding of the CBD to crystalline cellulose would thus involve precisely oriented, contrasting hydrophobic and hydrophilic interactions between the reciprocally flat surfaces of the protein and the carbohydrate substrate. Together they provide a selective biological interaction, which contributes to the specificity that a CBD exhibits towards its structure.

In contrast to the interaction with the crystalline cellulose surface, other CBMs seem to interact with single cellulose chains. The family-3c and family-4 CBDs preferentially bind to noncrystalline forms of cellulose and clearly have a different function in nature (Johnson et al., 1996; Sakon et al., 1997; Tomme et al., 1996). For example, the role of family-4 CBD may be to recognize, bind to and deliver an appropriate catalytic module to a cellulose chain, which has been loosened or liberated from a more ordered arrangement within the cellulose microfibril. The binding of the family-3c CBD to single cellulose chains and its remarkable role in cellulose hydrolysis will be discussed later (Fig. 9).

The Family-9 Cellulases: An Example

This section pertains to enzyme diversity and how a single type of catalytic module can be modified by the class of helper module(s) that flank its *C*- or *N*-terminus. We are only at the beginning in our understanding of how the modular arrangement affects the overall activity and function of a given enzyme.

In its simplest form, an enzyme would presumably consist of a single catalytic domain, usually with a standard CBM, which would target the enzyme to the crystalline substrate. Indeed, this is the norm for many individual glycosyl hydrolase families. However, in others, e.g., the family-9 cellulases, the catalytic domains commonly occur in tandem with a number of accessory modules. Although the story is still rather incomplete, we can discuss the currently available information regarding family 9 and draw several interesting conclusions from the few publications on this currently developing subject.

Family-9 Theme and Variations. The crystal structure of the family-9 catalytic module is known and displays an $(\alpha/\alpha)6$-barrel fold and inverting catalytic machinery. However, few of the prokaryotic family-9 enzymes consist of a solitary catalytic module (Fig. 7A). Actually, there are numerous family-9 cellulases of plant origin, the great majority of which are such lone catalytic modules that lack accessory modules. Another type of eukaryotic family-9 cellulase that lacks helper modules is produced by the termite. The prokaryotic family-9 enzymes, however, are almost invariably decorated with a variety of subsidiary modules that modulate the activity of the catalytic module.

Microbial family-9 cellulases commonly conform to one of the themes shown in Fig. 7. In one of these, the catalytic module is followed immediately downstream by a fused family-3c CBM (Fig. 7B). This particular type of CBM imparts special characteristics to the enzyme (see below). A second theme consists of an immunoglobulin-like (Ig) domain (of unknown function) immediately upstream to the catalytic domain (Fig. 7C). A variation of the latter theme includes a family-4 CBM at the *N*-terminus of the enzyme, followed by an Ig domain and family-9 catalytic domain (Fig. 7D). In addition to the above-described modular arrangement, each of the free prokaryotic enzyme systems includes a standard CBD that binds strongly to crystalline cellulose.

Until very recently, there has been but one example in the prokaryotic world of a family-9 enzyme that contains no helper domain. This is the family-9 glycosyl hydrolase of the cellulosomal scaffoldin from the cellulolytic anaerobic bacterium, *Acetivibrio cellulolyticus* (Ding et al., 1999). The *A. cellulolyticus* enzyme forms part of a multimodular scaffoldin, but the catalytic module appears to be a functionally distinct entity that lacks adjoining helper modules. The other modules are conventional scaffoldin-associated modules, e.g., cohesins and a true cellulose-binding CBD. More recently, a dockerin-containing cellulosomal family-9 enzyme from *Clostridium cellulovorans* has been sequenced and also seems to lack adjoining helper modules (Tamaru et al., 2000b).

This thematic arrangement of the family-9 cellulases is mirrored in the respective sequences of the catalytic modules. The divergent sequences are reflected by the phylogenetic relationship of the parent cellulases (Fig. 8). Thus, the simplest cellulases (the group A eukaryotic cellulases from plants) that lack adjacent helper modules are all phylogenetically related (theme A). Interestingly, the catalytic module of CipV from *A. cellulolyticus* is distinct from the other groups designated in Fig. 8, but closest to the plant enzymes, as might be anticipated from its lack of a helper module. In a similar manner, catalytic modules from cellulases that are fused to a family-3c CBD (group B), all map within the same branch (theme B). On the other hand, the catalytic modules that bear an adjacent Ig-like domain all fall into a cluster on the opposite side of the tree. Cellulases which have the Ig-like domain only (theme C) occupy a small separate branch and those that also include a family-4 CBD (theme D) that develop distally to form a separate subcluster.

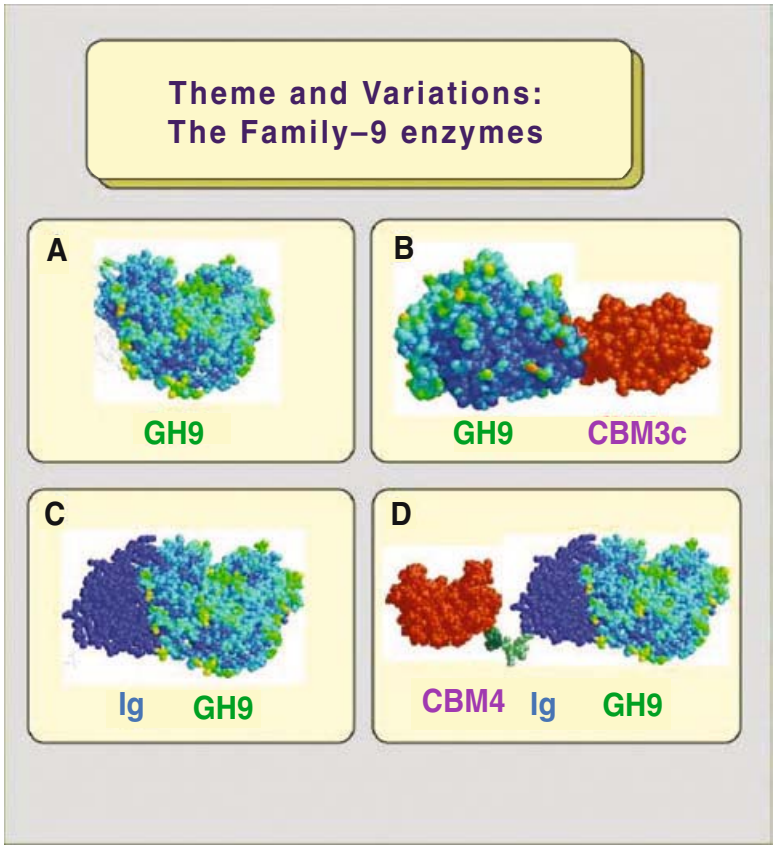

Fig. 7. Theme and variations: schematic view of the modular arrangement of the family-9 glycosyl hydrolases. A. The solitary catalytic domain. B. The catalytic domain and fused family-3c cellulose-binding domain (CBD). *C. Immunoglobulin*-like (Ig) domain, fused to the catalytic domain. D. Successive family-4 CBD, Ig and catalytic domains. The representations of the different modules are based on their known structures and are presented sequentially, left-to-right, from the *N*- to *C*-terminus. Structures (Ribbon diagrams produced by RasMol 2.6) in "A" and "B" are derived from cellulase E4 from *Thermomonospora fusca* (PDB code, 1TF4), those in "C" and "D" are from the CelD endoglucanase of *C. thermocellum* (PDB code, 1CLC). The figure used for the family-4 CBD in "D" is derived from the nuclear magnetic resonance (NMR) structure of the *N*-terminal CBD of *Cellulomonas fimi* β-1,4-glucanase CenC (PDB code, 1ULO). The structures in "B" and "C" are authentic views of the respective crystallized bi-domain protein components. The CBD in "D" has been placed manually to indicate its *N*-terminal position in the protein sequence, but its spatial position in the quaternary structure and the structure of the linker segment remains unknown.

Theme A enzymes: CipV Acece, CipV scaffoldin from the cellulolytic bacterium, *A. cellulolyticus* (AF155197); and plant (eukaryotic) cellulases from *Prunus persica* (X96853), *Populus alba* (D32166), *Citrus sinensis* (AF000135), *Persea americana* (M17634), *Pinus radiata* (X96853), *Arabidopsis thaliana* (X98543), *Phaseolus vulgaris* (M57400), *Capsicum annuum* (X97189), *Lycopersicon esculentum* (U20590).

Theme B enzymes: CelF Clotm, endoglucanase F from *Clostridium thermocellum* (X60545); CelZ Closr, exoglucanase Z from *Clostridium stercorarium* (X55299); CelA Calsa, cellulase A from Caldocellum saccharolyticum (L32742); CelG Cloce, endoglucanase G from *Clostridium cellulolyticum* (M87018); CelI Clotm, endoglucanase I from *Clostridium thermocellum* (L04735); CelB Celfi, endoglucanase B from *Cellulomonas fimi* (M64644); E4 Thefu, endo/exoglucanase E4 from *Thermomonospora fusca* (M73322).

Theme C enzymes: CelJ Clotm, cellulase J from *Clostridium thermocellum* (D83704); CelD Clotm, endoglucanase D from *Clostridium thermocellum* (X04584); CelC Butfi, endoglucanase C from *Butyrivibrio fibrisolvens* (X55732).

Theme D enzymes: CbhA Clotm, cellobiohydrolase A from *Clostridium thermocellum* (X80993); CelA Psefl, endoglucanase A from *Pseudomonas fluorescens* (X12570); CelC Celfi, endoglucanase C from *Cellulomonas fimi* (X57858); CelI Strre, endoglucanase I from *Streptomyces reticuli* (X65616); E1 Thefu, endoglucanase E1 from *Thermomonospora fusca* (L20094).

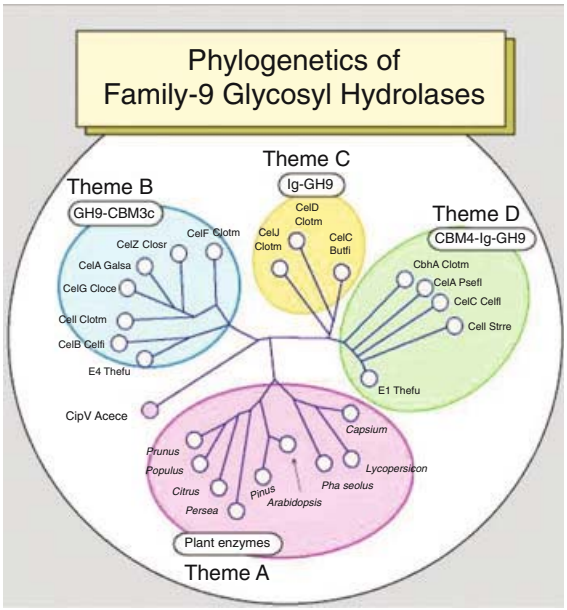

Fig. 8. Phylogenetic analysis of the *N*-terminal family-9 catalytic module of CipV and its relationship with other family-9 members. The various theme groupings roughly follow the groups shown in Fig. 7. Theme A (group A) enzymes lack associated helper modules. Theme B (group B) enzymes carry a fused family-3c cellulose-binding domain (CBD) downstream to the catalytic module. Theme C (group C) and theme D (group D) enzymes carry an immunoglobulin-like (Ig) domain upstream to the catalytic module, the theme D enzymes having an additional *N*-terminal family-4 CBM.

The analysis of the designated catalytic modules was performed using GenBee, based on the respective GenBank sequences (accession codes in parentheses).

Family-9 Crystal Structures. Two crystal structures of family-9 cellulases have been elucidated, representing two subtypes of this particular family of glycosyl hydrolase. These are cellulase E4 from Thermomonospora fusca (recently reclassified as *Thermobifida fusca*; Sakon et al., 1997) and CelD from *Clostridium thermocellum* (Juy et al., 1992). These two examples are architecturally distinct—the E4 cellulase being an example of a theme B family-9 enzyme (see Figs. 7B and 8) and the CelD cellulase being a theme C enzyme. Fortunately, in both cases, one of the neighboring modules co-crystallized with the catalytic module, thus providing primary insight into their combined structures. In the case of *T. fusca* E4, the catalytic domain and neighboring family-3c CBM were found to be interconnected by a long, rigid linker sequence, which envelops about half of the catalytic domain until it connects to the adjacent CBM (Fig. 9A). In contrast, in the *C. thermocellum* CelD, the catalytic domain is adjoined at its *N*-terminus by a 7-stranded immunoglobulin-like (Ig) domain of unknown function. The comparison between the E4 and CelD cellulases indicates that a given type of catalytic module can be structurally and functionally modulated by different types of accessory domains.

Helper Modules. The family-3c CBM is special. To date, this particular type of CBM has been found in nature associated exclusively with the family-9 catalytic domain. Structurally, the CBM is homologous to the other family-3 CBMs, but contains substitutions in many important surface residues. The three-dimensional crystal structure of the E4 cellulase revealed the close interrelationship between the family-9 catalytic domain and the family-3c CBM, thus suggesting a functional role as a helper module. This CBM seems not to bind directly to crystalline cellulose but appears to act in concert with the catalytic domain by binding transiently to the incoming cellulose chain, which is then fed into the active-site cleft pending hydrolysis (Gal et al., 1997; Irwin et al., 1998; Sakon et al., 1997; Fig. 9B).

The information derived from the family-9 enzymes suggests that the activity of catalytic domains can be modulated by accessory modules. The accessory modules can either supplement or otherwise alter the overall properties of an enzyme (Bayer et al., 1998). The recurrent appearance in nature of a given type of module adjacent to a specific type of neighboring catalytic domain may indicate a functionally significant theme. These observations raise the possibility of a more selective role for certain types of CBM and other modules, whereby their association with certain types of catalytic domains could signify a "helper" role. The helper module would provide hydrolytic efficiency and alter the catalytic character of the enzyme.

New Developments in Cellulase Analysis

The biochemical characterization of cellulases is in many cases a difficult task owing to the large variety of enzyme types and modes of action. At first glance, it is an intriguing phenomenon that for such a simple reaction (i.e., the hydrolysis of the β-1,4-glucose linkage in a linear glucose chain), Nature has evolved so many types of cellulases. The vast varieties of enzymes are found not only among the different species of cellulolytic bacteria but also within the same organism. The reason for this extensive diversity comes from the insoluble nature of cellulose and the fact that, although the chemical com-

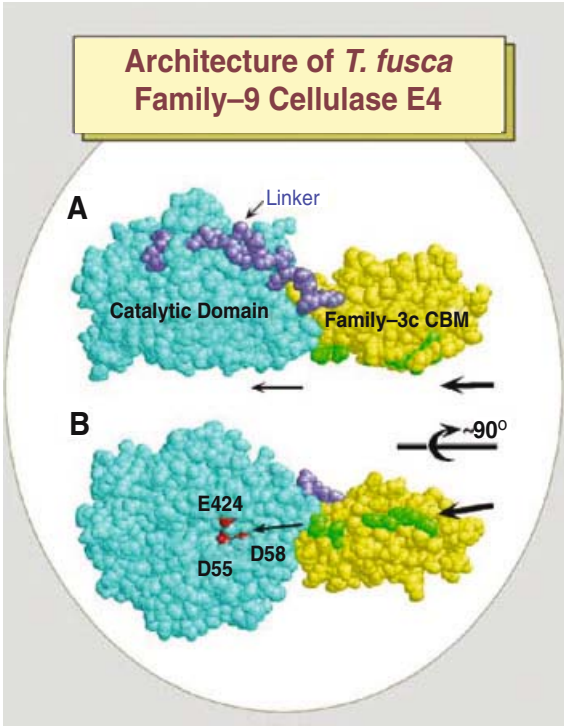

Fig. 9. Structural aspects of family-9 theme-B cellulase E4 from *Thermobifida fusca*. A. "Side view" of the E4 molecule, drawn using RasMol in spacefill mode. Shown are the family-9 catalytic module (turquoise, at left), the family-3c carbohydrate-binding module (CBM; in yellow, at right) and the intermodular linker (dark blue strip). The presumed path of a single cellulose chain, from the CBM to the catalytic domain, is shown at the bottom of the structure (arrows). The enzyme also possesses a fibronectin-like domain (FN3) and a cellulose-binding family-2 CBM (not shown). Note that the linker appears to serve a defined structural role by which the family-3c CBM is clamped tightly to the catalytic domain. Selected surface residues on the catalytic domain along the interface of both the linker and the CBM3c also serve to fasten both features tightly to the catalytic module. B. "Bottom view" of the E4 molecule (~90° rotation of "A"). From this perspective, the proposed catalytic residues (red), positioned in the active site cleft, are clearly visible. The path of the cellulose chain (arrows) passes through a succession of polar residues (green) on the bottom surface of the CBM, which would conceivably bind to the incoming cellulose chain and serve to direct it towards the active-site acidic residues of the catalytic domain.

position of the homopolymer is rather trivial, the physical and three-dimensional arrangement of the chains within the crystalline and amorphous regions of the microfibril can differ significantly.

Regarding the enzymes that degrade the substrate, the modular nature of the cellulases contributes additional degrees of complexity in our quest to characterize a given enzyme. Thus, the number, types, and arrangement of the accessory modules vis-à-vis the catalytic domain are important structural features that modulate the overall activity of the enzyme in question. This descriptive information should always be defined for a recombinant enzyme. Whenever possible, it is desirable to determine the relative contribution of the individual accessory modules to the activity of the enzyme. In this regard, the affiliation of a given module, e.g., CBM, into a defined family does not necessarily define its contribution to enzyme activity, as different specificities and functions have been attributed to different members of the same family of module. Moreover, sequences for almost 70 different "X" modules (i.e., modules for which the function remains undefined) are currently available (Coutinho and Henrissat, 1999b; Coutinho and Henrissat, 1999c), most of which probably play a binding or processing role in assisting the catalytic domain(s) in its capacity to hydrolyze the substrate.

A decade ago, the range of cellulases and hemicellulases within a given species was assessed mainly by biochemical techniques. In some cases, individual enzymes were isolated and their properties assessed using desired insoluble or soluble substrates. Another approach involved electrophoretic separation of cell-derived or cell-free extracts, and analysis of desired activities using zymograms. There are advantages and disadvantages with each of these strategies, and the employment of combined complementary approaches is always advisable.

More recently, molecular biology techniques have been used to reveal cellulase and hemicellulase genes, which can often be characterized on the basis of sequence homology with related, known genes (Béguin, 1990; Hazlewood and Gilbert, 1993). If further information is required on the structure or action of a given enzyme, the gene can then be expressed in an appropriate host organism, and the properties of the product can be characterized.

It is always instructive to compare the properties of an expressed gene product with those of the same protein isolated from the original bacterial culture. The results may be surprising; there are hazards inherent to both approaches. Expression of a gene may yield preparations with reduced or altered enzymatic properties. In this context, the expressed gene product may not have been folded properly. It is of course assumed that the investigator has taken the time and trouble to sequence the cloned gene to ensure no mutations have occurred. Unlike a gene expressed in a host cell environment, the native counterpart may have undergone post-translational modifications (e.g., glycosylation, proteolytic truncation, etc.) that improve its

physicochemical properties. Moreover, since the cellulase system in the native environment includes numerous enzyme types, often exhibiting similar molecular masses and other physical characteristics, the reputed purification of a given extracellular cellulase may still include contaminating enzymes that alter (usually increasing greatly due to synergistic action of two or more enzymes) the true enzymatic properties of the desired enzyme. The onus belongs to the conscience of the investigating scientist when publishing the properties of a given enzyme. Too often, erroneous data that enter the scientific literature are taken as fact. One should particularly be wary of comparing enzymatic activities of the same or similar types of enzymes (e.g., members of the same family) that have been published at different times and by different laboratories.

The assessment of cellulase activity is indeed a complicated undertaking, and there is no clear or standard methodology for doing so. This predicament apparently reflects a combination of factors, including the complex nature of the substrate, the multiplicity of enzymes and their synergistic action, and the variety of products formed. The fact that cellulose is an insoluble substrate converted to lower-order cello-oligosaccharide products is a further complication. It must be noted that as the cello-oligomers increase in length, they become less soluble, such that cello-octaose of 8 glucose units is no longer soluble in aqueous solutions. Moreover, the accumulation of one (particularly cellobiose) or more of the cellulose degradation products may be inhibitory towards enzymatic activity.

Today, the study of cellulase action usually includes, in addition to conventional biochemical assays, the analysis of the primary structure and the assignment of the various domains into known families. The catalytic domains can usually be assigned into one of the known glycoside hydrolase families (Henrissat and Bairoch, 1996; Henrissat and Davies, 1997). Whenever the sequence of a known polysaccharide-degrading enzyme failed to match a known family, a new family of glycosyl hydrolase was established. This approach was extensively developed in the last decade, owing to the increasing number of available DNA sequences and bioinformatics analysis tools. At the same time, an increasing number of crystal or solution structures of various catalytic and accessory domains were published that allow us to examine a new protein sequence in light of its structure. Sometimes, the publication of the structure of an accessory domain precedes determination of its function.

We can divide the analysis of a newly described prospective cellulase into several stages, such that a variety of complementary approaches are currently in use to classify the enzyme. Some of the questions one may ask are:

1) What is the primary structure (the amino acid sequence) of the enzyme? What are the binding residues and/or binding module(s) associated with the enzyme? What are its other accessory domains and their respective role(s) in catalysis or stability?
2) Is the enzyme a "true" cellulase, i.e., its preferred substrate is cellulose or cellulose degradation products, or whether the enzyme can act alone on insoluble cellulose.
3) What is the mode of action? Does the enzyme act as an endoglucanase, an exoglucanase or a processive enzyme?
4) What is the stereochemistry of the reaction? Does the enzyme exhibit an inverting or retaining mechanism?
5) What are the catalytic residues: the acid/base residue and the nucleophile that characterize a glycosyl hydrolase?

In the past ten years, several extensive reviews and book chapters dealing with different assays of cellulose degradation have been published (Ghose, 1987; Wood and Kellogg, 1988). In this treatise, we will briefly summarize the various approaches currently in use and direct the reader to the relevant literature.

While characterizing the activity of a new enzyme preparation, one has to bear in mind several secondary or indirect issues, such as the purity of the protein preparation, the sensitivity of the assay used, and the crossreactivity of the expected enzymatic activities. In some cases, only detailed kinetic analysis can provide appropriate characterization of the enzyme. As for many other types of glycosyl hydrolases, cellulases can exhibit crossreactivity with substrates of similar structure. This is particularly true when using, for example, *p*-nitrophenyl derivatized substrates that provide highly sensitive assays. However, in many cases such a soluble synthetic chromogenic substrate can fit the active-site pocket of a related but atypical enzyme, which catalyzes its hydrolysis. For example, family-10 glycosyl hydrolases are typically xylanases but can readily hydrolyze *p*-nitrophenyl cellobioside, which is a typical cellulase substrate. Without a detailed comparative kinetic analysis (kcat/Km) using different substrates, the true specificity of the enzyme might be overlooked. Given the amino acid sequence of the protein, its assignment to a given glycosyl hydrolase family will in many cases provide a reasonable general indication of its activity. The description of the modular structure provides additional knowledge that can imply how the catalytic function might be modulated, but this knowledge can also be

misleading. In the final analysis, there is no substitute for extensive biochemical and biophysical characterization of the given protein (recombinant or native) and its catalytic properties.

General procedures for assaying for cellulase and hemicellulase activities are very well documented in the *Methods in Enzymology* volume 160 (Wood and Kellogg, 1988). Conventional procedures for cellulase assay have been defined precisely by the International Union of Pure and Applied Chemists (IUPAC; Ghose, 1987). However, owing to the complexity of the substrate and enzyme systems, these procedures can only provide a starting point for understanding the true nature of the enzyme in question.

Since the publication of Part A of this treatise (Coughlan and Mayer, 1992), many of the previously reported assays of cellulase activity are still in common use. These include the use of soluble, derivatized forms of cellulose, e.g., carboxymethyl cellulose and hydroxymethyl cellulose as conventional substrates for determining endoglucanase activity. In addition, a derivatized, colored form of insoluble cellulose, i.e., azure cellulose, is frequently used as an indication of cellulase activity. Zymograms with such colored embedded substrates are useful in detecting endoglucanase or xylanase activities (Béguin, 1983). Individual soluble cello-oligomers (cellotetraose, cellopentaose, cellohexaose, etc.) are still used as substrates for analyzing enzyme action, but the reliance on these substrates as determinants for assessing cellulase activity is no longer a definitive approach. In the past decade or so, newly developed substrate analogues and reagents include thioglycoside substrates (Driguez, 1997), fluoride-derivatized sugars (Williams and Withers, 2000), chromophoric and fluorescent cello-oligosaccharides (Claeyssens and Henrissat, 1992; O'Neill et al., 1989; van Tilbeurgh et al., 1985). Recently, an ultraviolet-spectrophotometric method and an enzyme-based biosensor have been described (Bach and Schollmeyer, 1992; Hilden et al., 2001). In addition, a novel and intriguing bifunctionalized fluorogenic tetrasaccharide has been developed as an effective reagent for measuring the kinetic constants of cellulases by resonance energy transfer (Armand et al., 1997).

The thio-oligosaccharides serve as competitive inhibitors that mimic natural substrates but are enzyme resistant (Driguez, 1997). In this type of oligosaccharide, the oxygen of a bond to be cleaved is replaced by sulfur. The thio-oligodextrins are sometimes more soluble than the native cellodextrins and longer chains can be synthesized. The modified sugars can be used in biochemical studies or crystallographic studies to gain some information about the geometry of the active site or determine the mechanism of action of an enzyme.

Determination of "True" Cellulase Activity: Solubilization of Crystalline Cellulose Substrates True cellulase activity is usually defined as the ability to solubilize to an appreciable degree insoluble, "crystalline" forms of cellulose. The extent of hydrolysis can be evaluated by turbidity assays, weight loss of insoluble material, generation of reducing power, and accumulation of soluble sugars. It is important to realize that crystalline cellulose is not of uniform composition and therefore the rate of catalysis is in most cases not linear with time or enzyme concentration. Notably, the different preparations of crystalline cellulose contain varying levels of loosely associated loops and chains. The latter are readily accessible to hydrolysis by a given enzyme and lead to relatively high initial rates of activity, which do not reflect the actual degree of true cellulase activity. For example, such loose chains can be degraded by a relatively ineffectual enzyme, whereas the crystalline portions of the substrate will be immune to further hydrolysis by the same enzyme. To overcome these difficulties, IUPAC suggests determining the amount of enzyme required to achieve digestion of 5.2% of the insoluble substrate (e.g., filter paper) in 16 h (Ghose, 1987; Irwin et al., 1993).

Cellulose substrates commonly in use include Avicel, filter paper, cotton, Solka Floc, and more recently bacterial cellulose from *Acetobacter aceti* and algal cellulose prepared from *Valonia*. Consequently, these assays should be treated as a relative and not quantitative assessment. The nature of the original substrate selected—especially its extent of crystallinity—should always be taken into account. Proper controls and reference substrates should always be used. One should be wary about comparison among results reported by different laboratories and even by different researchers in the same laboratory. Nevertheless, such assays give an excellent indication of whether a given enzyme preparation exhibits substantial activity towards crystalline cellulose substrates.

Endoglucanase Versus Exoglucanase Activity As discussed earlier in this chapter, the cellulases have traditionally been divided into either endoglucanases or exoglucanases (Fig. 4). The biochemical or enzymatic assays that discriminate between these two modes of action usually involve soluble forms of cellulose, i.e., carboxymethyl or hydroxymethyl derivatives of cellulose. The action of a given enzyme on these substrates is followed by determining the

amount of reducing ends generated by the enzyme and the degree of polymerization (DP). The reducing power is usually determined either by using reagents such as 3,5-dinitrosalicylic acid (DNS; Miller et al., 1960), ferricyanide (Kidby and Davidson, 1973), or copper-arseno molybdate (Green et al., 1989; Marais et al., 1966).

Despite their traditional popularity, these two methods are intrinsically disadvantageous, owing to interference by metal ions and certain buffers. Moreover, such assays are sensitive to the chain length of the reducing end. A more recent approach involves the use of disodium 2,2′-bicinchoninate (BCA) for determination of reducing sugar. This procedure is more sensitive than the conventional methods and gives comparable values of reducing sugars for cellodextrins of different lengths (Doner and Irwin, 1992; Garcia et al., 1993; Vlasenko et al., 1998; Waffenschmidt and Jaenicke, 1987).

Viscosity-based measurements represent the most common approach for assessing the degree of polymerization. This approach is highly sensitive for internal bond cleavage, which leads to significant reduction of the average molecular weight of the substrate. The comparison between the amount of reducing sugars generated and the average molecular weight (i.e., viscosity or fluidity of the soluble cellulose substrate) gives a very good indication whether an enzyme is essentially exo- or endo-acting.

The average degree of polymerization also can be evaluated by size-exclusion chromatography either alone (Srisodsuk et al., 1998; Teeri, 1997) or combined with multiangle laser light scattering (Vlasenko et al., 1998). Mass spectometric procedures also can be applied to determine the identity and distribution of degradation products following hydrolysis of cellulosic substrates by an enzyme (Hurlbert and Preston, 2001; Rydlund and Dahlman, 1997). The mode of enzymatic action also can be appraised by determining the increase in reducing power associated with the insoluble versus the soluble fraction of the substrate. Increase in the proportion of reducing sugars associated with the soluble fraction indicates an exo-type of activity whereas a relatively large increase in the insoluble fraction would suggest an endo-type of activity (Barr et al., 1996).

Exocellulases can exhibit different specificities depending on their preference for the reducing or nonreducing end of the cellulose chain (Barr et al., 1996; Teeri, 1997). This feature of an exocellulase can be determined either by using oligosaccharide substrates labeled by tritium or 18O at the reducing end. Other procedures involve NMR, HPLC and/or mass spectrometric analysis of products released from native (unlabeled) cello-oligosaccharides. Within the past decade, the 3-D structures of enzyme-substrate complexes have been obtained, and the specificities of the enzyme can be interpreted directly from the data (Davies and Henrissat, 1995; Davies et al., 1998; Divne et al., 1998; Juy et al., 1992; Notenboom et al., 1998; Parsiegla et al., 1998; Rouvinen et al., 1990; Sakon et al., 1997; Zou et al., 1999).

Processivity One of the major recent conceptual advances in assessing the mode of enzymatic action of a cellulase is the concept of "processivity." Processive enzyme action can be defined as the sequential cleavage of a cellulose chain by an enzyme. In effect, exoglucanases are by nature and structure processive enzymes. Their tunnel-like active site thus allows processive action on the cellulose chain. Endoglucanases, however, were thought to be intrinsically nonprocessive. However, the traditional distinction between exo- and endocellulases was modified recently.

Experiments combining two or more purified cellulases have shown that synergism can even be detected upon mixing two different types of exo-acting enzymes. Such experiments led to the recognition that the exo enzymes can operate on both ends (i.e., the reducing and nonreducing ends) of the cellulose chain. Some enzymes, however, exhibit both endo and exo activities, although in such cases, the endocellulase activity is usually very low. In attempts to explain these phenomena, the concept of processivity was proposed, by which the activity of the enzyme is characterized by the sequential hydrolysis of the cellulose chain. Implicit in this concept is the notion that the catalytic site of the enzyme remains in continual and intimate contact with a given chain of the cellulose substrate.

A more complete mechanistic picture of the processive nature of such cellulases was revealed with the advent of high-resolution 3-D structures. It was thus demonstrated that the cellulose chain makes contact with the protein at multiple sites, either via a tunnel-shaped structural element (such as that observed in the family-48 enzymes) or by a special type of CBM (such as the family-9 theme B cellulases). These arrangements allow the threading of the cellulose chain into the active site, and, following initial cleavage at the end of the chain, the enzyme can move along the chain and position itself for the next cleavage. In addition to this processive nature of the active site, these enzymes also can make classic endo cleavages thus generating new ends.

Biochemically processive enzymes exhibit characteristics between endo- and exoenzymes. They have low but detectable endo activity towards soluble derivatives of cellulose (i.e.,

CMC), and may or may not possess exo activity on such substrates. With insoluble substrates, they will generate reducing power with a ratio between the soluble to the insoluble fractions of about 7. Endocellulases usually give a ratio of less than 2, whereas exocellulases produce a ratio of 12 to 23 (Irwin et al., 1998).

Once the processive nature of an enzyme has been indicated experimentally, molecular insight into the mechanisms responsible for this feature can be gained by determining the 3-D crystal structure of the active site together with model cellodextrins. In the case of the cellulases, the crystal structure of the catalytic domain together with the fused module, combined with accumulating enzymatic activity data, allowed further postulation as to the accessory role of the fused module. The fused CBM presumably interacts with a single cellulose chain and feeds it into the active site. Interestingly, this domain does not bind crystalline cellulose, but is inferred to act in dynamic binding of the single cellulose chain prior to its hydrolysis, thereby imparting the quality of processivity to the enzyme. Once such a property is associated with a given type of enzyme, the primary structure of the protein can now be used as an indication for all such enzymes. In the case of the family-9 theme B enzymes, it is now possible to identify the catalytic domain (e.g., glycosyl hydrolase family 9) and the additional accessory domains (in this case, family-3c CBM). Thus, the primary structure may by itself give a strong indication of the nature of the enzyme itself. Of course, the ultimate identification as to the mechanism of enzyme activity will come from the detailed 3-D structure of the enzyme-substrate complex.

An intriguing recent development in the analysis of the cellulolytic action of a given cellulase or a mixture of cellulase is the direct transmission electron microscopic (TEM) observation of the enzymatic action on bacterial cellulose ribbons. The approach provides information as to the endo or exo preference of the enzyme, the extent of processivity as well as the directionality of hydrolysis (i.e., from the reducing to the nonreducing ends or vice versa). This strategy has been used to study the hydrolysis of bacterial cellulose ribbons by individual purified enzymes, mixtures of purified enzymes, and intact cellulosomes.

Mechanism of Catalysis The mechanism of catalysis of cellulases address issues such as stereochemistry, binding and active-site residues and transition state intermediates. Excellent reviews have been published recently covering many of these (Ly and Withers, 1999; McCarter and Withers, 1994; Rye and Withers, 2000; Sinnott, 1990; White and Rose, 1997; Withers, 2001; Withers and Aebersold, 1995; Zechel and Withers, 2000). The fact that the stereochemistry and catalytic residues are conserved between members of the same family allows the putative identification of these elements if one member of the given (glycosyl hydrolase) has been characterized biochemically (Henrissat and Bairoch, 1996; Henrissat et al., 1995; Henrissat and Davies, 1997).

The sterochemistry of the reaction can in most cases be determined by proton NMR spectroscopy or by using chromatography systems that allow the resolution of anomeric species. In the case of NMR, the reaction between the test enzyme and its substrate is carried out in deuterated water (D_2O) and the appearance of the anomeric proton can be easily detected. Thus, for the degradation of cellulose, a retaining enzyme would produce a product in the β configuration whereas an inverting enzyme would yield the α-sugar.

The catalytic residues can be identified by performing site-directed mutagenesis on conserved acidic residues and studying the catalytic properties of the mutants with substrates bearing different leaving groups. Commonly used phenol substituents include the following, listed in order of leaving group ability (pKa values shown parenthetically): 2,4-dinitro (3.96) > 2,5-dinitro (5.15) > 3,4-dinitro (5.36) > 2-chloro-4-nitro (5.45) > 4-nitro (7.18) > 2-nitro (7.22) > 3,5-dichloro (8.19) > 3-nitro (8.39) > 4-cyano (8.49) > 4-bromo (9.34; Tull and Withers, 1994). In retaining enzymes, the nucleophilic residue can be identified directly by trapping the intermediate with an appropriate inhibitor. Such inhibitors include model saccharides containing a fluorine substituent in the 2- or 5-position and a good leaving group, such as fluoride or dinitrophenolate (Williams and Withers, 2000). The substituted substrate forms a relatively stable covalent substrate-enzyme complex, involving the nucleophile residues. The complex is then subjected to proteolytic cleavage and sequencing of the glycosylated peptide. Recently, the use of protocols involving combined liquid chromatography and mass spectrometry has facilitated the identification of the modified residues.

The acid-base residue in a retaining enzyme can be identified by a combination of kinetics-based methodologies. Mutation of this residue (usually to alanine) should affect the rate of both chemical steps, i.e., glycosylation and deglycosylation, though the effect on each step should be different. The effect on the glycosylation step will depend strongly on the leaving group ability of the aglycon. Thus, rates of hydrolysis for substrates with a poor leaving group should be affected much more strongly than those with a

good leaving group. The deglycosylation step, however, will be affected equally for all substrates carrying different leaving groups, because the same glycosyl enzyme intermediate is hydrolyzed during this step. Thus, detailed kinetic analysis (i.e., determination of k_{cat} and K_m) with substrates bearing different leaving groups can reveal whether the corresponding mutation is the acid-base residue. It should be noted that this approach requires synthetic substrates that are not necessarily recognized by all families of enzymes and are not necessarily commercially available. For example, the family-11 xylanases fail to hydrolyze *p*-nitrophenyl xylobioside, which is an excellent substrate for the family-10 xylanases. The assignment of the acid-base catalyst can also be examined by use of external nucleophilic anions, such as azide. In this approach, termed "azide rescue," the small azide anion enters the vacant space created by alanine replacement of the acidic amino acid residue. The azide reacts with the anomeric carbon instead of a water molecule to form the corresponding β-glycosyl azide product. In the absence of an acid-base catalyst, which normally provides general base catalysis during the second step, the deglycosylation step is severely affected. Thus, the acceleration of the reaction by the mutant enzymes in the presence of these external anions (provided that the second step is rate limiting) is a good indication that a mutant residue is the acid-base catalyst. Finally, the assignment of the acid-base catalyst can be tested by comparing the pH-dependence profiles for the wild-type and mutant enzymes. The profile for the native enzyme would approximate a perfect bell shape curve, reflecting the ionization of the two active site carboxylic acids, whereas the no reduction of activity at high pH values would be observed for the mutant. This pH dependency approach is also applicable for identifying the nucleophile residues and the catalytic residues in inverting enzymes.

Prokaryotic Cellulase Systems

The cellulolytic bacteria produce a variety of different cellulases and related enzymes, which together convert the plant cell wall polysaccharides to simple soluble sugars that can subsequently be assimilated. The complement of cellulases and hemicellulases that are synthesized by a given bacterium for this purpose is referred to as its "cellulase system." Different bacteria exploit different strategies for the ultimate degradation of their substrates. The given strategy is reflected by the complement and type(s) of enzymes produced by a given bacterium. The bacterial cellulase system may be characterized by free enzymes, cell-bound enzymes, multifunctional enzymes, cellulosomes, or any combination of the latter.

Cellulase enzyme systems are comprised of several different types of components, each type may exist in a multiplicity of forms. To add to the complexity, the same component may exist as free individual entities in the culture fluid, as individual entities bound to cellulose, or associated with the cell surface. Alternatively, an individual component may be organized as part of a multicomponent cellulosome complex attached to the cell surface, to the cellulose, to both, or as free complexes in the culture fluid. Furthermore, the situation existing during growth under one set of conditions (e.g., pH, temperature, distribution of carbon source, etc.) may not exist under another, or may change considerably during the course of cultivation. The bacterium reacts to these changes and its production of cellulases and/or cellulosomes may reflect the dynamics of the growth conditions.

Free Enzymes

As mentioned earlier in this chapter, the free enzymes in their simplest form comprise a catalytic module alone with no accessory domains or modules. Such enzymes often specialize in degrading soluble oligosaccharide breakdown products. Alternatively, such single-modular enzymes may rely on an intrinsic association with insoluble polysaccharide substrate such as cellulose, perhaps related to the active site of the enzyme.

A higher order level of organization and activity are free enzymes composed of a polypeptide chain that includes both a catalytic domain together with a CBM. This basic bi-modular arrangement can be further extended by the inclusion of additional types of modules or repeating units of the same module, all of which serve to modulate the activity of the catalytic domain on the substrate. The intact free enzyme, however, remains unattached to other enzymes and can work in an independent manner on a given substrate.

Cell-Bound Enzymes

Some enzymes are connected directly to the cell wall. In Gram-positive bacteria, this is frequently accomplished via a specialized type of module, the SLH (S-layer homology) module, previously shown to be associated with the cell surface of Gram-positive bacteria (Lupas et al., 1994). This arrangement may have evolved to provide a more economic degradation of insoluble substrates and to reduce competition with other bac-

teria for the soluble products, subject to diffusion in the media. As opposed to free enzymes, diffusion of an attached enzyme would itself be prevented.

Examples of enzymes, which are bound to the cell surface via an SLH module include, a family-5 cellulase and family-13 amylase-pullulanase from *Bacillus*, a family-10 xylanase from *Caldicellulosiruptor* (Saul et al., 1990), a family-5 endoglucanase from *Clostridium josui*, a family-16 lichenase and family-10 xylanase from *Clostridium thermocellum* (Jung et al., 1998), and a variety of enzymes (family-10 xylanases, a family-5 mannanase and a family-13 amylase-pullulanase) from different species of *Thermoanaerobacter* (Matuschek et al., 1996). The modular architecture of these enzymes may be particularly complicated, containing several different modules in a single polypeptide chain, thus forming extremely large enzymes sometimes comprising over 2,000 amino acids (Fig. 10).

Multifunctional Enzymes

Some cellulases exhibit a more complex architecture in that more than one catalytic domain and/or CBD may be included in the same protein. Examples of such enzymes are the very similar cellulases from *Anaerocellum thermophilum* (Zverlov et al., 1998) and *Caldocellum saccharolyticum* (Te'o et al., 1995), both of which contain a family-9 and a family-48 catalytic domain. Other paired catalytic domains include those from family 44 and either family 5 or 9. Such an arrangement might indicate a close cooperation between two particular catalytic domains, which may lead to synergistic action on the cellulosic substrate, thus portending on a smaller scale the advent of cellulosomes.

Like the cellulases, xylanases also tend to exhibit a modular structure, being composed of multiple domains joined by linker sequences. Family-10 and -11 xylanases may be linked in the same polypeptide chain either to each other, to catalytic domains from families 5, 16 and 43 or to carbohydrate esterases (Flint et al., 1993; Laurie et al., 1997). One particularly interesting combination of multifunctional catalytic modules that appear in the same polypeptide chain is a typical xylanase together with a feruloyl esterase. Such a combination would allow the rapid cleavage of hemicellulose from the lignin in natural systems, i.e., the plant cell wall (see Fig. 3). In this manner, the xylan chain would be severed by the xylanase component (Xyn in Fig. 3) and the lignin-xylan association would be disconnected simultaneously by the feruloyl acid esterase (Fae in Fig. 3).

Indeed, some xylanases are extremely complex in their modular architecture (Fig. 11). In addition to multiple catalytic modules, these enzymes often contain several different types of CBMs. Why would such a xylanase contain several types of CBM? And why would a xylanase contain a cellulose-specific CBD? Unlike the case of various cellulases, for which the CBD is usually essential for degrading insoluble crystalline cellulose, the CBMs of a hemicellulase do not necessarily bind the hemicellulose component (xylan). In some cases, its CBM is in fact an authentic CBD that situates the hemicellulase on the insoluble plant cell wall material by utilizing the most abundant and most stable cell-wall component—cellulose. Indeed, the three family-3 CBDs (CBM3) shown in Fig. 11 apparently bind to crystalline cellulose. Why would this xylanase require three tandem copies of the same type of CBD is yet another mystery that should eventually be addressed experimentally. At any rate, once bound via the

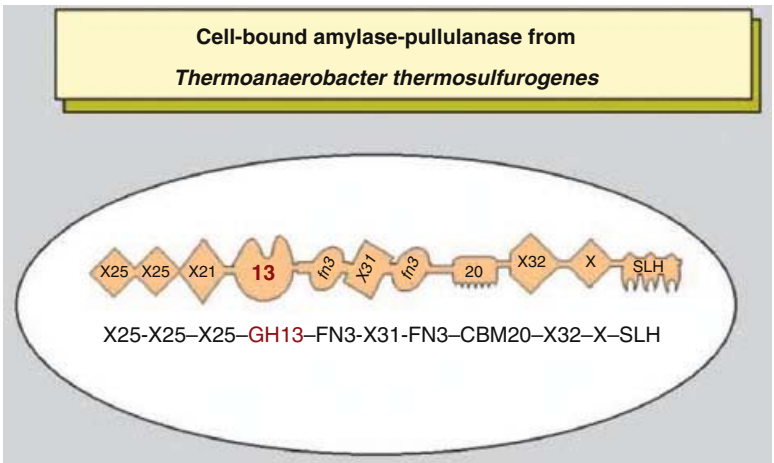

Fig. 10. A very large, cell-surface enzyme from *Thermoanaerobacter thermosulfurogenes*. The 1861-residue enzyme contains an SLH module, which is believed to mediate the attachment of the enzyme to the cell surface in Gram-positive bacteria. The enzyme contains a multiplicity of modules, which apparently serve to regulate the hydrolytic action of its single family-13 catalytic module with the complex substrate. Several X domains of unknown function may either represent as yet undescribed catalytic functions, carbohydrate-binding activities or structural entities.

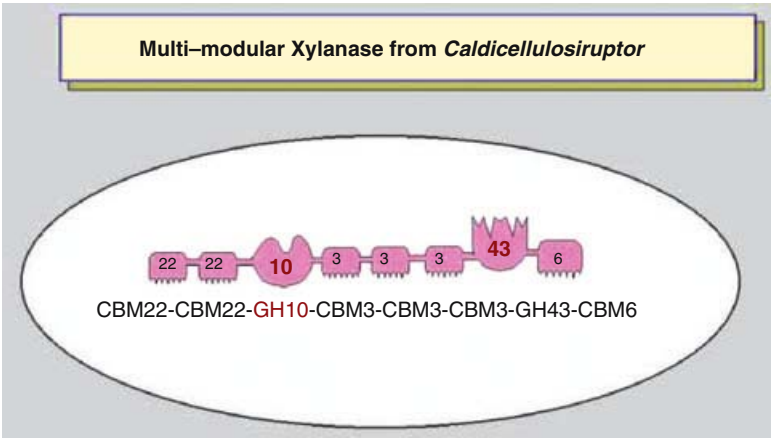

Fig. 11. A very large, multimodular xylanase from *Caldicellulosiruptor*. The 1,795-residue enzyme contains 8 separate modules, including 2 catalytic modules from families 10 (invariably a xylanase) and 43 (frequently an arabinofuranosidase). These are modulated by numerous carbohydrate-binding modules, which include 3 from family 3 (likely for binding to crystalline cellulose), 2 from family 22 (newly classified and shown to function in xylan binding and one from family 6.

cellulose component of the plant cell wall composite substrate, the immobilized enzyme then acts on the accessible and appropriate hemicellulose components. Once thus situated on the plant cell wall, another type of CBM on the same molecule would then assist in the binding to the xylan (or mannan, etc.) component to direct the appropriate catalytic module to its true substrate. Hence, the modular proximity of the xylanase shown in Fig. 11 would presumably indicate that the two CBM22s would modulate the action of the family-10 catalytic module, and the C-terminal CBM6 would facilitate the catalysis by the family-43 module. Together, the two catalytic modules would act synergistically to degrade susceptible plant cell wall components. In this context, the complex architecture of a xylanase would reflect the complex chemistry of its substrate and the neighboring polymers of its immediate environment in the plant cell wall.

Cellulosomes

Cellulosomes are multienzyme complexes, which bind to and catalyze the efficient degradation of cellulosic substrates. The first cellulosome was discovered while studying the anaerobic thermophilic bacterium, *Clostridium thermocellum* (Bayer et al., 1983; Lamed et al., 1983). Since its initial description in the literature, the cellulosome concept has been subject to numerous reviews (Bayer et al., 1996; Béguin and Lemaire, 1996; Belaich et al., 1997; Doi et al., 1994; Doi and Tamura 2001; Felix and Ljungdahl, 1993; Karita et al., 1997; Lamed and Bayer, 1988; Lamed and Bayer, 1991; Lamed and Bayer, 1993; Lamed et al., 1983; Shoham et al., 1999).

Cellulosomes in *C. thermocellum* exist in both cell-associated and extracellular forms, the cell-associated form being associated with polycellulosomal protuberance-like organelles on the cell surface. Later, cellulosomes were detected in other cellulolytic organisms (Lamed et al., 1987; Mayer et al., 1987), including *Acetivibrio cellulolyticus*, *Bacteroides cellulosolvens*, *Clostridium cellulovorans* and *Ruminococcus albus*, all of which contained protuberance-like organelles on their surfaces (Bayer et al., 1994; Lamed and Bayer, 1988; Fig. 12).

The cellulosomes contain numerous components, many of which were shown to display enzymatic activity. They also contain a characteristic nonenzymatic high-molecular-weight component. This component proved to be highly antigenic and glycosylated (Bayer et al., 1985). The cellulosomal enzymatic subunits from this organism showed a broad range of different cellulolytic and xylanolytic activities (Morag et al., 1990). Ultrastructural evidence indicated the multisubunit nature of the cellulosome (Fig. 13).

Eventually, genetic engineering techniques led to the sequencing of cellulosomal genes in *C. thermocellum* and several other bacteria, thus confirming the existence of cellulosomes as a major paradigm of prokaryotic degradation of cellulose and related plant cell wall polysaccharides.

Clostridium Thermocellum Cellulosomal Subunits and Their Modules

A simplified schematic view of the cellulosome from *C. thermocellum* and its interaction with its substrate is shown in Fig. 14. The cellulosomal enzyme subunits were found to be united into a complex by means of a unique class of nonenzymatic, multimodular polypeptide subunit, termed "scaffoldin" (Bayer et al., 1994). The

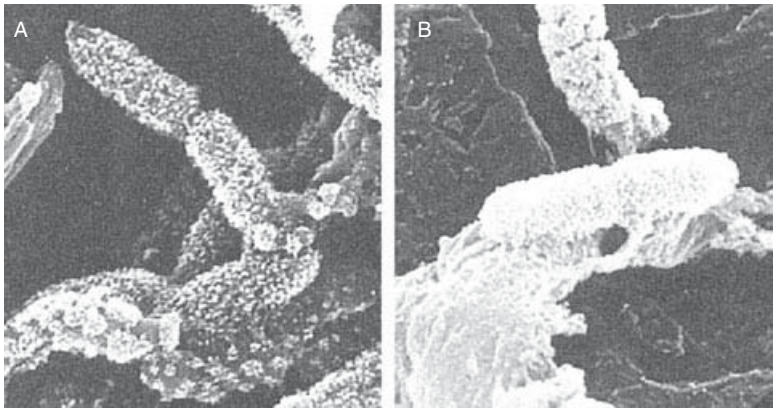

Fig. 12. Scanning electron microscopy (SEM) of *Acetivibrio cellulolyticus* showing the presence of large characteristic protuberance-like structures on the cell surface. Cells are shown in the free state (A) or bound to cellulose (B). Cell preparations were treated with cationized ferritin before processing. Cationized ferritin has been shown to stabilize such surface structures, thus allowing their ultrastructural visualization (Lamed et al., 1987a; Lamed et al., 1987b). Without pretreatment with cationized ferritin, these structures are invisible. In (B), the cellulose-bound cells appear to be connected to the substrate via structural extensions of the cell-surface protuberances. Such a mechanism was originally observed for other cellulolytic prokaryotes, e.g., *C. thermocellum* (Bayer and Lamed, 1986).

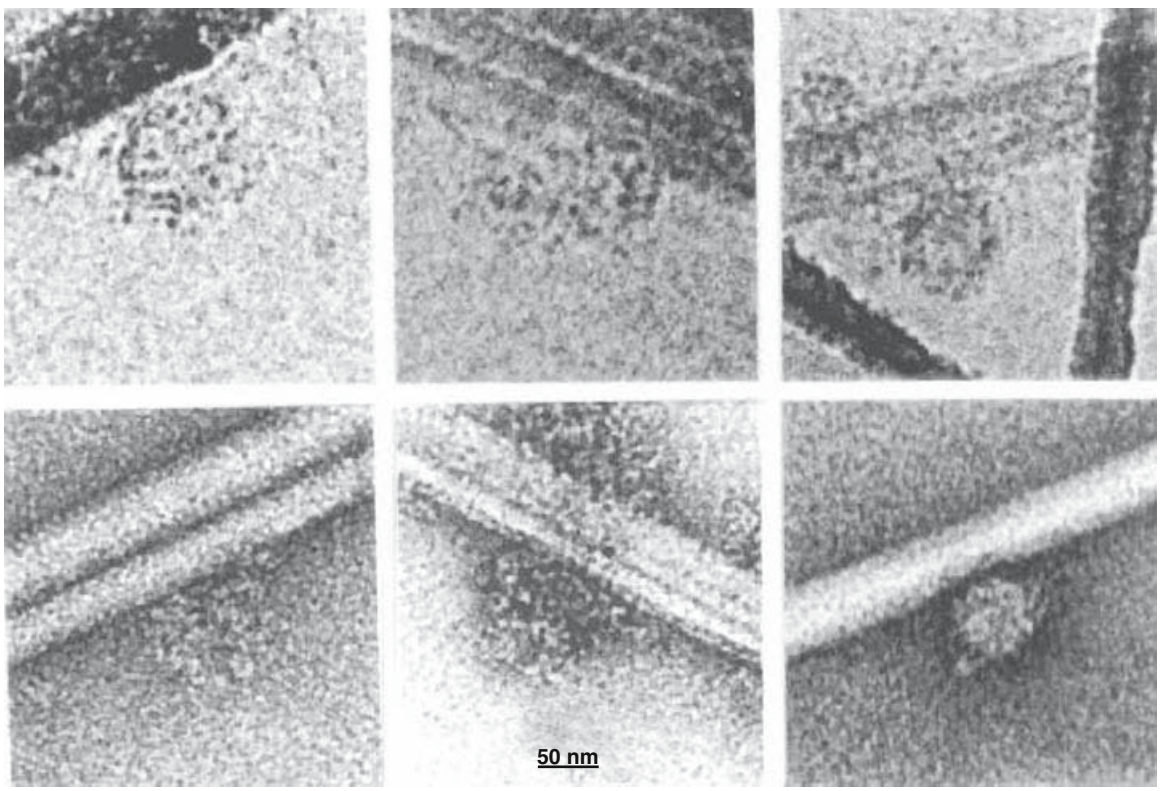

Fig. 13. Comparison between negative staining (bottom) and cryo images (top) of the purified cellulosome from *C. thermocellum*, adsorbed on cellulose microcrystals from the algae, *Valonia ventricosa*. The images illustrate the diversity of shapes of the cellulosomes, which adopt either compact or loosely organized ultrastructure. In the cryo images, the subunits of the cellulosomes (i.e., the individual enzymatic components) are clearly visible. Micrographs courtesy of Claire Boisset and Henri Chanzy (CNRS—CERMAV, Grenoble, France).

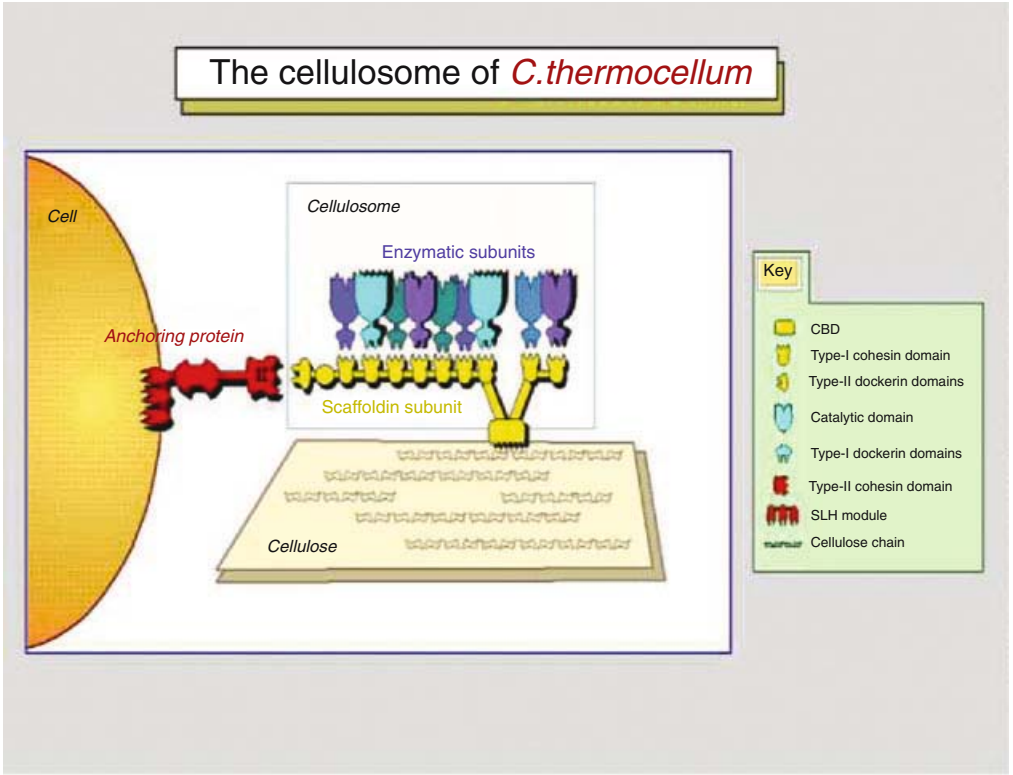

Fig. 14. Simplified schematic view of the molecular disposition of the cellulosome and one of the associated anchoring proteins on the cell surface of *C. thermocellum*. The key defines the symbols used for the modules, from which the different cellulosomal proteins are fabricated. The progression of cell to anchoring protein to cellulosome to cellulose substrate is illustrated. The SLH module links the parent anchoring protein to the cell. The cellulosomal scaffoldin subunit performs three separate functions, each mediated by its resident functional domains: 1) its multiple type-I cohesins integrate the cellulosomal enzymes into the complex via their resident type-I dockerins, 2) its family-IIIa CBD binds to the cellulose surface, and 3) its type-II dockerin interacts with the type-II cohesin of the exocellular anchoring protein.

scaffoldins usually contain a family-3 CBD that provides the cellulose-binding function. The scaffoldins also contain multiple copies of a definitive type of module, called "the cohesin domain." The cellulosomal enzyme subunits, on the other hand, contain a complementary type of module, called "the dockerin domain." The interaction between the cohesin and dockerin domains provides the definitive molecular mechanism that integrates the enzyme subunits into the cellulosome complex (Salamitou et al., 1994; Tokatlidis et al., 1991; Tokatlidis et al., 1993). Cohesin and dockerins are considered to be cellulosome "signature sequences"—i.e., their presence is a good indication of a cellulosome in a given bacterium (Bayer et al., 1998).

The major difference between free enzymes and cellulosomal enzymes is that the free enzymes usually contain a CBD for guiding the catalytic domain to the substrate, whereas the cellulosomal enzymes carry a dockerin domain that incorporates the enzyme into the cellulosome complex. Otherwise, both the free and cellulosomal enzymes contain very similar types of catalytic domains. The cellulosomal enzymes rely on the Family-3a CBD of the scaffoldin subunit for collective binding to crystalline cellulose.

The incorporation of the multiplicity of enzyme subunits into the cellulosome complex is a function of the repeated copies of the cohesin module borne by the scaffoldin subunit. For most species of scaffoldin, the cohesins have been classified as type-I on the basis of sequence homology. The cohesin module is composed of about 150 amino acid residues. The basic structure of the cohesin is known and comprises a nine-stranded β sandwich with a jelly-roll topology (Shimon et al., 1997; Spinelli et al., 2000; Tavares et al., 1997).

The dockerin domain contains about 70 amino acids and is distinguished by a 22-residue duplicated sequence (Chauvaux et al., 1990), which bears similarity to the well-characterized EF-hand motif of various calcium-binding proteins (e.g., calmodulin and troponin C). Within this repeated sequence is a 12-residue calcium-binding loop, indicating that calcium-binding is an important characteristic of the dockerin

domain. This assumption was eventually confirmed experimentally (Yaron et al., 1995). The specificity characteristics of the cohesin-dockerin interaction also have been investigated. The results showed that four suspected residues may serve as recognition codes for interaction with the cohesin domain (Mechaly et al., 2000; Mechaly et al., 2001; Pagès et al., 1997). The three-dimensional solution structure of the 69-residue dockerin domain of a *Clostridium thermocellum* cellulosomal cellulase subunit was recently determined (Lytle et al., 2001). As predicted earlier (Bayer et al., 1998; Lytle et al., 2000; Pagès et al., 1997), the structure consists of two Ca^{2+}-binding loop-helix motifs connected by a linker; the E helices entering each loop of the classical EF-hand motif are absent from the dockerin domain.

The scaffoldin of *C. thermocellum* also contains a special type of dockerin domain. This dockerin failed to bind to the cohesins from the same scaffoldin subunit, but instead interacted with a different type of cohesin—termed "type-II cohesins"—identified on the basis of sequence homology (Salamitou et al., 1994). These cohesins are somewhat different than those of type I, having an additional segment and diversity in the latter half of the sequence. The type-II cohesins were discovered as component parts of a group of noncatalytic cell-surface "anchoring" proteins on *C. thermocellum* (Leibovitz and Béguin, 1996; Leibovitz et al., 1997; Lemaire et al., 1995; Salamitou et al., 1994). The three known anchoring proteins in *C. thermocellum* contain different copy-numbers of the type-II cohesins as illustrated in Fig. 15. Each of these anchoring proteins also contains an S-layer homology (SLH) module, analogous to those of the cell-bound enzymes mentioned above. The intervening sequences, however, between the cohesins and SLH domains are different. In any case, the type-II cohesins selectively bind the type-II dockerins, and the cellulosome (i.e., the scaffoldin subunit together with all of its enzyme subunits) is thereby incorporated into the cell surface of *C. thermocellum*.

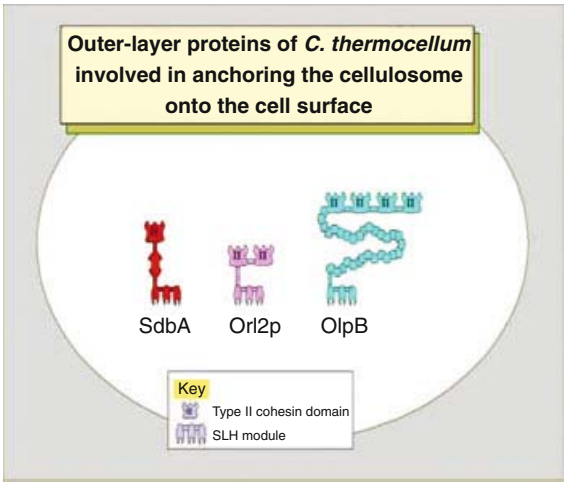

Fig. 15. Schematic representation of the known anchoring proteins of the *C. thermocellum* cell surface. Each protein bears an SLH domain that connects the protein to the cell surface via yet undefined surface components. The different proteins carry different numbers of type-II cohesins. SdbA has one cohesin, Orf2p has 2 and OlpB has 4, presumably allowing the corresponding number of scaffoldins (i.e., cellulosomes) to be attached to the given protein.

Similarity and Diversity of Scaffoldins from Different Species

The modular architecture of the known scaffoldins and their comparison to that of *Clostridium thermocellum* is presented in Fig. 16. Two new scaffoldins have recently been described for *Acetivibrio cellulolyticus* and *Bacteroides cellulosolvens* that, like *C. thermocellum*, carry dockerin domains at their C terminus (Ding et al., 1999; Ding et al., 2000). The *A. cellulolyticus* genome also includes a gene (immediately downstream of the scaffoldin gene) that contains type-II cohesins that may represent an anchoring protein. It thus seems that the arrangement of the cellulosome on the cell surface of these latter strains may be analogous to that of *C. thermocellum*. It is interesting to note that the cohesins of the *Bacteroides cellulosolvens* scaffoldin are clearly type-II cohesins and not of type I. This infers that there is not a clear linkage between the type-II cohesins and anchoring proteins.

The scaffoldins from the other clostridial species thus far described all lack "type-II dockerin" domains, the inference being that cells of *C. cellulovorans*, for example, would apparently not bear anchoring proteins that contain type-II cohesins. It thus follows that either their cellulosomes are not surface bound or, if indeed they are surface components, then their anchoring thereto is accomplished via an alternative molecular mechanism. Recently (Doi and Tamura, 2001; Tamaru and Doi, 1999a; Tamaru et al., 1999b), a cell-surface binding function has been proposed for a domain of unknown function, designated "X2" (Coutinho and Henrissat, 1999b; Coutinho and Henrissat, 1999c) of the scaffoldin from *C. cellulovorans*. On the basis of sequence alignment of a few conserved identical amino acids with S-layer proteins from Mycoplasma hyorhinis and Plasmodium reichenowi, the authors consider that this domain may be recognized as an SLH domain. The four X2 domains of the *C. cellulovorans* scaffoldin are very similar in sequence to the X-domains from the scaffoldins of *C. cellulolyticum* and *C. josui*,

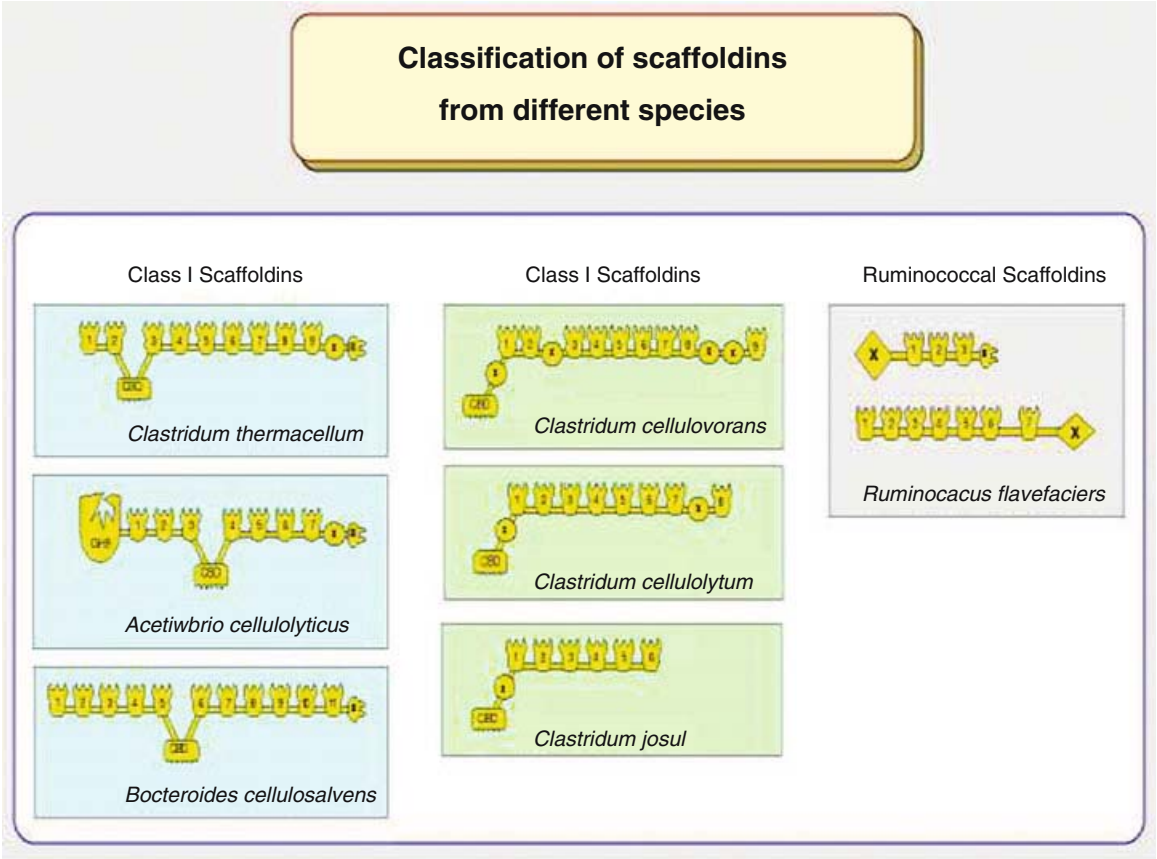

Fig. 16. Schematic view of the modular similarity and diversity of scaffoldins from different cellulosome species. Class-I scaffoldins feature an internal CBD and a *C*-terminal type-II dockerin domain. Class-II scaffoldins exhibit an *N*-terminal CBD and lack a dockerin domain. The newly described scaffoldins from *Ruminococcus flavefaciens* lack a defined CBD. The functional role of the two different X domains in the two *R. flavefaciens* scaffoldins is currently unknown. All of these scaffoldins contain multiple copies of cohesin domains.

which contain only two and one copies of this domain, respectively. If this domain functions in attaching the scaffoldin with its complement of enzymes to the cell surface, it is unclear why there would be different copy numbers of the domain in the different scaffoldins. Likewise, one of the *C. cellulovorans* cellulosomal enzyme components (EngE) also contains a triplicated segment of unknown function, designated "X48" (Coutinho and Henrissat, 1999b; Coutinho and Henrissat, 1999c) that the authors consider to be involved in cell-surface attachment (Tamaru and Doi, 1999a). In any case, final proof of the function of the X2 and X48 domains awaits biochemical examination, as has been clearly achieved for the SLH domain of the *C. thermocellum* anchoring proteins (Chauvaux et al., 1999; Lemaire et al., 1998).

Finally, two new scaffoldins have recently been sequenced from the rumen bacterium, *Ruminococcus flavefaciens* (Ding et al., 2001). Although each of the two proteins contains multiple cohesins, their sequences indicate that they are neither of type-I or type-II, but occupy their own phylogenetic branch. Interestingly, the ruminococcal scaffoldins lack a known type of CBD. Both have dissimilar X domains of unknown function, the sequences of which bear no resemblance to any other known module. Both X domains were expressed, but the resultant proteins failed to bind to cellulose. The lack of a scaffoldin CBD raises the question as to how the ruminococcal cellulosome(s) and/or the bacterium bind to the substrate. Perhaps it does so like another closely related species, *R. albus*, which binds cellulose via a noncellulosomal cell-surface protein (Pegden et al., 1998).

Schematic Comparison of Prokaryotic Cellulase Systems

In this section, we will describe schematically the similarity and diversity of representative enzyme systems, demonstrating different strategies, from

different plant cell wall degrading bacteria. It is emphasized that the accumulating information is based on what is known currently from biochemical data combined with gene sequencing and bioinformatics. The information is still rather sketchy but quite revealing when compared among different bacteria. As time progresses and the entire genomes of cellulolytic microorganisms become known, the data concerning the complement of enzymes produced by a given bacterium will be complete, and we will be able to speculate with heightened certainty how the various cellulase systems might have evolved. A survey of genes, however, does not inform us how a given bacterial system is regulated and what role(s) the bacterium and its enzyme system may play in nature. The explosive development of molecular biology techniques, however revealing, cannot supplant the fundamental contribution of biochemical and ecological approaches to the study of microbial degradation of cellulose and other plant cell wall polysaccharides.

Free Enzyme Systems

Many cellulolytic microorganisms show a very similar pattern in the types of enzymes that comprise the complement of their cellulase system. For the purposes of this discussion, the concept of "cellulase system" will include the complement of all plant cell wall hydolyzing enzymes and other glycosyl hydrolases, including the different cellulases per se, the hemicellulases (e.g., xylanases and mannanases), etc.

The cellulase system of the mesophilic cellulolytic aerobe, *Cellulomonas fimi*, is one of the first studied, and has since been one of the most studied bacterial cellulase systems (O'Neill et al., 1986; Shen et al., 1995; Whittle et al., 1982). The enzymes of this bacterium are essentially free enzymes, which allowed their early isolation and characterization. Moreover, the genes of the cellulases from this bacterium were of the earliest to have been sequenced. To date, about 10 glycosyl hydrolases have been sequenced from *Cellulomonas fimi*. Their modular composition and family associations are shown symbolically in Fig. 17. As an example of a free enzyme system, most of the enzymes bear a substrate-targeting CBM—in this bacterium, most of the CBMs are from family 2. Several of the enzymes have multiple copies of the fibronectin 3 (FN3) domain, the function of which is still unknown.

The *Cellulomonas* system includes two family-6 enzymes—an endoglucanase and an exoglucanase (cellobiohydrolase) of the types described in Fig. 4. The modularity of the endoglucanase is very simple, having the family-6 catalytic module together with a family-2 CBM. The cellobiohydrolase is a bit more complex with three additional FN3 domains that separate the same two types of modules. Another cellobiohydrolase (that exhibits processive cleavage of the substrate) is from family 48. Its general modular architecture is similar to that of the family-6 cellobiohydrolase with the substitution of the catalytic module from a different family. The cellulase system from this organism also includes two family-9 cellulases with modular themes B and D, familiar to us from the earlier description (Fig. 7). In addition, a simple family-5 cellulase and an interesting cell-borne family-26 mannanase are components of the system. The fact that an enzyme bears an SLH domain and is presumably cell-associated would underscore its importance to the cell. Finally, three xylanases are currently known for Cellulomonas fimi. One of these xylanases is a simple enzyme consisting of a family-10 catalytic domain connected to a family-2 CBM. The other two are more complicated, each containing two catalytic domains—either a family-10 or -11 domain and a carbohydrate esterase (in both cases, probably an acetyl xylan esterase; Fig. 3)—plus several CBMs. This rather complex system is probably not nearly complete, and more enzymes will inevitably be described in the future.

A second example of a free enzyme system, from the aerobic thermophilic bacterium *Thermobifida fusca* (formerly classified as *Thermomonospora fusca*), has also been studied extensively (Wilson, 1992; Wilson and Irwin, 1999). A brief comparison of its known enzyme components (Fig. 18) shows a striking resemblance to those of *Cellulomonas* (compare Figs. 17 and 18). According to known data, both species produce similar types of cellulases from families 5, 6, 9 and 48 plus xylanases from families 10 and 11. Nevertheless, the modular repertoire of the corresponding enzyme in *T. fusca* is generally somewhat simpler. For example, two of the *T. fusca* cellulases include single FN3 domains, whereas several *Cellulomonas* cellulases harbor multiple copies of the same domain. Some *T. fusca* enzymes lack accessory modules other than a cellulose-binding CBM, whereas the corresponding *Cellulomonas* enzyme is elaborated by multiple copies of accessory modules. In some cases though, the respective CBMs appear on opposite termini of the polypeptide chain (i.e., the family-48 and family-5 cellulases).

The complement of enzymes and their modular content of the free enzyme systems from *Cellulomonas* and *T. fusca* are not necessarily similar in other free enzyme systems. Many free enzyme systems, such as those of *Butyrivibrio fibrisolvens*, *Pseudomonas fluorescens*, *Fibro-*

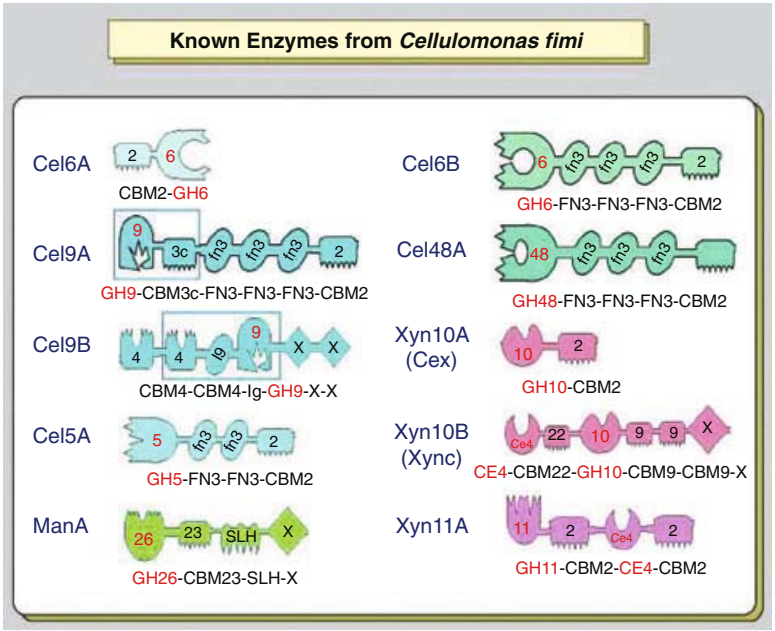

Fig. 17. *Cellulomonas fimi* cellulase system: Symbolic view of the enzyme components and their modular architecture. An example of a cell-free enzyme system. The modular content of the enzymes in this and subsequent figures is shown from (left to right) the *N*-terminus to the *C*-terminus of the polypeptide chain. The family numbers of the given domains are enumerated, the catalytic modules given in red. Key to symbols: GH, glycosyl hydrolase (e.g., cellulase, xylanase and mannanase); CE, carbohydrate esterase (e.g., acetyl xylan esterase and ferulic acid esterase); CBM, carbohydrate-binding module (e.g., CBD, cellulose-binding domain); SLH, S-layer homology (domain); FN3, fibronectin-3 (domain); Ig, immunoglobulin-like domain; and X, domain of unknown function.

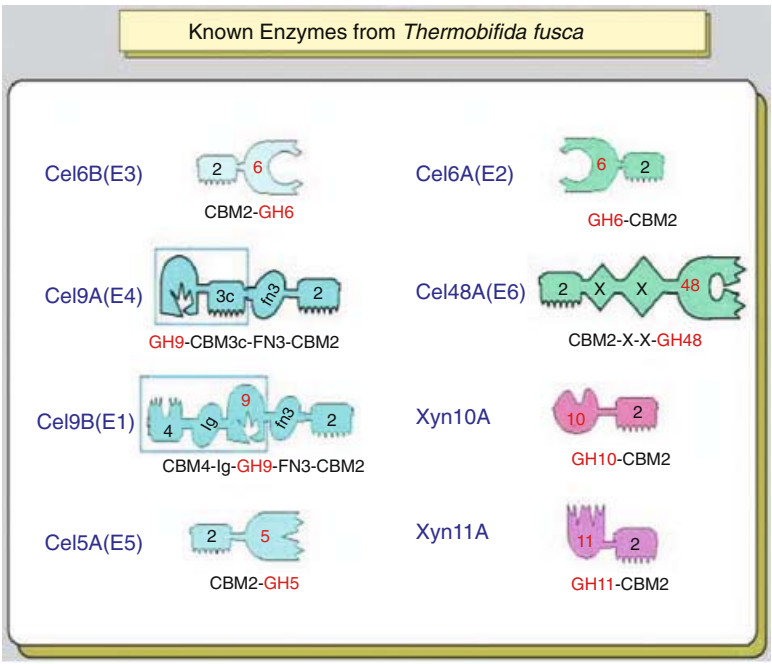

Fig. 18. *Thermobifida fusca* cellulase system. A cell-free enzyme system. The modular content of the enzy*mes is shown from (left to right) the N-terminus to the C*-terminus of the polypeptide chain. Compare with the *Cellulomonas* system (Fig. 17). Key to symbols: GH, glycosyl hydrolase (e.g., cellulase, xylanase and mannanase); CBM, carbohydrate-binding module (e.g., CBD, cellulose-binding domain); FN3, fibronectin-3 (domain); Ig, immunoglobulin-like domain; and X, domain of unknown function.

bacter succinogenes, various species of *Streptomyces*, *Erwinia* and *Thermatoga*, appear to have several cellulases, xylanases and mannanases from the common families, together with other glycosyl hydrolases, e.g., arabinosidases, lichenases, amylases, pullulanases, galactanases, polygalacturonase, glucuronidases and pectate lyases. In many of these bacterial enzymes, the family-2 CBM appears to predominate as a common cellulose-binding domain, but in others (e.g., *Erwinia*) relevant enzymes usually bear a cellulose-binding CBM from family-3. Nevertheless, in many of the free systems, many enzymes are characterized by CBMs from other families as well as other noncatalytic domains of unknown function (X domains). Once again, until the genome sequences of cellulolytic prokaryotes are widely available, we are still lim-

ited in our capacity to compare among the enzyme systems because our knowledge of their enzyme sequences is incomplete.

Multifunctional Enzyme Systems

In an extremely thermophilic bacteria, classified as *Caldicellulosiruptor*, the enzymes currently characterized in this system also appear to be free enzymes, but their modular organization is of a higher order (Daniel et al., 1996; Gibbs et al., 2000; Reeves et al., 2000). Many of the enzymes of this system are bifunctional in that they contain two separate catalytic modules in the same polypeptide chain (Fig. 19). As mentioned earlier, the appearance of two catalytic modules in the same enzyme would infer a distinctive synergistic action between the two. Thus, in CelA, the family-9 and -48 catalytic modules would be expected to work in concerted fashion on crystalline cellulose. In another type of enzyme, the family-10 xylanase and family-5 cellulase would likely be most effective on regions of the plant cell wall that are characterized by cellulose-xylan junctions. The diversity in the modular architecture of the family-10 xylanases is particularly striking, and the various combinations of this type of catalytic module are apparently important to the sustenance of the bacterium in its environment. One of these xylanases appears to be attached to the cell surface via SLH domains. In contrast to the *Cellulomonas* and *T. fusca* enzymes that often harbor a family-2 CBM, the module responsible for binding to cellulosic substrates in *Caldicellulosiruptor* enzymes is usually one or more copies of a family-3 CBM.

Other bacterial strains that include at least one free bifunctional enzyme in their enzyme systems are *Anaerocellum thermophilum*, *Bacillus stearothermophilus*, *Fibrobacter succinogenes*, *Prevotella ruminicola*, *Ruminococcus albus*, *Ruminococcus flavefaciens*, *Streptomyces chattanoogensis* and the thermophilic anaerobe NA10. Unlike the *Caldicellulosiruptor* system, most of the free bifunctional enzymes in the latter strains appear to be isolated cases in the given system, rather than being a common character of their enzymes.

Cellulosomal Systems

The inclusion of enzymes into a cellulosome via the noncatalytic scaffoldin subunit represents a higher level of organization. The association of complementary enzymes into a complex is considered to contribute sterically to their synergistic action on cellulose and other plant cell wall polysaccharides. As mentioned earlier, in the case of *Clostridium thermocellum*, *Acetivibrio cellulolyticus* and *Bacteroides cellulosolvens*, the cellulosomes appear to be attached to the cell surface. The cellulosomes of *C. cellulolyticum*, *C. cellulovorans* and *C. josui* may also be cell-associated, but if so, the lack of a scaffoldin-borne dockerin and reciprocal anchoring protein would suggest an alternative mechanism.

The cellulosomes of *C. cellulolyticum*, *C. cellulovorans* and *C. josui* are very similar. The genes encoding for many or most of the enzymes in all

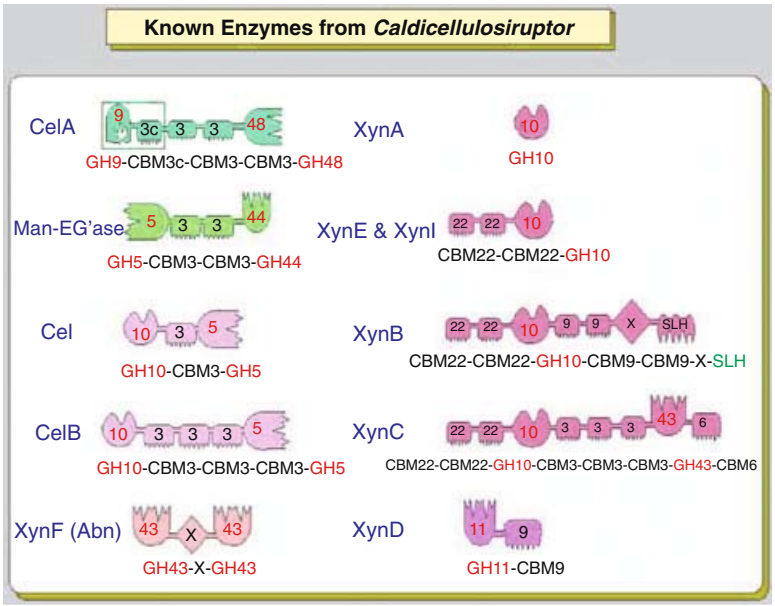

Fig. 19. Caldicellulosiruptor enzyme system. An example of a cell-free enzyme system that includes several multifunctional enzymes. The modular content of the enzymes is shown from (left to right) the *N*-terminus to the *C*-terminus of the polypeptide chain. Key to symbols: GH, glycosyl hydrolase (e.g., cellulase, xylanase and mannanase); CBM, carbohydrate-binding module (e.g., CBD, cellulose-binding domain); and SLH, S-layer homology (domain).

three cellulosomal systems are arranged in a large cluster on the chromosome. Some of the cellulosomal genes, however, are located outside of the cluster in other regions of the chromosome. The majority of the cellulosome gene clusters from *C. cellulolyticum* and *C. cellulovorans* have been sequenced (Bagnara-Tardif et al., 1992; Belaich et al., 1999; Tamaru et al., 2000b). In contrast, the cellulosomal genes from *C. thermocellum* are generally scattered over a large portion of the chromosome (Guglielmi and Béguin, 1998). A few small clusters of cellulosomal genes are apparent in the genome, including a scaffoldin-containing cluster that also contains several cell-surface anchoring proteins (Fujino et al., 1993). The following descriptive analysis serves to compare the cellulosomal system of these three microorganisms.

Cellulosomal components from *Clostridium cellulolyticum*. All of the sequenced enzymes from this organism are relatively common cellulases (Belaich et al., 1999). None of the known cellulosomal enzymes yet described for this species contains more than one catalytic module (Fig. 20). The largest one, CelE (estimated at 94 kDa), is a theme-D family-9 cellulase (Gaudin et al., 2000). The critical family-48 cellulase (CelF) is also a major cellulosome component (Reverbel-Leroy et al., 1997). Interestingly, the gene cluster of *C. cellulolyticum* contains three copies of other family-9 cellulases (CelG, CelH and CelJ), all of which contain the theme-B fused family-3c CBM (Belaich et al., 1998; Fig. 8). The currently known cellulosome system in this bacterium also contains two family-5 cellulases (CelA and CelD), a family-5 mannanase (ManK, which bears an *N*-terminal rather than *C*-terminal dockerin) and a family-8 cellulase (CelC).

Biochemical characterization of the *C. cellulolyticum* cellulosome demonstrated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) a 160-kDa scaffoldin band and up to 16 smaller bands, representing putative enzyme subunits (Gal et al., 1997). Many of these were clearly identified as known gene products. Only two cellulosomal cellulase genes are currently known to be located outside of the gene cluster. Further work on the enzyme system of this species may yet provide more complicated multimodular enzymes and/or other types of enzymes, such as hemicellulases. In this context, recent biochemical evidence has suggested that xylanases from *C. cellulolyticum* are also organized in a cellulosome-like complex, but defined xylanase sequences are still lacking from this organism (Mohand-Oussaid et al., 1999). The known activity of this organism on other plant cell wall

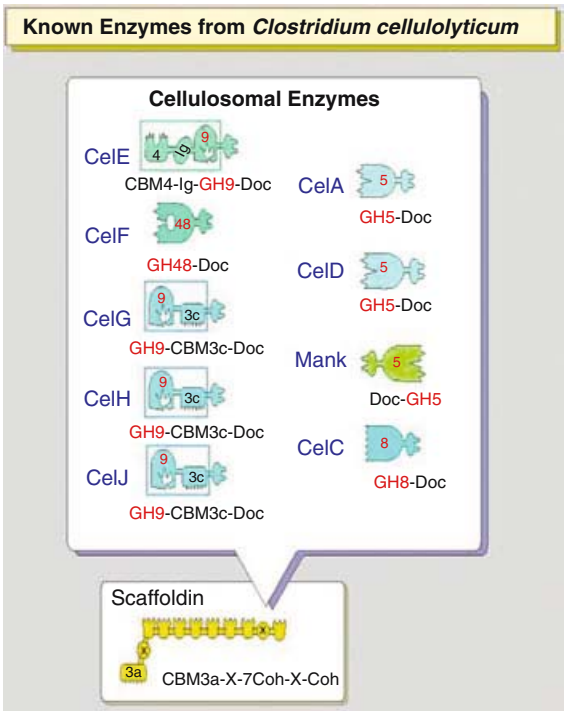

Fig. 20. *Clostridium cellulolyticum* enzyme system. An example of a cellulosomal system. The modular content of the enzymes is shown from (left to right) the *N*-terminus to the *C*-terminus of the polypeptide chain. Key to symbols: GH, glycosyl hydrolase (e.g., cellulase, xylanase and mannanase); CBM, carbohydrate-binding module (e.g., CBD, cellulose-binding domain); and Doc, dockerin domain.

polysaccharides would indicate that numerous other enzymes, either cellulosomal or not, remain as yet undiscovered.

Cellulosomal components from *Clostridium cellulovorans*. Like *C. cellulolyticum*, the cellulases from this organism are relatively simple (Fig. 21). In addition to the cellulosomal enzymes thus described, at least three non-cellulosomal endoglucanases have also been partially or totally sequenced (Doi et al., 1998; Tamaru et al., 1999b).

Several of the cellulosomal enzymes are architecturally synonymous to those of the *C. cellulolyticum* system (compare Figs. 20 and 21). This includes the critical family-48 cellulase (ExgS; Liu and Doi, 1998), two copies of the theme-B family-9 cellulase (EngH and EngY), a family-5 endoglucanase and a family-5 mannanase that bears an N-terminal dockerin (Tamaru and Doi, 2000a). Rather than a single theme-D family-9 cellulase as in *C. cellulolyticum*, the *C. cellulovorans* system contains two such enzymes (EngK and EngM). The *C. cellulovorans* cellulosome also appears to contain an unusual theme-A

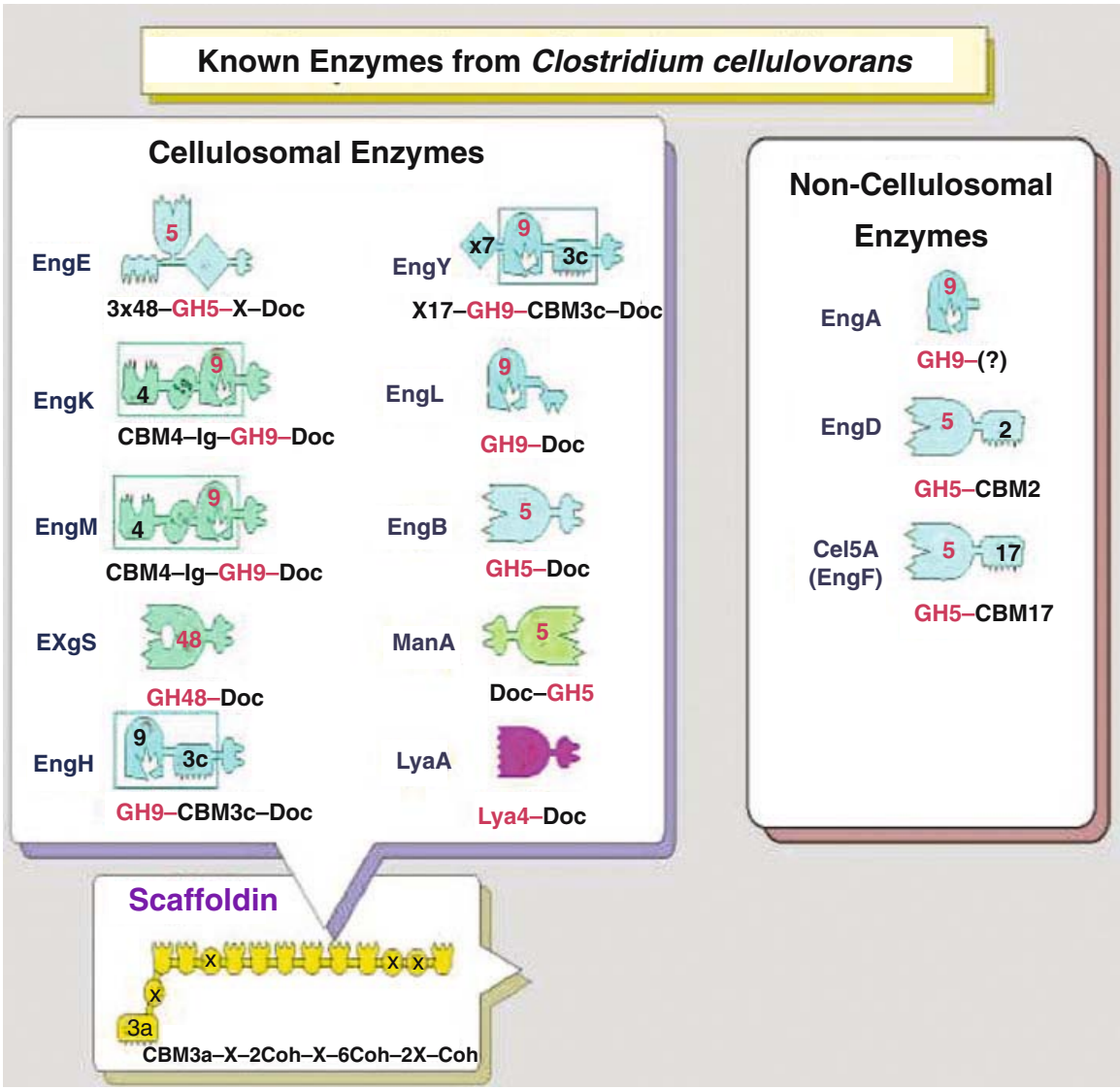

Fig. 21. *Clostridium cellulovorans*: A second cellulosomal system. The modular content of the enzymes is shown from (left to right) the *N*-terminus to the *C*-terminus of the polypeptide chain. Key to symbols: GH, glycosyl hydrolase (e.g., cellulase, xylanase and mannanase); CBM, carbohydrate-binding module (e.g., CBD, cellulose-binding domain); Doc, dockerin domain; SLH, S-layer homology (domain); Ig, immunoglobulin-like domain; and X, domain of unknown function.

family-9 cellulase (EngL) that lacks helper domains. The remaining two known cellulosomal enzymes are thus far unique to *C. cellulovorans*. A dockerin-bearing pectate lyase (LyaA) infers that the bacterium would degrade pectin (Tamaru and Doi, 2001). Indeed, early evidence (Sleat et al., 1984) indicated that, in addition to cellulose, *C. cellulovorans* is capable of assimilating a wide variety of other plant cell wall polysaccharides, including, xylans, pectins and mannans. As in the case of *C. celluolyticum*, it seems that future work will yield new sequences of many other types of cellulosomal and noncellulosomal enzymes.

More significant to the cellulosomal system of *C. cellulovorans*, perhaps, is the large family-5 enzyme that purportedly comprises both an *N*-terminal SLH domain and a *C*-terminal dockerin (Tamaru and Doi, 1999a). This arrangement may imply that the entire cellulosome is bound to the cell surface via this enzyme. If this proves to be the case, it is interesting to speculate whether the *C. cellulolyticum* and *C. josui* cellulosomes are also connected to the cell surface by a similar, but as yet undiscovered enzyme that bears both SLH and dockerin domains.

Cellulosomal components from *Clostridium thermocellum*. Compared to the cellulosomal

systems of *C. cellulovorans* and *C. cellulolyticum*, the enzymes from *C. thermocellum* are relatively large proteins, ranging in molecular size from about 40–180 kDa (Bayer et al., 1998; Bayer et al., 2000; Béguin and Lemaire, 1996; Felix and Ljungdahl, 1993; Lamed and Bayer, 1988; Shoham et al., 1999). Examination of Fig. 22 reveals why these enzymes are so big—many of the larger ones contain multiple types of catalytic domains as well as other functional modules as an integral part of a single polypeptide chain (see Table I in Bayer et al., 1998, for a list of relevant references). In addition to the cellulosomal enzymes, several noncellulosomal enzymes have also been described from this organism (Morag et al., 1990). These include two free enzymes (one of which lacks a CBM) and two cell-associated (SLH-containing) enzymes. Consequently, the potent cellulose- and plant cell wall-degrading activities of *C. thermocellum* are clearly reflected in its cellulase system, which displays an exceptional wealth, diversity and intricacy of enzymatic components, thus representing the premier cellulose-degrading organism currently known.

Many of the *C. thermocellum* cellulosomal enzymes are cellulases, which include both endo- and exo-acting β-glucanases. Some of the important exoglucanases and processive cellulases include CelS, CbhA, CelK and CelF. The CelS subunit is a member of the family-48 glycosyl hydrolases, and this particular family is now recognized as a critical component of bacterial cellulosomes (Morag et al., 1991; Morag et al., 1993; Wang et al., 1993; Wang et al., 1994; Wu et al., 1988). Several other processive cellulases are members of the family-9 glycosyl hydrolases. CelF and CelN are theme-B family-9 enzymes (Navarro et al., 1991; Fig. 7). The other two are remarkably similar theme-D enzymes, which exhibit nearly 95% similarity along their common regions (Kataeva et al., 1999a; Kataeva et al., 1999b; Zverlov et al., 1998; Zverlov et al., 1999). The main difference between CbhA and CelK is the presence in the former of three extra modules (a family-3 CBD and two modules of unknown function). The functional significance of these supplementary modules to the activity of CbhA has not been elucidated.

The fact that the cellulosome from this organism contains many different types of cellulases is, of course, to be expected if we consider that growth of *C. thermocellum* is restricted to cellulose and its breakdown products, particularly cellobiose. Consequently, it is surprising to discover, in addition to the cellulases, at least five classic xylanases, i.e., those belonging to glycosyl hydrolase families 10 and 11. In addition, two of the larger enzymes, CelH and CelJ, contain hemicellulase components, i.e., family-26 and -44 catalytic modules (a mannanase and a xylanase, respectively), together with a standard cellulase module in the same polypeptide chain (Ahsan et al., 1996; Yagüe et al., 1990). It is also interesting to note the presence of carbohydrate esterases together with xylanase or cellulase modules in some of the enzyme subunits (i.e., XynU/A, XynY, XynZ and CelE), thus conferring the capacity to hydrolyze acetyl or feruloyl groups from hemicellulose substrates (Blum et al., 2000; Fernandes et al., 1999). Finally, the *C. thermocellum* cellulosome includes a typical family-16 lichenase, a family-26 mannanase and a family-18 chitinase.

The non-cellulosomal enzymes include another theme-B family-9 cellulase (CelI), and cell-bound forms of a xylanase (XynX) and a lichenase (LicA), both of which contain multiple CBMs adjacent to the catalytic module. In the midst of all this complexity, the *C. thermocellum* non-cellulosomal cellulase system includes a simple family-5 cellulase, CelC, which is completely devoid of additional accessory modules.

Why does this bacterium—which subsists exclusively on cellulosic substrates—need all these hemicellulases? The inclusion of such an impressive array of non-cellulolytic enzymes in a strict cellulose-utilizing species would suggest that their major purpose would be to collectively purge the unwanted polysaccharides from the milieu and to expose the preferred substrate—cellulose. The ferulic acid esterases, in concert with the xylanase components of the parent enzymes, could grant the bacterium a relatively simple mechanism by which it could detach the lignin component from the cellulose-hemicellulose composite. The lichenase (LicB) and chitinase (ChiA) are also intriguing components of the cellulosome. The former would provide the bacterium with added action on cell-wall β-glucan components from certain types of plant matter. It is not clear whether the presence of the latter cellulosomal enzyme would reflect chitin-derived substrates from the exoskeletons of insects and/or from fungal cell walls. Whatever the source, the chitin breakdown products, like those of the hemicelluloses, would presumably not be utilized by the bacterium itself, but would be passed on to appropriate satellite bacteria for subsequent assimilation.

Phylogenetics of Cellulase and Cellulosomal Systems

Early in the history of the development and establishment of the cellulosome concept, it was noted that the apparent occurrence of cellulo-

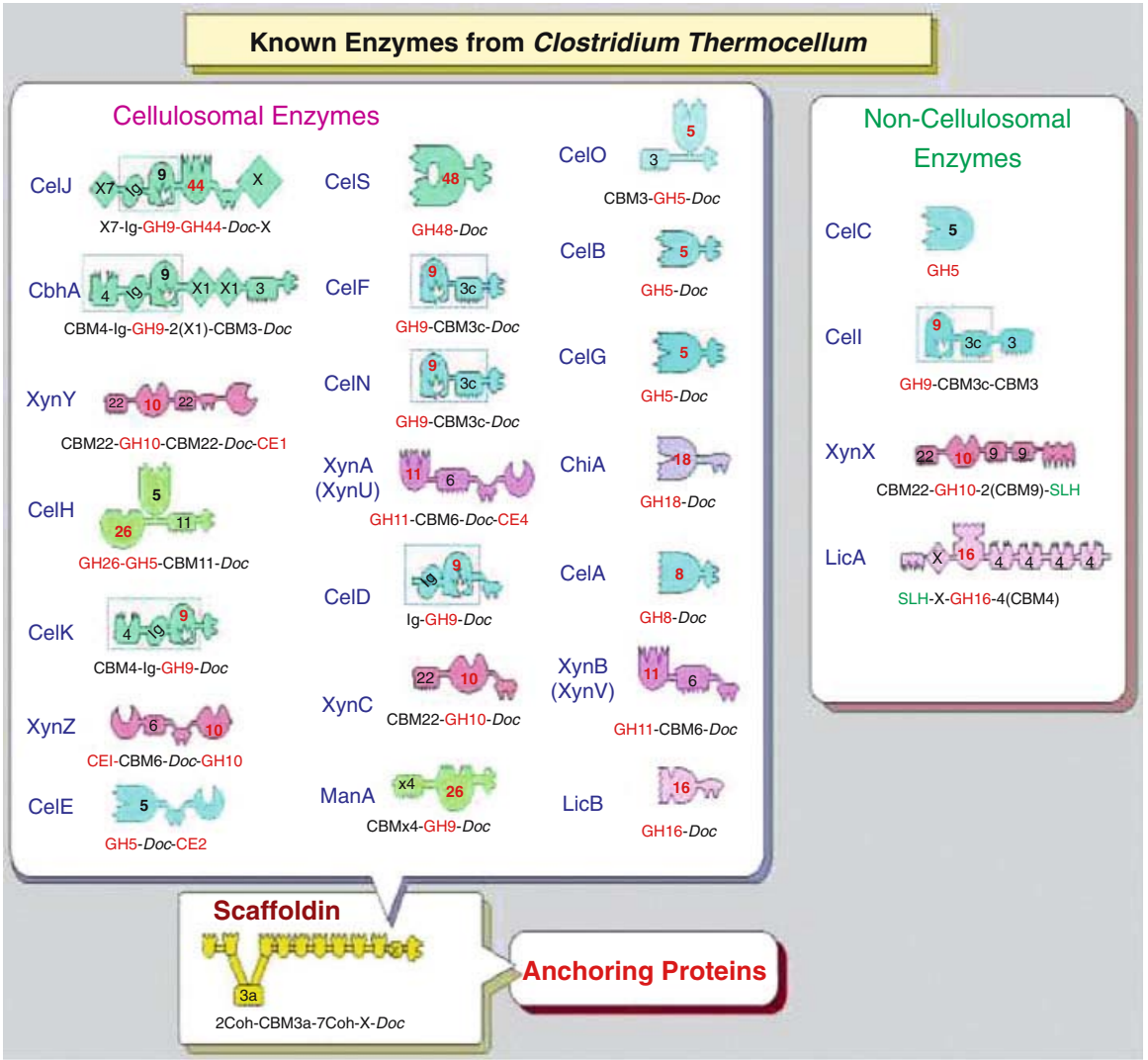

Fig. 22. *Clostridium thermocellum*: A very complex cellulosomal system. The modular content of the enzymes is shown from (left to right) the *N*-terminus to the *C*-terminus of the polypeptide chain. Key to symbols: GH, glycosyl hydrolase (e.g., cellulase, xylanase and mannanase); CE, carbohydrate esterase (e.g., acetyl xylan esterase and ferulic acid esterase); CBM, carbohydrate-binding module (e.g., CBD, cellulose-binding domain); Doc, dockerin domain; SLH, S-layer homology (domain); Ig, immunoglobulin-like domain; and X, domain of unknown function.

somes in different microorganisms tended to cross ecological, physiological and evolutionary boundaries (Lamed et al., 1987). Initial biochemical and immunochemical evidence to this effect has been supported by the accumulated molecular biological studies.

Various lines of evidence indicate that the modular enzymes that degrade plant cell wall polysaccharides have evolved from a restricted number of common ancestral sequences. Much of the information in this direction remains as a legacy, inherently encoded in the sequences of the functional domains that comprise the different enzymes. By comparing sequences of the various cellulosomal and noncellulosomal enzymes within and among the different strains, we can gain insight into the evolutionary rationale of the multigene families that comprise the glycosyl hydrolases.

Horizontal Gene Transfer

It is clear that very similar enzymes which comprise a given glycosyl hydrolase family are prevalent among a variety of different bacteria and fungi, thus indicating that they were not inherited through conventional evolutionary processes. The widespread occurrence of such conserved enzymes among phylogenetically different species argues that horizontal transfer of

genes has been a major process by which a given microorganism can acquire a desirable enzyme. Once such a transfer event has taken place, the newly acquired gene would then be subjected to environmental pressures of its new surroundings, i.e., the genetic and physiological constitution of the cell itself. Following such selective pressure, the sequence of the gene would be adjusted to fit the host cell.

Gene Duplication

Sequence comparisons have also revealed the presence of very similar genes within a genome that may have very similar or even identical functions. One striking example is the tandem appearance of *cbhA* and *celK* genes in the chromosome of *Clostridium thermocellum*. Other examples are *xynA* and *xynB* also of *C. thermocellum* and *xynA* of the anaerobic fungus *Neocallimastix patriciarum*, which includes two very similar copies of family-11 catalytic modules within the same polypeptide chain. These examples imply a mechanism of gene duplication (Chen et al., 1998; Gilbert et al., 1992), whereby the duplicated gene can serve as a template for secondary modifications that could result in two very similar enzymes with different properties, such as substrate and product specificities. A similar process could also account for the multiplicity of other types of modules (i.e., CBDs, cohesins or helper modules) within a polypeptide chain. Comparison of the modular architectures of similar genes from different species would suggest that individual modules can undergo a duplication process. This is exemplified by the multiple copies of FN3 in CelB from *Cellulomonas fimi* versus the single copy of the same domain in cellulase E4 from *Thermobifida fusca*. But innumerable other examples are evident from the databases, whenever multiple copies of the same modular type exist in the same protein.

Domain Shuffling

Another observation from the genetic composition of the glycosyl hydrolases argues for an alternative type of process, which would propagate new or modified types of enzymes. It is clear that many microbial enzyme systems contain individual hydrolases that carry very similar catalytic domains but include different types of accessory modules (Gilkes et al., 1991). An example that demonstrates this phenomenon is the observed species preference of otherwise very similar glycosyl hydrolases for a given family of crystalline cellulose-binding CBD, which is entirely independent of the type of catalytic module borne by the complete enzyme. In this context, as we have seen above, the free enzymes of some bacteria, such as *Cellulomonas fimi*, *Pseudomonas fluorescens* and *Thermomonospora fusca*, invariably include a family-2 CBD, irrespective of the type of catalytic domain. In contrast, those of other bacteria, e.g., *Bacillus subtilis*, *Caldocellum saccharolyticum*, *Erwinia carotovora* and various clostridia, appear to prefer family-3 CBDs. Moreover, the position of the CBD in the gene may be different for different genes. For example, the CBD may occur upstream or downstream from the catalytic domain; it may be positioned either internally (sandwiched between two other modules) or at one of the termini of the polypeptide chain. The same pattern is characteristic of several other kinds of modules associated with the plant cell wall hydrolases. This is particularly evident in family-9 cellulases and family-10 xylanases, where the number and types of accessory modules may vary greatly within a given species. It seems that individual domains can be transferred en bloc and incorporated independently into appropriate enzymes. Once again, the modular architectures and sequence similarities between *Clostridium thermocellum* cellulosomal enzyme pairs (CbhA and CelK; XynA and XynB) are particularly revealing: in both cases, following an apparent gene duplication event, one or more additional modules appear to have been incorporated into the duplicated enzyme. Taken together, the information suggests that domain shuffling is an important process by which the properties of such enzymes can be modified and extended.

Proposed Mechanisms for Acquiring Cellulase and Cellulosomal Genes

Like the free enzyme systems, the phylogeny of cellulosomal components seems to have been driven by processes that include horizontal gene transfer, gene duplication and domain shuffling. In cellulolytic/hemicellulolytic ecosystems, the resident microorganisms are usually in close contact, often under difficult conditions and in competition or cooperation with one another toward a common goal: the rapid degradation of recalcitrant polysaccharides and assimilation of their breakdown products.

A possible scenario for the molecular evolution of a cellulase/hemicellulase system in a prospective bacterium could involve the initial transfer of genetic material from one microbe to another in the same ecosystem. The size and type of transferred material could vary, such as a gene or part of gene (e.g., selected functional modules) or even all or part of a gene cluster. The process could then be sustained by gene duplica-

tion, which would propagate the insertion of repeated modules, e.g., the multiple cohesin domains in the scaffoldins, or even smaller units, such as the linker sequences or the duplicated calcium-binding loop of the dockerin domain. Domain shuffling can account for the observed permutations in the arrangement of domains in scaffoldin subunits from different species (Fig. 16). Finally, conventional mutagenesis would then render such products more suitable for the cellular environment or for interaction with other components of the cellulase system.

The available data suggest that there are no set of rules, which would, at this stage, enable us to anticipate the nature of a given cellulase system from a given microorganism. It seems that phylogenetically dissimilar organisms can possess similar types of cellulosomal or non-cellulosomal enzyme systems, whereas phylogenetically related organisms that inhabit similar niches may be characterized by different types of enzyme systems. It is clear that to shed further light on this apparent enigma, we require more information about more types of enzyme systems. In addition to more sequences and structures, we will need more information—biochemical, physiological and ecological—to sharpen existing notions regarding the enzymatic degradation of plant cell wall polysaccharides or to formulate new ones.

Acknowledgments. Grants from the Israel Science Foundation (administered by the Israel Academy of Sciences and Humanities, Jerusalem) and the Minerva Foundation (Germany) are sincerely appreciated.

Literature Cited

Ahsan, M. M., T. Kimura, S. Karita, K. Sakka, and K. Ohmiya. 1996. Cloning, DNA sequencing, and expression of the gene encoding Clostridium thermocellum cellulase CelJ, the largest catalytic component of the cellulosome. J. Bacteriol. 178:5732–5740.

Armand, S., S. Drouillard, M. Schulein, B. Henrissat, and H. Driguez. 1997. A bifunctionalized fluorogenic tetrasaccharide as a substrate to study cellulases. J. Biol. Chem. 272:2709–2713.

Atalla, R. H., and D. L. VanderHart. 1984. Native cellulose: A composite of two distinct crystalline forms. Science 223:283–285.

Atalla, R. H. 1999. Celluloses. *In:* B. M. Pinto (Ed.) Comprehensive Natural Products Chemistry. Elsevier. Cambridge, UK. 3:529–598.

Bach, E., and E. Schollmeyer. 1992. An ultraviolet-spectrophotometric method with 2-cyanoacetamide for the determination of the enzymatic degradation of reducing polysaccharides. Analyt. Biochem. 203(2):335–339.

Bagnara-Tardif, C., C. Gaudin, A. Belaich, P. Hoest, T. Citard, and J.-P. Belaich. 1992. Sequence analysis of a gene cluster encoding cellulases from Clostridium cellulolyticum. Gene 119:17–28.

Barr, B. K., Y.-L. Hsieh, B. Ganem, and D. B. Wilson. 1996. Identification of two functionally different classes of exocellulases. Biochemistry 35:586–592.

Bayer, E. A., R. Kenig, and R. Lamed. 1983. Adherence of Clostridium thermocellum to cellulose. J. Bacteriol. 156:818–827.

Bayer, E. A., E. Setter, and R. Lamed. 1985. Organization and distribution of the cellulosome in Clostridium thermocellum. J. Bacteriol. 163:552–559.

Bayer, E. A., and R. Lamed. 1986. Ultrastructure of the cell surface cellulosome of Clostridium thermocellum and its interaction with cellulose. J. Bacteriol. 167:828–836.

Bayer, E. A., and R. Lamed. 1992. The cellulose paradox: Pollutant par excellence and/or a reclaimable natural resource?. Biodegradation 3:171–188.

Bayer, E. A., E. Morag, and R. Lamed. 1994. The cellulosome—a treasure-trove for biotechnology. Trends Biotechnol. 12:378–386.

Bayer, E. A., E. Morag, Y. Shoham, J. Tormo, and R. Lamed. 1996. The cellulosome: A cell-surface organelle for the adhesion to and degradation of cellulose. *In:* M. Fletcher (Ed.) Bacterial Adhesion: Molecular and Ecological Diversity. Wiley-Liss. New York, NY. 155–182.

Bayer, E. A., H. Chanzy, R. Lamed, and Y. Shoham. 1998a. Cellulose, cellulases and cellulosomes. Curr. Opin. Struct. Biol. 8:548–557.

Bayer, E. A., E. Morag, R. Lamed, S. Yaron, and Y. Shoham. 1998b. Cellulosome structure: Four-pronged attack using biochemistry, molecular biology, crystallography and bioinformatics. *In:* M. Claeyssens, W. Nerinckx, and K. Piens (Eds.) Carbohydrases from Trichoderma reesei and Other Microorganisms. The Royal Society of Chemistry. London, 39–67.

Bayer, E. A., L. J. W. Shimon, R. Lamed, and Y. Shoham. 1998c. Cellulosomes: Structure and ultrastructure. J. Struct. Biol. 124:221–234.

Bayer, E. A., Y. Shoham, and R. Lamed. 2000. The cellulosome—an exocellular organelle for degrading plant cell wall polysaccharides. *In:* R. J. Doyle (Ed.) Glycomicrobiology. Kluwer Academic/Plenum Publishers. New York, NY. 387–439.

Béguin, P. 1983. Detection of cellulase activity in polyacrylamide gels using Congo red-stained agar replicas. Analyt. Biochem. 131:333–336.

Béguin, P. 1990. Molecular biology of cellulose degradation. Ann. Rev. Microbiol. 44:219–248.

Béguin, P., and J.-P. Aubert. 1994. The biological degradation of cellulose. FEMS Microbiol. Lett. 13:25–58.

Béguin, P., and M. Lemaire. 1996. The cellulosome: An exocellular, multiprotein complex specialized in cellulose degradation. Crit. Rev. Biochem. Molec. Biol. 31:201–236.

Belaich, J.-P., C. Tardif, A. Belaich, and C. Gaudin. 1997. The cellulolytic system of Clostridium cellulolyticum. J. Biotechnol. 57:3–14.

Belaich, A., J.-P. Belaich, H.-P. Fierobe, C. Gaudin, S. Pagès, C. Reverbel-Leroy, and C. Tardif. 1998. Cellulosome analysis and cellulases CelF and CelG from Clostridium cellulolyticum. *In:* M. Claeyssens, W. Nerinckx, and K. Piens (Eds.) Carbohydrases from Trichoderma reesei and Other Microorganisms. The Royal Society of Chemistry. London, 73–86.

Belaich, J.-P., A. Belaich, H.-P. Fierobe, L. Gal, C. Gaudin, S. Pagès, C. Reverbel-Leroy, and C. Tardif. 1999. The cellulolytic system of Clostridium cellulolyticum. *In:* K. Ohmiya, K. Hayashi, K. Sakka, Y. Kobayashi, S. Karita, and T. Kimura (Ed.) Genetics, Biochemistry and Ecology of Cellulose Degradation. Uni Publishers. Tokyo, 479–487.

Bhat, M. K. 2000. Cellulases and related enzymes in biotechnology. Biotechnol. Adv. 18:355–383.

Biely, P. 1985. Microbial xylanolytic systems. Trends Biotechnol. 3:285–290.

Blum, D. L., I. A. Kataeva, X. L. Li, and L. G. Ljungdahl. 2000. Feruloyl esterase activity of the Clostridium thermocellum cellulosome can be attributed to previously unknown domains of XynY and XynZ. J Bacteriol 182(5):1346–1351.

Boraston, A. B., B. W. McLean, J. M. Kormos, M. Alam, N. R. Gilkes, C. A. Haynes, P. Tomme, D. G. Kilburn, and R. A. Warren. 1999. Carbohydrate-binding modules: Diversity of structure and function. *In:* H. J. Gilbert, G. J. Davies, B. Henrissat, and B. Svensson (Eds.) Recent Advances in Carbohydrate Bioengineering. The Royal Society of Chemistry. Cambridge, 202–211.

Borneman, W. S., L. G. Ljungdahl, R. D. Hartley, and D. E. Akin. 1993. Feruloyl and p-coumaroyl esterases from the anaerobic fungus Neocallimastix strain MC-2 properties and functions in plant cell wall degradation. *In:* M. P. Coughlan and G. P. Hazlewood (Eds.) Hemicellulose and Hemicellulases. Portland Press. London, 85–102.

Brun, E., F. Moriaud, P. Gans, M. Blackledge, F. Barras, and D. Marion. 1997. Solution structure of the cellulose-binding domain of the endoglucanase Z secreted by Erwinia chrysanthemi. Biochemistry 36:16074–16086.

Carpita, N. C., and D. M. Gibeaut. 1993. Structural models of primary cell walls in flowering plants: Consistency of molecular structure with the physical properties of the walls during growth. Plant J. 3:1–30.

Chanzy, H. 1990. Aspects of cellulose structure. *In:* J. F. Kennedy, G. O. Philips, and P. A. Williams (Eds.) Cellulose Sources and Exploitation: Industrial Utilization, Biotechnology and Physico-chemical Properties. Ellis Horwood. New York, NY. 3–12.

Chauvaux, S., P. Béguin, J.-P. Aubert, K. M. Bhat, L. A. Gow, T. M. Wood, and A. Bairoch. 1990. Calcium-binding affinity and calcium-enhanced activity of Clostridium thermocellum endoglucanase D. Biochem. J. 265:261–265.

Chauvaux, S., M. Matuschek, and P. Béguin. 1999. Distinct affinity of binding sites for S-layer homologous domains in Clostridium thermocellum and Bacillus anthracis cell envelopes. J Bacteriol. 181:2455–2458.

Chen, H., X. L. Li, D. L. Blum, and L. G. Ljungdahl. 1998. Two genes of the anaerobic fungus Orpinomyces sp. strain PC-2 encoding cellulases with endoglucanase activities may have arisen by gene duplication. FEMS Microbiol. Lett. 159:63–68.

Claeyssens, M., and B. Henrissat. 1992. Specificity mapping of cellulolytic enzymes: Classification into families of structurally related proteins confirmed by biochemical analysis. Protein Sci. 1(10):1293–1297.

Coughlan, M. P., and F. Mayer. 1992. The cellulose-decomposing bacteria and their enzyme systems. *In:* A. Balows, H. G. Trüper, M. Dworkin, W. Harder, and K.-H. Schleifer (Eds.) The Prokaryotes, 2nd ed. Springer. New York, NY. 1:459–516.

Coughlan, M. P., and G. P. Hazlewood. 1993. β-1,4-D-xylan-degrading enzyme systems: Biochemistry, molecular biology and applications. Biotechnol. Appl. Biochem. 17:259–289.

Coutinho, P. M., and B. Henrissat. 1999a. Carbohydrate-active enzymes: An integrated database approach. *In:* H. J. Gilbert, G. J. Davies, B. Henrissat, and B. Svensson (Eds.) Recent Advances in Carbohydrate Bioengineering. The Royal Society of Chemistry. Cambridge, 3–12.

Coutinho, P. M., and B. Henrissat. 1999b. Carbohydrate-Active enZYmes and associated MODular Organization server. CAZyModO Website afmb.cnrs-mrs.fr/.

Coutinho, P. M., and B. Henrissat. 1999c. The modular structure of cellulases and other carbohydrate-active enzymes: An integrated database approach. *In:* K. Ohmiya, K. Hayashi, K. Sakka, Y. Kobayashi, S. Karita, and T. Kimura (Eds.) Genetics, Biochemistry and Ecology of Cellulose Degradation. Uni Publishers. Tokyo, 15–23.

Daniel, R. M., H. S. Toogood, and P. L. Bergquist. 1996. Thermostable proteases. Biotechnol. Genet. Engin. Rev. 13:51–100.

Das, N. N., S. C. Das, and A. K. Mukerjee. 1984. On the ester linkage between lignin and 4-O-methyl-D-glucurono-D-xylan in jute fiber (Corchorus capsularis). Carbohydr. Res. 127:345–348.

Davies, G., and B. Henrissat. 1995. Structures and mechanisms of glycosyl hydrolases. Structure 3:853–859.

Davies, G. J., L. Mackenzie, A. Varrot, M. Dauter, M. Brzozowski, M. Schülein, and S. G. Withers. 1998. Snapshots along an enzymatic reaction coordinate: Analysis of a retaining β-glycoside hydrolase. Biochemistry 37:11707–11713.

Ding, S.-Y., E. A. Bayer, D. Steiner, Y. Shoham, and R. Lamed. 1999. A novel cellulosomal scaffoldin from Acetivibrio cellulolyticus that contains a Family-9 glycosyl hydrolase. J. Bacteriol. 181:6720–6729.

Ding, S.-Y., E. A. Bayer, D. Steiner, Y. Shoham, and R. Lamed. 2000. A scaffoldin of the Bacteroides cellulosolvens cellulosome that contains 11 type II cohesins. J. Bacteriol. 182:4915–4925.

Ding, S.-Y., M. T. Rincon, R. Lamed, J. C. Martin, S. I. McCrae, V. Aurilia, Y. Shoham, E. A. Bayer, and H. J. Flint. 2001. Cellulosomal scaffoldin-like proteins from Ruminococcus flavefaciens. J. Bacteriol. 183:1945–1953.

Divne, C., J. Stahlberg, T. Teeri, and T. Jones. 1998. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from Trichoderma reesei. J. Molec. Biol. 275:309–325.

Doi, R. H., M. Goldstein, S. Hashida, J. S. Park, and M. Takagi. 1994. The Clostridium cellulovorans cellulosome. Crit. Rev. Microbiol. 20:87–93.

Doi, R. H., J. S. Park, C. C. Liu, L. M. Malburg, Y. Tamaru, A. Ichiishi, and A. Ibrahim. 1998. Cellulosome and noncellulosomal cellulases of Clostridium cellulovorans. Extremophiles 2:53–60.

Doi, R. H., and Y. Tamura. 2001. The Clostridium cellulovorans cellulosome: An enzyme complex with plant cell wall degrading activity. Chem. Rec. 1:24–32.

Doner, L. W., and P. L. Irwin. 1992. Assay of reducing end-groups in oligosaccaride homolgues with 2,2′-bicinchoninate. Analyt. Biochem. 202:50–53.

Driguez, H. 1997. Thiooligosaccharides in glycobiology. Topics Curr. Chem. 187:85–116.

Eriksson, K.-E. L., R. A. Blanchette, and P. Ander. 1990. Biodegradation of hemicelluloses. Microbial and Enzymatic Degradation of Wood and Wood Components. Springer-Verlag. Heidelberg, 181–397.

Felix, C. R., and L. G. Ljungdahl. 1993. The cellulosome—the exocellular organelle of Clostridium. Ann. Rev. Microbiol. 47:791–819.

Fernandes, A. C., C. M. Fontes, H. J. Gilbert, G. P. Hazlewood, T. H. Fernandes, and L. M. A. Ferreira. 1999. Homologous xylanases from Clostridium thermocellum: Evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. Biochem. J. 342:105–110.

Flint, H. J., J. Martin, C. A. McPherson, A. S. Daniel, and J. X. Zhang. 1993. A bifunctional enzyme, with separate xylanase and β(1,3-1,4)-glucanase domains, encoded by the xynD gene of Ruminococcus flavefaciens. J. Bacteriol. 175:2943–2951.

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

Gal, L., C. Gaudin, A. Belaich, S. Pagès, C. Tardif, and J.-P. Belaich. 1997a. CelG from Clostridium cellulolyticum: A multidomain endoglucanase acting efficiently on crystalline cellulose. J. Bacteriol. 179:6595–6601.

Gal, L., S. Pagès, C. Gaudin, A. Belaich, C. Reverbel-Leroy, C. Tardif, and J.-P. Belaich. 1997b. Characterization of the cellulolytic complex (cellulosome) produced by Clostridium cellulolyticum. Appl. Environ. Microbiol. 63:903–909.

Garcia, E., D. Johnston, J. R. Whitaker, and S. P. Shoemaker. 1993. Assessment of endo-1,4-beta-D-glucanase activity by a rapid colorimetric assay using disodium 2,2′-bicinchoninate. J. Food Biochem. 17:135–145.

Gaudin, C., A. Belaich, S. Champ, and J. P. Belaich. 2000. CelE, a multidomain cellulase from Clostridium cellulolyticum: A key enzyme in the cellulosome? J. Bacteriol. 182:1910–1915.

Ghose, T. K. 1987. Measurments of cellulase activity. Pure Appl. Chem. 59:257–268.

Gibbs, M. D., R. A. Reeves, G. K. Farrington, P. Anderson, D. P. Williams, and P. L. Bergquist. 2000. Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1. Curr. Microbiol. 40(5):333–40.

Gilbert, H. J., G. P. Hazlewood, J. I. Laurie, C. G. Orpin, and G. P. Xue. 1992. Homologous catalytic domains in a rumen fungal xylanase: Evidence for gene duplication and prokaryotic origin. Molec. Microbiol. 6:2065–2072.

Gilbert, H. J., and G. P. Hazlewood. 1993. Bacterial cellulases and xylanases. J. Gen. Microbiol. 139:187–194.

Gilkes, N. R., B. Henrissat, D. G. Kilburn, R. C. J. Miller, and R. A. J. Warren. 1991. Domains in microbial β-1,4-glycanases: Sequence conservation, function, and enzyme families. Microbiol. Rev. 55(2):303–315.

Green 3rd, F., C. A. Clausen, and T. L. Highley. 1989. Adaptation of the Nelson-Somogyi reducing-sugar assay to a microassay using microtiter plates. Analyt. Biochem. 182(2):197–9.

Guglielmi, G., and P. Béguin. 1998. Cellulase and hemicellulase genes of Clostridium thermocellum from five independent collections contain few overlaps and are widely scattered across the chromosome. FEMS Microbiol. Lett. 161:209–215.

Haigler, C. H., and P. J. Weimer. 1991. *In:* C. H. Haigler and P. J. Weimer (Eds.) Biosynthesis and Biodegradation of Cellulose. Marcel Dekker. New York, NY.

Hazlewood, G. P., and H. J. Gilbert. 1993. Molecular biology of hemicellulases. *In:* M. P. Coughlan and G. P. Hazlewood (Eds.) Hemicellulose and Hemicellulases. Portland Press. London, 103–126.

Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 280:309–316.

Henrissat, B., I. Callebaut, S. Fabrega, P. Lehn, J.-P. Mornon, and G. Davies. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc. Natl. Acad. Sci. USA 92:7090–7094.

Henrissat, B., and A. Bairoch. 1996. Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. 316:695–696.

Henrissat, B., and G. Davies. 1997. Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. 7:637–644.

Henrissat, B., T. T. Teeri, and R. A. J. Warren. 1998. A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. FEBS Lett. 425:352–354.

Higuchi, T. 1990. Lignin biochemistry, biosynthesis and biodegradation. Wood Sci. Technol. 24:23–63.

Hilden, L., L. Eng, G. Johansson, S. E. Lindqvist, and G. Pettersson. 2001. An amperometric cellobiose dehydrogenase-based biosensor can be used for measurement of cellulase activity. Analyt. Biochem. 290(2):245–50.

Himmel, M. E., M. F. Ruth, and C. E. Wyman. 1999. Cellulase for commodity products from cellulosic biomass. Curr. Opin. Biotechnol. 10(4):358–364.

Hurlbert, J. C., and J. F. Preston 3rd. 2001. Functional characterization of a novel xylanase from corn strain of Erwinia chrysanthemi. J. Bacteriol. 183:2093–2100.

Irwin, D., L. Walker, M. Spezio, and D. Wilson. 1993. Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. Biotech. Bioengin. 42:1002–1013.

Irwin, D., D.-H. Shin, S. Zhang, B. K. Barr, J. Sakon, P. A. Karplus, and D. B. Wilson. 1998. Roles of the catalytic domain and two cellulose binding domains of Thermomonospora fusca E4 in cellulose hydrolysis. J. Bacteriol. 180:1709–1714.

Johnson, P., M. Joshi, P. Tomme, D. Kilburn, and L. McIntosh. 1996. Structure of the N-terminal cellulose-binding domain of Cellulomonas fimi Cen C determined by nuclear magnetic resonance spectroscopy. Biochemistry 35:14381–14394.

Jung, K. H., K. M. Lee, H. Kim, K. H. Yoon, S. H. Park, and M. Y. Pack. 1998. Cloning and expression of a Clostridium thermocellum xylanase gene in Escherichia coli. Biochem. Molec. Biol. Intl. 44(2):283–292.

Juy, M., A. G. Amit, P. M. Alzari, R. J. Poljak, M. Claeyssens, P. Béguin, and J.-P. Aubert. 1992. Crystal structure of a thermostable bacterial cellulose-degrading enzyme. Nature 357:39–41.

Karita, S., K. Sakka, and K. Ohmiya. 1997. Cellulosomes, cellulase complexes, of anaerobic microbes: Their structure models and functions. *In:* R. Onodera, H. Itabashi, K. Ushida, H. Yano, and Y. Sasaki (Eds.) Rumen

Microbes and Digestive Physiology in Ruminants. Japanese Scientific Society Press and S. Karger. Tokyo/Basle, Germany. 14:47–57.

Kataeva, I., X.-L. Li, H. Chen, S. K. Choi, and L. G. Ljungdahl. 1999a. Cloning and sequence analysis of a new cellulase gene encoding CelK, a major cellulosome component of Clostridium thermocellum: Evidence for gene duplication and recombination. J. Bacteriol. 181:5288–5295.

Kataeva, I., X.-L. Li, H. Chen, and L. G. Ljungdahl. 1999b. CelK—a new cellobiohydrolase from Clostridium thermocellum cellulosome: Role of N-terminal cellulose-binding domain. *In:* K. Ohmiya, K. Hayashi, K. Sakka, Y. Kobayashi, S. Karita, and T. Kimura (Eds.) Genetics, Biochemistry and Ecology of Cellulose Degradation. Uni Publishers. Tokyo, 454–460.

Kidby, D. K., and D. J. Davidson. 1973. A convenient ferricyanide estimation of reducing sugars in the nanomole range. Analyt. Biochem. 55(1):321–325.

Kuhad, R. C., A. Singh, and K.-E. Eriksson. 1997. Microorganisms and enzymes involved in the degradation of plant fiber cell walls. Adv. Biochem. Engin. 57:45–125.

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

Lamed, R., E. Setter, R. Kenig, and E. A. Bayer. 1983b. The cellulosome—a discrete cell surface organelle of Clostridium thermocellum which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. Biotechnol. Bioeng. Symp. 13:163–181.

Lamed, E., J. Naimark, E. Morgenstern, and E. A. Bayer. 1987a. Scanning electron microscopic delineation of bacterial surface topology using cationized ferritin. J. Microbiol. Methods 7:233–240.

Lamed, R., J. Naimark, E. Morgenstern, and E. A. Bayer. 1987b. Specialized cell surface structures in cellulolytic bacteria. J. Bacteriol. 169:3792–3800.

Lamed, R., and E. A. Bayer. 1988a. The cellulosome concept: Exocellular/extracellular enzyme reactor centers for efficient binding and cellulolysis. *In:* J.-P. Aubert, P. Beguin, and J. Millet (Eds.) Biochemistry and Genetics of Cellulose Degradation. Academic Press. London, 101–116.

Lamed, R., and E. A. Bayer. 1988b. The cellulosome of Clostridium thermocellum. Adv. Appl. Microbiol. 33:1–46.

Lamed, R., and E. A. Bayer. 1991. Cellulose degradation by thermophilic anaerobic bacteria. *In:* C. H. Haigler and P. J. Weimer (Eds.) Biosynthesis and Biodegradation of Cellulose and Cellulose Materials. Marcel Dekker. New York, NY. 377–410.

Lamed, R., and E. A. Bayer. 1993. The cellulosome concept—a decade later!. *In:* K. Shimada, S. Hoshino, K. Ohmiya, K. Sakka, Y. Kobayashi, and S. Karita (Eds.) Genetics, Biochemistry and Ecology of Lignocellulose Degradation. Uni Publishers. Tokyo, 1–12.

Laurie, J. I., J. H. Clarke, A. Ciruela, C. B. Faulds, G. Williamson, H. J. Gilbert, J. E. Rixon, J. Millward-Sadler, and G. P. Hazlewood. 1997. The NodB domain of a multidomain xylanase from Cellulomonas fimi deacylates acetylxylan. FEMS Microbiol. Lett. 148:261–264.

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

Leibovitz, E., H. Ohayon, P. Gounon, and P. Béguin. 1997. Characterization and subcellular localization of the Clostridium thermocellum scaffoldin dockerin binding protein SdbA. J. Bacteriol. 179:2519–2523.

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

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

Lewis, N. G., and E. Yamamoto. 1990. Lignin: Occurrence, biogenesis and biodegradation. Ann. Rev. Plant Physiol. Plant Molec. Biol. 41:455–496.

Linder, M., and T. T. Teeri. 1997. The roles and function of cellulose-binding domains. J. Biotechnol. 57:15–28.

Liu, C. C., and R. H. Doi. 1998. Properties of exgS, a gene for a major subunit of the Clostridium cellulovorans cellulosome. Gene 211:39–47.

Ljungdahl, L. G., and K.-E. Eriksson. 1985. Ecology of microbial cellulose degradation. Adv. Microb. Ecol. 8:237–299.

Lupas, A., H. Engelhardt, J. Peters, U. Santarius, S. Volker, and W. Baumeister. 1994. Domain structure of the Acetogenium kivui surface layer revealed by electron crystallography and sequence analysis. J. Bacteriol. 176:1224–1233.

Ly, H. D., and S. G. Withers. 1999. Mutagenesis of glycosidases. Ann. Rev. Biochem. 68:487–522.

Lynd, L. R., J. H. Cushman, R. J. Nichols, and C. E. Wyman. 1991. Fuel ethanol from cellulosic biomass. Science 251:1318–1323.

Lytle, B., B. F. Volkman, W. M. Westler, and J. H. D. Wu. 2000. Secondary structure and calcium-induced folding of the Clostridium thermocellum dockerin domain determined by NMR spectroscopy. Arch. Biochem. Biophys. 379:237–244.

Lytle, B. L., B. F. Volkman, W. M. Westler, M. P. Heckman, and J. H. Wu. 2001. Solution structure of a type i dockerin domain, a novel prokaryotic, extracellular calcium-binding domain. J. Molec. Biol. 307(3):745–753.

Marais, J. P., J. L. De Wit, and G. V. Quicke. 1966. A critical examination of the Nelson-Somogyi method for the determination of reducing sugars. Analyt. Biochem. 15(3):373–381.

Mattinen, M.-L., M. Kontteli, J. Kerovuo, M. Linder, A. Annila, G. Lindeberg, T. Reinikainen, and T. Drakenberg. 1997. Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from Trichoderma reesei. Protein Sci. 6:294–303.

Matuschek, M., K. Sahm, A. Zibat, and H. Bahl. 1996. Characterization of genes from Thermoanaerobacterium thermosulfurigenes EM1 that encode two glycosyl hydrolases with conserved S-layer-like domains. Molec. Gen. Genet. 252(4):493–496.

Mayer, F., M. P. Coughlan, Y. Mori, and L. G. Ljungdahl. 1987. Macromolecular organization of the cellulolytic enzyme complex of Clostridium thermocellum as revealed by electron microscopy. Appl. Environ. Microbiol. 53:2785–2792.

McCarter, J. D., and S. G. Withers. 1994. Mechanisms of enzymatic glycoside hydrolysis. Curr. Opin. Struct. Biol. 4:885–892.

Mechaly, A., S. Yaron, R. Lamed, H.-P. Fierobe, A. Belaich, J.-P. Belaich, Y. Shoham, and E. A. Bayer. 2000. Cohesin-dockerin recognition in cellulosome assembly: Experiment versus hypothesis. Proteins 39:170–177.

Mechaly, A., H.-P. Fierobe, A. Belaich, J.-P. Belaich, R. Lamed, Y. Shoham, and E. A. Bayer. 2001. Cohesin-dockerin interaction in cellulosome assembly: A single hydroxyl group of a dockerin domain distinguishes between non-recognition and high-affinity recognition. J. Biol. Chem. 276:9883–9888.

Miller, G. L. R., W. E. Blum, and A. L. Burton. 1960. Measurements of carboxymethylcellulase activity. Analyt. Biochem. 2:127–132.

Mohand-Oussaid, O., S. Payot, E. Guedon, E. Gelhaye, A. Youyou, and H. Petitdemange. 1999. The extracellular xylan degradative system in Clostridium cellulolyticum cultivated on xylan: Evidence for cell-free cellulosome production. J. Bacteriol. 181:4035–4040.

Morag, E., E. A. Bayer, and R. Lamed. 1990. Relationship of cellulosomal and noncellulosomal xylanases of Clostridium thermocellum to cellulose-degrading enzymes. J. Bacteriol. 172:6098–6105.

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

Morag, E., E. A. Bayer, G. P. Hazlewood, H. J. Gilbert, and R. Lamed. 1993. Cellulase S_S (CelS) is synonymous with the major cellobiohydrolase (subunit S8) from the cellulosome of Clostridium thermocellum. Appl. Biochem. Biotechnol. 43:147–151.

Morag, E., A. Lapidot, D. Govorko, R. Lamed, M. Wilchek, E. A. Bayer, and Y. Shoham. 1995. Expression, purification and characterization of the cellulose-binding domain of the scaffoldin subunit from the cellulosome of Clostridium thermocellum. Appl. Environ. Microbiol. 61:1980–1986.

Navarro, A., M.-C. Chebrou, P. Béguin, and J.-P. Aubert. 1991. Nucleotide sequence of the cellulase gene celF of Clostridium thermocellum. Res. Microbiol. 142:927–936.

Notenboom, V., C. Birsan, R. Warren, S. Withers, and D. Rose. 1998. Exploring the cellulose/xylan specificity of the β-1,4-glycanase Cex from Cellulomonas fimi through crystallography and mutation. Biochemistry 37:4751–4758.

Ohmiya, K., K. Sakka, S. Karita, and T. Kimura. 1997. Structure of cellulases and their applications. Biotechnol. Genet. Engin. Rev. 14:365–414.

O'Neill, G., S. H. Goh, R. A. Warren, D. G. Kilburn, and R. C. Miller. 1986. Structure of the gene encoding the exoglucanase of Cellulomonas fimi. Gene 44(2–3):325–30.

O'Neill, R. A., A. Darvill, and P. Albersheim. 1989. A fluorescence assay for enzymes that cleave glycosidic linkages to produce reducing sugars. Analyt. Biochem. 177(1):11–5.

O'Sullivan, A. C. 1997. Cellulose: The structure slowly unravels. Cellulose 4:173–207.

Pagès, S., A. Belaich, J.-P. Belaich, E. Morag, R. Lamed, Y. Shoham, and E. A. Bayer. 1997. Species-specificity of the cohesin-dockerin interaction between Clostridium thermocellum and Clostridium cellulolyticum: Prediction of specificity determinants of the dockerin domain. Proteins 29:517–527.

Parsiegla, G., M. Juy, C. Reverbel-Leroy, C. Tardif, J. P. Belaich, H. Driguez, and R. Haser. 1998. The crystal structure of the processive endocellulase CelF of Clostridium cellulolyticum in complex with a thiooligosaccharide inhibitor at 2.0 Å resolution. EMBO J. 17:5551–5562.

Pegden, R. S., M. A. Larson, R. J. Grant, and M. Morrison. 1998. Adherence of the Gram-positive bacterium Ruminococcus albus to cellulose and identification of a novel form of cellulose-binding protein which belongs to the Pil family of proteins. J. Bacteriol. 180:5921–5927.

Puls, J., and J. Schuseil. 1993. Chemistry of hemicellulases: Relationship between hemicellulose structure and enzymes required for hydrolysis. *In:* M. P. Coughlan and G. P. Hazlewood (Eds.) Hemicellulose and Hemicellulases. Portland Press. London, 1–27.

Reese, R. T. 1976. History of the cellulase program at the U.S. Army Natick Development Center. Biotechnol. Bioeng. Symp. 6:9–20.

Reeves, R. A., M. D. Gibbs, D. D. Morris, K. R. Griffiths, D. J. Saul, and P. L. Bergquist. 2000. Sequencing and expression of additional xylanase genes from the hyperthermophile Thermotoga maritima FjSS3B.1. Appl. Environ. Microbiol. 66(4):1532–1537.

Reverbel-Leroy, C., S. Pagés, A. Belaich, J.-P. Belaich, and C. Tardif. 1997. The processive endocellulase CelF, a major component of the Clostridium cellulolyticum cellulosome: Purification and characterization of the recombinant form. J. Bacteriol. 179:46–52.

Rouvinen, J., T. Bergfors, T. Teeri, J. K. C. Knowles, and T. A. Jones. 1990. Three-dimensional structure of cellobiohydrolase II from Trichoderma reesei. Science 279:380–386.

Rydlund, A., and O. Dahlman. 1997. Oligosaccharides obtained by enzymatic hydrolysis of birch kraft pulp xylan: Analysis by capillary zone electrophoresis and mass spectrometry. Carbohydr. Res. 300:95–102.

Rye, C. S., and S. G. Withers. 2000. Glycosidase mechanisms. Curr. Opin. Chem. Biol. 4:573–580.

Sakon, J., D. Irwin, D. B. Wilson, and P. A. Karplus. 1997. Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca. Nature Struct. Biol. 4:810–818.

Salamitou, S., M. Lemaire, T. Fujino, H. Ohayon, P. Gounon, P. Béguin, and J.-P. Aubert. 1994a. Subcellular localization of Clostridium thermocellum ORF3p, a protein carrying a receptor for the docking sequence borne by the catalytic components of the cellulosome. J. Bacteriol. 176:2828–2834.

Salamitou, S., O. Raynaud, M. Lemaire, M. Coughlan, P. Béguin, and J.-P. Aubert. 1994b. Recognition specificity of the duplicated segments present in Clostridium thermocellum endoglucanase CelD and in the cellulosome-integrating protein CipA. J. Bacteriol. 176:2822–2827.

Saul D. J., L. C. Williams, R. A. Grayling, L. W. Chamley, D. R. Love, and P. L. Bergquist. 1990. celB, a gene coding for a bifunctional cellulase from the extreme thermophile "Caldocellum saccharolyticum". Appl. Environ. Microbiol. 56(10):3117–3124.

Schülein, M. 1997. Enzymatic properties of cellulases from Humicola insolens. J. Biotechnol. 57:71–81.

Shen, H., A. Meinke, P. Tomme, H. G. Damude, E. Kwan, D. G. Kilburn, R. C. Miller Jr., R. A. J. Warren, and N. R. Gilkes. 1995. Cellulomonas fimi cellobiohydrolases. *In:* J. N. Saddler and M. H. Penner (Eds.) Enzymatic Degradation of Insoluble Polysaccharides. American Chemical Society. Washington DC, 174–196.

Shimon, L. J. W., E. A. Bayer, E. Morag, R. Lamed, S. Yaron, Y. Shoham, and F. Frolow. 1997. The crystal structure at 2.15 Å resolution of a cohesin domain of the cellulosome from Clostridium thermocellum. Structure 5:381–390.

Shoham, Y., R. Lamed, and E. A. Bayer. 1999. The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. Trends Microbiol. 7:275–281.

Simpson, H. D., and F. Barras. 1999a. Functional analysis of the carbohydrate-binding domains of Erwinia chrysanthemi Cel5 (Endoglucanase Z) and an Escherichia coli putative chitinase. J Bacteriol. 181(15):4611–4616.

Simpson, P. J., D. N. Bolam, A. Cooper, A. Ciruela, G. P. Hazlewood, H. J. Gilbert, and M. P. Williamson. 1999b. A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity. Structure Fold. Des. 7:853–864.

Sinnott, M. L. 1990. Catalytic mechanisms of enzymic glycosyl transfer. Chem. Rev. 90:1171–1202.

Sleat, R., R. A. Mah, and R. Robinson. 1984. Isolation and characterization of an anaerobic, cellulolytic bacterium, Clostridium cellulovorans, sp. nov. Appl. Environ. Microbiol. 48:88–93.

Spinelli, S., H. P. Fierobe, A. Belaich, J. P. Belaich, B. Henrissat, and C. Cambillau. 2000. Crystal structure of a cohesin module from Clostridium cellulolyticum: Implications for dockerin recognition. J. Molec. Biol. 304(2):189–200.

Srisodsuk, M., K. Kleman-Leyer, S. Keranen, T. K. Kirk, and T. T. Teeri. 1998. Modes of action on cotton and bacterial cellulose of a homologous endoglucanase-exoglucanase pair from Trichoderma reesei. Eur. J. Biochem. 251(3):885–892.

Stutzenberger, F. 1990. Bacterial cellulases. *In:* W. M. Fogarty and C. T. Kelly (Eds.) Microbial Enzymes and Biotechnology. Elsevier Applied Science. London, New York, 37–70.

Tamaru, Y., and R. H. Doi. 1999a. Three surface layer homology domains at the N terminus of the Clostridium cellulovorans major cellulosomal subunit EngE. J. Bacteriol. 181:3270–3276.

Tamaru, Y., C.-C. Liu, A. Ichi-ishi, L. Malburg, and R. H. Doi. 1999b. The Clostridium cellulovorans cellulosome and non-cellulosomal cellulases. *In:* K. Ohmiya, K. Hayashi, K. Sakka, Y. Kobayashi, S. Karita, and T. Kimura (Eds.) Genetics, Biochemistry and Ecology of Cellulose Degradation. Uni Publishers. Tokyo, 488–494.

Tamaru, Y., and R. H. Doi. 2000a. The engL gene cluster of Clostridium cellulovorans contains a gene for cellulosomal ManA. J. Bacteriol. 182:244–247.

Tamaru, Y., S. Karita, A. Ibrahim, H. Chan, and R. H. Doi. 2000b. A large gene cluster for the Clostridium cellulovorans cellulosome. J Bacteriol. 182(20):5906–5910.

Tamaru, Y., and R. H. Doi. 2001. Pectate lyase A, an enzymatic subunit of the Clostridium cellulovorans cellulosome. Proc. Natl. Acad. Sci. USA 20:20.

Tavares, G. A., P. Béguin, and P. M. Alzari. 1997. The crystal structure of a type I cohesin domain at 1.7 Å resolution. J. Molec. Biol. 273:701–713.

Teeri, T. T., T. Reinikainen, L. Ruohonen, T. A. Jones, and J. K. C. Knowles. 1992. Domain function in Trichoderma reesei cellulases. J. Biotechnol. 24:169–176.

Teeri, T. T. 1997. Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. Trends Biotechnol. 15:160–167.

Te'o, V. S., D. J. Saul, and P. L. Bergquist. 1995. CelA, another gene coding for a multidomain cellulase from the extreme thermophile Caldocellum saccharolyticum. Appl. Microbiol. Biotechnol. 43:291–296.

Timell, T. E. 1967. Recent progress in the chemistry of wood hemicelluloses. Wood Sci. Technol. 1:45–70.

Tokatlidis, K., S. Salamitou, P. Béguin, P. Dhurjati, and J.-P. Aubert. 1991. Interaction of the duplicated segment carried by Clostridium thermocellum cellulases with cellulosome components. FEBS Lett. 291:185–188.

Tokatlidis, K., P. Dhurjati, and P. Béguin. 1993. Properties conferred on Clostridium thermocellum endoglucanase CelC by grafting the duplicated segment of endoglucanase CelD. Protein Engin. 6(8):947–952.

Tomme, P., R. A. J. Warren, and N. R. Gilkes. 1995a. Cellulose hydrolysis by bacteria and fungi. Adv. Microb. Physiol. 37:1–81.

Tomme, P., R. A. J. Warren, R. C. Miller, D. G. Kilburn, and N. R. Gilkes. 1995b. Cellulose-binding domains: Classification and properties. *In:* J. M. Saddler and M. H. Penner (Eds.) Enzymatic Degradation of Insoluble Polysaccharides. American Chemical Society. Washington DC, 142–161.

Tomme, P., A. L. Creagh, D. G. Kilburn, and C. A. Haynes. 1996. Interaction of polysaccharides with the N-terminal cellulose-binding domain of Cellulomonas fimi CenC. Biochemistry 35:13885–13894.

Tormo, J., R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham, and T. A. Steitz. 1996. Crystal structure of a bacterial family-III cellulose-binding domain: A general mechanism for attachment to cellulose. EMBO J. 15:5739–5751.

Tull, D., and S. G. Withers. 1994. Mechanisms of cellulases and xylanases: A detailed kinetic study of the exo-beta-1,4-glycanase from Cellulomonas fimi. Biochemistry 33(20):6363–6370.

van Tilbeurgh, H., G. Pettersson, R. Bhikabhai, H. De Boeck, and M. Claeyssens. 1985. Studies of the cellulolytic system of Trichoderma reesei QM 9414. Reaction specificity and thermodynamics of interactions of small substrates and ligands with the 1,4-beta-glucan cellobiohydrolase II. Eur. J. Biochem. 148(2):329–34.

Viikari, L., and T. Teeri. 1997. *In:* L. Viikari and T. Teeri (Eds.) Biochemistry and Genetics of Cellulases and Hemicellulases and Their Application.

Vlasenko, E. Y., A. I. Ryan, C. F. Shoemaker, and S. P. Shoemaker. 1998. The use of capillary viscometry, reducing end-group analysis, and size exclusion chromatography combined with multi-angle laser light scattering to characterize endo-1,4-β-D-glucanases on carboxymethylcellulose: A comparative evaluation of three methods. Enzyme Microb. Technol. 23:350–359.

Waffenschmidt, S., and L. Jaenicke. 1987. Assay of reducing sugars in the nanomole range with 2,2′-bicinchoninate. Analyt. Biochem. 165:337–340.

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

Wang, W. K., K. Kruus, and J. H. D. Wu. 1994. Cloning and expression of the Clostridium thermocellum cellulase

celS gene in Escherichia coli. Appl. Microbiol. Biotechnol. 42:346–352.

Warren, R. A. J. 1996. Microbial hydrolysis of polysaccharides. Ann. Rev. Microbiol. 50:183–212.

White, A., and D. R. Rose. 1997. Mechanism of catalysis by retaining β-glycosyl hydrolases. Curr. Opin. Struct. Biol. 7:645–651.

Whittle, D. J., D. G. Kilburn, R. A. Warren, and R. C. Miller. 1982. Molecular cloning of a Cellulomonas fimi cellulose gene in Escherichia coli. Gene 17(2):139–145.

Williams, S. J., and S. G. Withers. 2000. Glycosyl fluorides in enzymatic reactions. Carbohydr. Res. 327:27–46.

Williamson, M. P., P. J. Simpson, D. N. Bolam, G. P. Hazlewood, A. Ciruela, A. Cooper, and H. J. Gilbert. 1999. How the N-terminal xylan-binding domain from C. fimi xylanase D recognises xylan. *In:* H. J. Gilbert, G. J. Davies, B. Henrissat, and B. Svensson (Eds.) Recent Advances in Carbohydrate Bioengineering. The Royal Society of Chemistry. Cambridge, 212–220.

Wilson, D. B. 1992. Biochemistry and genetics of actinomycete cellulases. Crit. Rev. Biotechnol. 12:45–63.

Wilson, D. B., and D. C. Irwin. 1999. Genetics and properties of cellulases. Adv. Biochem. Engin. 65:1–21.

Withers, S. G., and R. Aebersold. 1995. Approaches to labeling and identification of active site residues in glycosidases. Protein Sci. 4(3):361–372.

Withers, S. G. 2001. Mechanisms of glycosyl transferases and hydrolases. Carbohydr. Res. 44:325–337.

Wood, W. A., and S. T. Kellogg. 1988. *In:* W. A. Wood and S. T. Kellogg (Eds.) Biomass. Part A: Cellulose and Hemicellulose. Academic Press. San Diego, CA. 160.

Wu, J. H. D., W. H. Orme-Johnson, and A. L. Demain. 1988. Two components of an extracellular protein aggregate of Clostridium thermocellum together degrade crystaline cellulose. Biochemistry 27:1703–1709.

Yagüe, E., P. Béguin, and J.-P. Aubert. 1990. Nucleotide sequence and deletion analysis of the cellulase-encoding gene celH of Clostridium thermocellum. Gene 89:61–67.

Yaron, S., E. Morag, E. A. Bayer, R. Lamed, and Y. Shoham. 1995. Expression, purification and subunit-binding properties of cohesins 2 and 3 of the Clostridium thermocellum cellulosome. FEBS Lett. 360:121–124.

Zechel, D. L., and S. G. Withers. 2000. Glycosidase mechanisms: Anatomy of a finely tuned catalyst. Acc. Chem. Res. 33(1):11–18.

Zou, J., G. J. Kleywegt, J. Stahlberg, H. Driguez, W. Nerinckx, M. Claeyssens, A. Koivula, T. T. Teeri, and T. A. Jones. 1999. Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Ce16A from trichoderma reesei. Struct. Fold. Des. 7(9):1035–1045.

Zverlov, V. V., S. Mahr, K. Riedel, and K. Bronnenmeier. 1998a. 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.

Zverlov, V. V., G. V. Velikodvorskaya, W. H. Schwarz, K. Bronnenmeier, J. Kellermann, and W. L. Staudenbauer. 1998b. Multidomain structure and cellulosomal localization of the Clostridium thermocellum cellobiohydrolase CbhA. J. Bacteriol. 180:3091–3099.

Zverlov, V. V., G. V. Velikodvorskaya, W. H. Schwarz, J. Kellermann, and W. L. Staudenbauer. 1999. Duplicated Clostridium thermocellum cellobiohydrolase gene encoding cellulosomal subunits S3 and S5. Appl. Microbiol. Biotechnol. 51:852–859.

Zverlov, V. V., I. Y. Volkov, G. A. Velikodvorskaya, and W. H. Schwarz. 2001. The binding pattern of two carbohydrate-binding modules of laminarinase Lam16A from Thermotoga neapolitana: Differences in beta-glucan binding within family CBM4. Microbiology 147(3):621–629.