Review Article

Bacterial cellulases and xylanases

HARRY J. GILBERT¹* and GEOFFREY P. HAZLEWOOD²

¹*Department of Biological and Nutritional Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK*

²*Department of Biochemistry, AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, UK*

Why study cellulases and xylanases?

Although this review focusses primarily on bacterial cellulases and xylanases, there is considerable overlap in the molecular biology and biochemistry of prokaryotic and fungal forms of these enzymes. Therefore, fungal plant cell wall hydrolases are discussed where comparisons with the corresponding bacterial enzymes are important. Cellulases and xylanases encompass a collection of enzymes whose primary function is to hydrolyse β -1,4-glycosidic linkages in the major plant structural polysaccharides, cellulose and xylan. In converting cellulose and xylan to their constituent sugars, these enzymes play an essential role in the digestive processes of herbivores and in the recycling of photosynthetically fixed carbon. The past decade has seen a burgeoning interest in all aspects of the biochemistry and molecular biology of cellulases and xylanases, which at first glance is rather perplexing. As a paradigm for β -1,4-glycanases, lysozyme is the enzyme of choice; it has been studied in greater detail than cellulases or xylanases and has a clearly understood catalytic mechanism. The recent discoveries from analysis of gene sequences and the three-dimensional structure of cellulase proteins show, however, that there is much to be understood about the enzymic hydrolysis of plant carbohydrate polymers beyond what can be adduced from analogy with lysozyme. To really appreciate the rationale behind current studies of microbial cellulases and xylanases, it is necessary to take a broad view that takes into account not just their intrinsic interest, but also their undoubted commercial potential. The structural polysaccharides cellulose and hemicellulose (xylan being the major

component of hemicellulose) together account for greater than 50% of plant biomass and are consequently the most abundant terrestrial organic molecules. The value of plant biomass as a renewable, and therefore inexhaustible, resource is immediately apparent if one considers that its estimated total energy content is equivalent to 640 billion tonnes of oil (Coughlan, 1985). Plant structural polysaccharides provide the major source of nutrients for ruminant livestock, and have the potential, through microbial fermentations, to provide renewable substrates for the chemical, pharmaceutical and feed industries.

A number of specific examples of how microbial cellulases and xylanases might be exploited in agricultural and industrial processes can be cited (Gilbert & Hazlewood, 1991).

(i) Pre-treatment of forage crops and other cellulosic biomass with cellulases and xylanases, to improve the nutritional quality and digestibility of ruminant feeds, or to facilitate composting.

(ii) Addition of cellulases and xylanases to pig and poultry cereal-based diets, to elicit a significant improvement in nutrient utilization, through the hydrolysis of barley β -glucan and arabinoxylans.

(iii) Improving the ensilage process by the introduction of cellulolytic capacity into silage bacteria, and pretreatment of ensiled herbage with cellulases and hemicellulases.

(iv) Enzymic saccharification of agricultural, industrial and municipal wastes to provide sugar syrups for human or animal consumption, or for the production of fine chemicals through industrial fermentations.

(v) Xylanase pretreatment of paper pulps to remove xylan while preserving cellulose content, and reducing dependence on chlorine in the brightening process.

(vi) Enzymic digestion of industrial wastes as an alternative to landfill deposition.

* Author for correspondence. Tel. (091) 222 6000 ext. 6962; fax (091) 222 6720.

Abbreviation: CBD, cellulose-binding domain.

(vii) Use of xylanase as a flour improver for bakery products.

It is apparent that plant biomass provides an extensive renewable resource; however, its exploitation has been limited by the cost of hydrolysing plant structural polysaccharides to their component sugars. Although microbial cellulases and xylanases can be produced by some organisms at very high levels, for example *Trichoderma reesei* produces 20 g of cellulase per litre of culture (Mandels, 1985), the enzymes have extremely low turnover rates when hydrolysing crystalline substrates. The specific activity of purified cellulase from *T. reesei* is 0.6 U (mg protein)⁻¹, which is 2–4 orders of magnitude lower than for most enzymes (Mandels, 1985). Furthermore, hydrolysis of the designated substrate is invariably incomplete. Thus, the production of enzymes which exhibit a relatively high specific activity against cellulose and xylan is crucial if the potential agricultural and industrial benefits of plant biomass are to be fully realized. The development of such enzymes is not beyond the scope of current technology, but will be possible only when we have a detailed understanding of the structure and catalytic mechanism of existing cellulases and xylanases.

Biochemical studies

Cellulose, a homopolymer of cellobiose units linked by β -1,4 bonds, is hydrolysed by essentially two types of enzyme systems defined as non-aggregating and aggregating enzymes. Micro-organisms that produce cellulases are almost invariably able to hydrolyse xylan also; there are few examples of organisms able to hydrolyse only one of these two structural polysaccharides. (Biely, 1991). Cellulases and xylanases from a wide range of micro-organisms have been purified, analysed and categorized according to their substrate specificities. Much of this work has been considered in earlier reviews (see, for example, Coughlan & Ljungdahl, 1988; Robson & Chambliss, 1989; Coughlan, 1990; Stutzenberger, 1990; Gilkes *et al.*, 1991*b*) and will not be touched upon further. For our purposes data accumulated from previous studies can be summarized as follows.

Non-aggregating enzymes

There are essentially three types of enzyme in this system: β -1,4-endoglucanase (EC 3.2.1.4), which cleaves internal β -1,4-glycosidic bonds; cellobiohydrolase (EC 3.2.1.91), which releases cellobiose from the non-reducing end of cellulose; and β -glucosidase (EC 3.2.1.21), which hydrolyses cellobiose to glucose (Coughlan & Ljungdahl, 1988; Wood, 1989). The three enzymes act in synergy to degrade cellulose. The rationale for the synergy is as follows: although cellulose is

chemically homogeneous, it is structurally diverse, and comprises amorphous regions where the cellulose chains are not closely linked, and crystalline areas where inter- and intra-chain hydrogen bonding results in an ordered array of cellulose microfibrils. It is this latter form of polysaccharide which is particularly recalcitrant to enzyme attack. It is generally agreed that the endo-acting enzymes hydrolyse bonds in the amorphous regions creating numerous non-reducing ends to the cellulose molecules. The exo-acting cellulase (cellobiohydrolase) initiates cellulose hydrolysis at these newly created termini and continues releasing cellobiose from the crystalline regions of the polymer. The β -glucosidase, by cleaving cellobiose, prevents the disaccharide from inhibiting primarily the cellobiohydrolase (Wood, 1989).

The above description of a non-aggregating cellulase system is by necessity an oversimplification. Single enzymes can display endo- and exo-glucanase activity and synergism is evident not only between enzymes with different substrate specificities, but also between cellulases with apparently the same mode of action (Coughlan & Ljungdahl, 1988; Coughlan, 1990). Furthermore, although cellobiohydrolases are clearly present in fungal cellulases, the evidence for this enzyme in comparable non-aggregating bacterial systems is less certain. In *Pseudomonas fluorescens* subsp. *cellulosa*, a bacterium capable of hydrolysing crystalline cellulose, endoglucanase (Gilbert *et al.*, 1987), cellobiohydrolase (cleaves oligosaccharides via an endo-mechanism; Ferreira *et al.*, 1991) and β -glucosidase (Rixon *et al.*, 1992) activities were detected but the organism did not synthesize a cellobiohydrolase. Although *Cellulomonas fimi* does express an exo-acting cellulase (O'Neill *et al.*, 1986), its classification as a true cellobiohydrolase remains uncertain. Likewise, current evidence for the synthesis of a true cellobiohydrolase by cellulolytic *Thermomonospora* spp. is tenuous (Wilson, 1988; Stutzenberger, 1990).

Aggregating system

In several cellulolytic anaerobic bacteria, secreted cellulases associate into high molecular mass multienzyme complexes, whose quaternary structure is pivotal in the rapid hydrolysis of crystalline cellulose. The complexes are generally associated with the cell surface and mediate attachment between cells and the insoluble substrate. The most extensively studied aggregating system is that of *Clostridium thermocellum* (Lamed & Bayer, 1988). The complex, termed the 'cellulosome', comprises at least 14 distinct polypeptides including numerous endoglucanases and xylanases and at least one β -glucosidase. These enzymically active polypeptides are associated with a 210 kDa non-catalytic scaffolding protein, termed S_L (or

S1), which mediates adherence to either the cell surface and/or the substrate (Bayer & Lamed, 1986; Wu *et al.*, 1988). Recently, a cellobiohydrolase was also shown to be a component of the cellulosome (Morag *et al.*, 1991). Individual components of the cellulosome display little activity against crystalline cellulose. A model for the action of this cellulase complex (Mayer *et al.*, 1987) proposes that cellulose is attacked simultaneously by regularly spaced catalytic subunits which are lined up along the cellulose molecule. However, it is not entirely clear how this model accounts for the cellobiohydrolase, or the synergy observed between *Trichoderma koningii* exocellulase and the *Clostridium* cellulase complex (Gow & Wood, 1988). Other cellulolytic bacteria which express cellulosome-like complexes include *Ruminococcus albus*, *Ruminococcus flavefaciens*, *Fibrobacter succinogenes* and other clostridia (Lamed *et al.*, 1987; Shoseyov & Doi, 1990). Cellobiohydrolase activity is apparently associated with the *Ruminococcus flavefaciens* and *Clostridium stercorarium* cellulase complexes (Béguin, 1990; Bronnenmeier *et al.*, 1991).

Xylanases

Xylan, the major hemicellulose component of the plant cell wall in most plant species, is hydrolysed by a repertoire of hydrolytic enzymes. Unlike cellulose, xylan is a complex polysaccharide comprising a backbone of xylose residues linked by β -1,4-glycosidic bonds. Depending on source, the xylopyranoside units are substituted with mainly acetyl, arabinosyl and glucuronosyl residues. Hardwood xylan is typically *O*-acetyl-4-*O*-methylglucuronoxylan with approximately 10% of xylose units α -1,2-linked to a 4-*O*-methylglucuronic acid side-chain, and 70% of xylose residues acetylated at the C-2 or C-3 positions. Softwood xylans are commonly arabinoxylans in which 10% of the xylose units are substituted with α -2,3-linked arabinofuranose residues (Whistler & Richards, 1970; Biely, 1985). Microbial enzymes act co-operatively to convert xylan to its constituent simple sugars: these enzymes include β -1,4-endoxylanases (xylanases; EC 3.2.1.8), which cleave internal glycosidic bonds within the xylan backbone; arabinofuranosidase (EC 3.2.1.55) which hydrolyses arabinose side-chains; α -glucuronidase which removes glucuronic acid side-chains from the xylose units; xylan esterases (EC 3.1.1.6) which release acetate groups; and finally xylosidase (EC 3.2.1.37), which hydrolyses xylobiose to xylose (Wong *et al.*, 1988). It is apparent that there is a considerable degree of synergy between these enzymes (Poutanen *et al.*, 1991). For example, many xylanases will not cleave glycosidic bonds between xylose units which are substituted. Thus, side-chains must be cleaved before the xylan backbone can be

completely hydrolysed (Lee & Forsberg, 1987). Conversely, several accessory enzymes will only remove side-chains from xylooligosaccharides. These enzymes therefore require xylanases to partially hydrolyse the plant structural polysaccharide, before side-chains can be cleaved (Poutanen *et al.*, 1991). Although the structure of xylan is more complex than cellulose, and thus requires a large number of different enzymes to elicit efficient hydrolysis, the polysaccharide does not form tightly packed structures and is thus more accessible to hydrolytic enzymes. Consequently, the specific activity of xylanases is 2–3 orders of magnitude greater than for cellulase hydrolysis of crystalline cellulose.

Multiplicity of cellulases and xylanases

Most micro-organisms capable of degrading cellulose and xylan synthesize a range of isoenzymes with very similar substrate specificities (Wong *et al.*, 1988). However, as extracellular plant structural polysaccharide hydrolases are often post-translationally modified through glycosylation and partial proteolysis, it is not clear, from biochemical studies, whether the apparent isoenzymes are the products of multiple genes or the result of the modification of a single enzyme (Wong *et al.*, 1988). Similarly, the definition of substrate specificity for cellulases and xylanases is also far from clear. For example, the classification of cellulases is less distinct than suggested earlier in this review, with some endoglucanases and cellobiohydrolases exhibiting limited exo- and endo-activity, respectively (Coughlan, 1990). Furthermore, several cellulases also appear to exhibit xylanase activity (see, for example, Hall *et al.*, 1988). The problem of defining the substrate specificity of plant cell wall hydrolases is compounded by the difficulty of obtaining reproducible and structurally characterized substrates for these enzymes. In addition, it could be argued that the apparent broad specificity of some plant cell wall hydrolases may reflect the contamination of purified enzymes with other cellulases and/or xylanases (Coughlan, 1990).

Molecular biology of cellulases and xylanases

Although biochemical analysis has generated a comprehensive insight into the action of cellulases and xylanases, several questions remain unanswered. (i) Are the substrate specificities of those enzymes which apparently display multiple activities a consequence of impure preparations? (ii) Are the isoenzymic xylanases and cellulases observed in bacterial and fungal systems a consequence of post-translational processing of a few enzymes or, almost unique in prokaryotes, are there extensive gene families encoding isoenzymes which hydrolyse cellulose and xylan? (iii) What are the

structures of cellulases and xylanases – have they evolved from common ancestral genes by divergent evolution, or is there evidence for convergent evolution? (iv) Can plant cell wall material be more efficiently exploited through the construction of novel enzymes or the introduction of cellulase and xylanase genes into industrial and agriculturally important organisms?

By enabling research workers to manipulate and modify genomic DNA *in vitro*, and to clone and express genes coding for single components of complex enzyme systems in *Escherichia coli*, recombinant DNA technology has provided the means to address several of the questions listed above. The following is not an exhaustive treatment of the topic, but is an attempt to illustrate how molecular studies have advanced our understanding of cellulases and xylanases.

Multiple genes

Since the early 1980s, when the isolation of cellulase genes was first reported, the primary structures of in excess of 70 cellulases and xylanases have been determined (Gilkes *et al.*, 1991a). Inspection of the amino acid sequences of these enzymes, coupled with hydrophobic cluster analysis (Henrissat *et al.*, 1989), has enabled the evolutionary relationships of these enzymes to be determined. The data have revealed the presence of nine families of cellulases and xylanases. It is apparent that the isoenzymic forms of cellulases and xylanases, observed in several organisms, are a consequence of large multigene families, and not solely the result of processing of a single gene product. For example at least fifteen cellulase, two xylanase and two β -glucosidase genes have been isolated from *Clostridium thermocellum* (Hazlewood *et al.*, 1988) and similarly, multiple cellulase and xylanase genes have been characterized from strains of *Ruminococcus*, *Pseudomonas*, *Cellulomonas*, *Bacillus*, *Fibrobacter* and *Butyrivibrio* (Gilkes *et al.*, 1991a). Single bacteria often contain cellulase and xylanase genes from different families, suggesting that micro-organisms have acquired multiple plant cell wall hydrolase genes not exclusively through gene duplication, but via extensive horizontal gene transfer. That genes in a single family can be derived from aerobic, anaerobic, Gram-positive and Gram-negative bacteria illustrates the high degree of cellulase and xylanase gene transfer between prokaryotes (Béguin 1990; Gilkes *et al.*, 1991a). In addition, families A, B, F and H contain both fungal and bacterial enzymes, while family E consists of prokaryotic and plant enzymes. This suggests that lateral transfer of cellulase and xylanase genes has also occurred.

Enzymes in a single family do not always display the same substrate specificity. For example, families B and C contain both cellobiohydrolases and endoglucanases;

similarly, the substrate specificity of enzymes in family A is also variable. Gilkes *et al.* (1991a) suggested that whether the activity observed is that of an exo- or endoglucanase may be a consequence of fine details of three-dimensional structure rather than of overall conformation. In contrast, those enzymes which predominantly hydrolyse xylan are in two discrete groups, F and G together with exoglucanases, which also display considerable xylanase activity. It appears, therefore, that true cellulases and xylanases have evolved from separate genes and are thus quite distinct enzymes. However, it should be emphasized that the cloning of cellulase and xylanase genes into non-cellulolytic/xylanolytic backgrounds confirmed previous biochemical data which revealed that some cellulases are also capable of hydrolysing xylan.

In general, cellulases and xylanases comprise a single catalytic domain although two notable exceptions to this rule have recently been reported: *Caldocellum saccharolyticum* CelB has an N-terminal exoglucanase and a C-terminal endoglucanase domain (Saul *et al.*, 1990), while *Ruminococcus flavefaciens* XYLA comprises two xylanase catalytic domains from groups F and G respectively (Zhang & Flint, 1992).

The rationale for the acquisition of multiple cellulases and xylanases by a single organism is an intriguing question. Presumably, this reflects the recalcitrant nature of plant cell walls to enzymic attack. For bacteria to efficiently hydrolyse plant structural polysaccharides, high level expression of cellulases and xylanases is required. Multiple genes encoding these enzymes would result in elevated synthesis of the required proteins. However, prokaryotes could also increase cellulase/xylanase synthesis through gene duplication or up mutations in promoter regions. Apart from the tandem repeat of an endoglucanase gene in *Bacillus* sp. strain N-4 (Fukumori *et al.*, 1986), there is little evidence for extensive gene duplication in prokaryotes. Thus, it could be argued that the acquisition, from diverse sources, of cellulase and xylanase by a single organism reflects the heterogeneous nature of the substrate. Although cellulose molecules are chemically homogeneous, structurally, they are quite diverse. It is possible, therefore, that various isoenzymes could hydrolyse β -1,4 bonds that are in different physical environments within the cellulose molecule. This argument can also be applied to xylan hydrolysis in view of the heterogeneous chemical and physical nature of the polymer. Two approaches could be employed to answer these questions: (i) synergy experiments could be performed, using recombinant cellulases and xylanases expressed in a non-cellulolytic background, to assess the minimum number of different enzymes required to reconstitute an active cellulase/xylanase system; (ii) the effect on plant cell wall

hydrolysis of inactivating specific cellulase/xylanase genes in a given bacterium could be evaluated. In a recent study by Seiboth *et al.* (1992), the cellobiohydrolase II gene of *T. reesei* was deleted and the capacity of the resulting strain to hydrolyse cellulose was assessed. The data suggested that the enzyme was important in inducing cellulase synthesis, but was not essential for degradation of the polysaccharide.

Non-catalytic domains

(a) *Linker sequences.* It is now apparent that the molecular architecture of cellulases and xylanases can be of two types; single domain or multidomain. In the former, the enzymes comprise a single catalytic domain. Examples of these single domain enzymes can be found in *Ruminococcus* and *Butyrivibrio* (Poole *et al.*, 1990; Hazlewood *et al.*, 1990). Most cellulases and xylanases, however, consist of multiple domains which are joined by characteristic linker sequences. The majority of these sequences are rich in either serine residues, as found in *Pseudomonas* cellulases and xylanases (Hall & Gilbert, 1988; Hall *et al.*, 1989; Gilbert *et al.*, 1990; Kellett *et al.*, 1990; Ferreira *et al.*, 1991), or a combination of proline and threonine residues which predominate in the corresponding *Clostridium* (Hall *et al.*, 1988; Béguin *et al.*, 1985; Grépinet *et al.*, 1988; Grépinet & Béguin, 1986) and *Cellulomonas* (Gilkes *et al.*, 1991a) enzymes. The sequences vary considerably in length (6–59 amino acids). A notable exception, however, is the linker sequence of *R. flavefaciens* XYLA, which consists of 374 residues and is extraordinarily rich in asparagine (45%) and glutamine (26%). In fact, only seven amino acids are represented in this region, the others being tryptophan (16%), glutamate, glycine, alanine and threonine (Zhang & Flint, 1992).

(b) *Repeated sequences.* Numerous cellulases, and some xylanases, contain reiterated regions which vary in size from 20 to 150 residues. In general, the role of these repeats is unknown, although it has been postulated that the large repeats in CelZ, CenB and CenC could constitute non-catalytic cellulose-binding domains (CBD). Recent studies by Tokatlidis *et al.* (1991) and Fujino *et al.* (1992) have provided an insight into the role of the 24-amino-acid duplicated sequence found in numerous *C. thermocellum* cellulases, xylanases and the non-catalytic subunit S1. The sequence interacts directly with the 147-residue repeated sequence (seven repeats) present in the non-catalytic scaffolding protein S1, and thus provides at least one of the mechanisms by which *C. thermocellum* cellulases are assembled into a multi-enzyme complex (Fujino *et al.*, 1992).

(c) *Cellulose-binding domains.* Numerous multidomain cellulases contain a CBD. In bacteria, the majority of

CBDs are characteristically 100 residues in length and can be located at either the N- or C-terminus. Considerable homology exists between these domains in which the following features are highly conserved: (i) two cysteines are present close to the N- and C-termini respectively; (ii) there are four very highly conserved tryptophan residues in addition to glycine and asparagine residues which also show identity in the various domains; (iii) there is a marked lack of charged amino acids (Ferreira *et al.*, 1991). By analogy with other sugar-binding proteins, it has been proposed that the interaction of the CBD with cellulose is mediated by the conserved aromatic residues, either through hydrogen-bonding or hydrophobic interactions (Béguin, 1990). It has been argued that CBDs confer cellulase specificity on β -1,4-glycanases (West *et al.*, 1989). However, it is difficult to reconcile this view with the location of these domains in xylanases, an arabinofuranosidase (Kellett *et al.*, 1990) and an esterase (H. J. Gilbert, unpublished data). In fungi, CBDs appear to be essential for the respective enzymes to attack crystalline cellulose. However, removal of CBDs from *Cellulomonas* cellulases resulted in only a small decrease in enzymic activity (Shen *et al.*, 1991), while a truncated *Pseudomonas* xylanase, lacking its endogenous CBD, exhibited similar catalytic activity to the full-length enzyme (Ferreira *et al.*, 1990). Thus, the precise role of bacterial CBDs in plant cell wall hydrolases is a matter of some debate. However, it would be wrong to assume that the conserved 100-residue CBDs found in many bacterial enzymes all have an equivalent role. For example, the *Cellulomonas*-derived CBDs can be eluted from cellulose with distilled water, while the corresponding *Pseudomonas* domain requires protein denaturants to effect dissociation from its ligand (Gilbert *et al.*, 1990). Furthermore, recent studies by Din *et al.* (1991) showed that a *Cellulomonas* CBD disrupts, and to a limited extent, solubilizes crystalline cellulose. In view of the widespread occurrence of CBDs in *Pseudomonas* plant cell wall hydrolases, we postulate that in these proteins the CBD enables a repertoire of enzymes to adhere to the plant cell wall. As the recalcitrant tissue is degraded, different polysaccharides are exposed at the surface and are hydrolysed by the appropriate enzyme(s), which are held in close proximity with their respective substrates through the CBD/cellulose interactions. The absence of xylan binding-domains probably reflects the heterogeneity of the polysaccharide, which might preclude the evolution of a protein domain which binds to all xylans, irrespective of their source.

Watanabe *et al.* (1992) have shown that a β -1,3-glucanase from *Bacillus circulans* comprises four structural domains; a 100-amino-acid N-terminal repeated sequence, followed by a 120-residue domain which is

linked to a C-terminal 350-amino-acid catalytic domain. The three N-terminal domains constitute a non-catalytic β -1,3-glucan-binding domain. Truncated derivatives of the enzyme consisting of only the catalytic region efficiently hydrolyse soluble glucan, but exhibit diminished activity against insoluble β -1,3-glucan, when compared to the full-length enzyme.

Bacterial CBDs that do not exhibit homology with the 100-residue CBD described above have been found recently in enzymes from *Cellulomonas* (Coutinho *et al.*, 1991; Meinke *et al.*, 1991) and *Clostridium* (Durrant *et al.*, 1991). Both domains bind weakly to cellulose. The *C. thermocellum* CBD decreases both the K_m and V_{max} of the enzyme for cellulose. Removal of the domain switches the properties of the cellulase to suit the availability of substrate. Thus, the full-length enzyme in early cellulolysis efficiently hydrolyses the low concentration of soluble cellulose polymers generated. As cellulose hydrolysis proceeds there is a build-up of cellulooligosaccharides and a concomitant increase in proteinase activity that removes the CBD from the enzyme, which then displays increased cellulase activity at the elevated substrate concentration. A further example of a CBD that does not conform to the corresponding *Pseudomonas* domain is found in the S1 subunit of the *Clostridium thermocellum* cellulosome (Poole *et al.*, 1992). The 167-amino-acid sequence, which exhibits homology with the non-catalytic domains of endoglucanase B from *Caldocellum saccharolyticum* (Saul *et al.*, 1990), Avicelase I from *Clostridium stercoararium* (Jauris *et al.*, 1990), an endoglucanase from *Bacillus subtilis* PAP115 (McKay *et al.*, 1986) and the non-catalytic scaffolding protein of the cellulosome produced by *Clostridium cellulovorans* (Shoseyov *et al.*, 1992), binds tightly to cellulose but not xylan. The S1 CBD is present in at least one other *C. thermocellum* endoglucanase (Hazlewood *et al.*, 1993).

Modular structure of cellulases and xylanases

It is apparent that many plant cell wall hydrolases have a modular structure in which the various domains are not located in equivalent positions in the different enzymes. In addition, catalytic domains which are structurally unrelated are linked, in some cases to identical CBDs. Thus, extensive modular shuffling has occurred during the course of evolution. The mechanism by which this has occurred is open to question. In eukaryotes, functional domains of proteins are often encoded by distinct exons (Doolittle, 1979). Extended introns enable exons to be excised, without interruption of coding sequence, and fused to other exons to generate novel proteins. The shuffling of functional domains within bacterial cellulases and xylanases points to the existence of a similar mechanism. The DNA sequences

encoding the linker sequences in these enzymes may fulfil a role analogous to that of introns, by enabling sequences encoding discrete domains to be excised and fused to other genes, thus generating novel hybrid enzymes. Support for the redundant role of linker sequences is provided by Ferreira *et al.* (1990), who showed that removal of these regions from a xylanase and cellulase, respectively, did not markedly affect the catalytic activity or cellulase-binding capacity of the enzymes. Alternatively, it could be argued that the linker sequences provide a flexible enzyme structure in which the polysaccharidase is anchored to the plant cell wall via the CBD/cellulose interaction and the catalytic domain can then locate and hydrolyse β -1,4-glycosidic bonds at a variable distance from the point of cellulose attachment.

Three-dimensional structure of cellulases and xylanases

The crystallization of multidomain enzymes has presented serious problems due to the flexible structure of the linker regions. However, low-angle X-ray scattering has revealed a tadpole-shaped structure for two multidomain cellulases in which the catalytic domain forms a globular structure while the CBD and linker sequences are more extended (Pilz *et al.*, 1990). Recently, the three-dimensional structures of the catalytic domains of two cellulases have been determined through X-ray crystallography (Rouvinen *et al.*, 1990; Juy *et al.*, 1992). Both enzymes contain a long groove in which the active site is located. The structures of these enzymes support the view that cellulases hydrolyse glycosidic bonds through an acid-base catalytic mechanism in which acidic amino acids act as proton donors and acceptors. Further support for this view is provided by several site-directed mutagenesis studies in which these acidic residues were shown to be essential for catalytic activity (Chauvaux *et al.*, 1992).

Conclusions

From the foregoing discussion, it is apparent that cellulases and xylanases are not just interesting proteins from an industrial and agricultural viewpoint, but are also of intrinsic scientific interest. In recent years, there has been substantial progress in the analysis of the molecular architecture of these enzymes, many of which have modular structures comprising non-catalytic and catalytic domains. Several of the non-catalytic domains constitute CBDs. Once the mechanism by which the CBDs elicit enhanced plant cell wall hydrolysis has been defined, the modular structure of these enzymes can be exploited to construct novel enzymes with increased capacity to degrade the plant cell wall. These engineered

enzymes can then be utilized in maximizing the agricultural and industrial exploitation of plant biomass.

References

- BAYER, E. A. & LAMED, R. (1986). Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *Journal of Bacteriology* **167**, 828–836.
- BÉGUIN, P. (1990). Molecular biology of cellulose degradation. *Annual Review of Microbiology* **44**, 219–248.
- BÉGUIN, P., CORNET, P. & AUBERT, J. P. (1985). Sequence of a cellulase gene of the thermophilic bacterium *Clostridium thermocellum*. *Journal of Bacteriology* **162**, 102–105.
- BIELY, P. (1985). Microbial xylanolytic systems. *Trends in Biotechnology* **3**, 286–291.
- BIELY, P. (1991). Biotechnological potential and production of xylanolytic systems free of cellulases. In *Enzymes in Biomass Conversion*, ACS symposium series 460, pp. 408–416. Edited by G. F. Leatham & M. E. Himmel. Washington, DC: American Chemical Society.
- BRONNENMEIER, K., RÜCKNAGEL, K. & STAUNDENBAUER, W. L. (1991). Purification and properties of a novel type of exo-1,4- β -glucanase (Avicelase II) from the cellulolytic thermophile *Clostridium stercoarum*. *European Journal of Biochemistry* **200**, 379–385.
- CHAUVAUX, S., BÉGUIN, P. & AUBERT, J. P. (1992). Site-directed mutagenesis of essential carboxylic residues in *Clostridium thermocellum* endoglucanase CelD. *Journal of Biological Chemistry* **267**, 4472–4478.
- COUGHLAN, M. P. (1985). Cellulases: production, properties and applications. *Biochemical Society Transactions* **13**, 405–406.
- COUGHLAN, M. P. (1990). Cellulose degradation by fungi. In *Microbial Enzymes and Biotechnology*, pp. 1–36. Edited by W. M. Fogarty & C. T. Kelly. London & New York: Elsevier Applied Science.
- COUGHLAN, M. P. & LJUNGAHL, L. G. (1988). Comparative biochemistry of fungal and bacterial cellulolytic enzyme systems. In *Biochemistry and Genetics of Cellulose Degradation*, FEMS Symposium no. 43, pp. 11–30. Edited by J.-P. Aubert, P. Béguin & J. Millet. London & San Diego: Academic Press.
- COUTINHO, J. B., MOSER, B., KILBURN, D. G., WARREN, R. A. J. & MILLER, R. C. (1991). Nucleotide sequence of the endoglucanase C gene (*cenC*) of *Cellulomonas fimi*, its high-level expression in *Escherichia coli*, and characterization of its products. *Molecular Microbiology* **5**, 1221–1233.
- DIN, N., GILKES, N. R., TEKANT, B., MILLER, R. C., WARREN, R. A. J. & KILBURN, D. G. (1991). Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. *Bio/Technology* **9**, 1096–1099.
- DOOLITTLE, R. F. (1979). Protein evolution. In *The Proteins*, vol. 4, pp. 1–118. Edited by H. Neurath & R. L. Hills. New York: Academic Press.
- DURRANT, A. J., HALL, J., HAZLEWOOD, G. P. & GILBERT, H. J. (1991). The non-catalytic C-terminal region of endoglucanase E from *Clostridium thermocellum* contains a cellulose-binding domain. *Biochemical Journal* **273**, 289–293.
- FERREIRA, L. M. A., DURRANT, A. J., HALL, J., HAZLEWOOD, G. P. & GILBERT, H. J. (1990). Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase. *Biochemical Journal* **269**, 261–264.
- FERREIRA, L. M. A., HAZLEWOOD, G. P., BARKER, P. J. & GILBERT, H. J. (1991). The cellodextrinase from *Pseudomonas fluorescens* subsp. *cellulosa* consists of multiple functional domains. *Biochemical Journal* **279**, 793–799.
- FUJINO, T., BÉGUIN, P. & AUBERT, J.-P. (1992). Cloning of a *Clostridium thermocellum* DNA fragment encoding polypeptides that bind the catalytic components of the cellulosome. *FEMS Microbiology Letters* **94**, 165–170.
- FUKUMORI, F. N., SASHIHARA, N., KUDO, T. & HORIKOSHI, K. (1986). Nucleotide sequences of two cellulase genes from alkalophilic *Bacillus* sp. strain N-4 and their strong homology. *Journal of Bacteriology* **168**, 479–485.
- GILBERT, H. J. & HAZLEWOOD, G. P. (1991). Genetic modification of fibre digestion. *Proceedings of the Nutrition Society* **50**, 173–186.
- GILBERT, H. J., JENKINS, G., SULLIVAN, D. A. & HALL, J. (1987). Evidence for multiple carboxymethylcellulase genes in *Pseudomonas fluorescens* subsp. *cellulosa*. *Molecular and General Genetics* **210**, 551–556.
- GILBERT, H. J., HALL, J., HAZLEWOOD, G. P. & FERREIRA, L. M. A. (1990). The N-terminal region of an endoglucanase from *Pseudomonas fluorescens* subspecies *cellulosa* constitutes a cellulose-binding domain that is distinct from the catalytic centre. *Molecular Microbiology* **4**, 759–767.
- GILKES, N. R., HENRISSAT, B., KILBURN, D. G., MILLER, R. C., JR & WARREN, R. A. J. (1991a). Domains in microbial β -1,4-glycanases: sequence conservation, function, and enzyme families. *Microbiological Reviews* **55**, 303–315.
- GILKES, N. R., KILBURN, D. G., MILLER, R. C. & WARREN, R. A. J. (1991b). Bacterial cellulases. *Bioresource Technology* **36**, 21–35.
- GOW, L. A. & WOOD, T. M. (1988). Breakdown of crystalline cellulose by synergistic action between cellulase components from *Clostridium thermocellum* and *Trichoderma koningii*. *FEMS Microbiology Letters* **50**, 247–252.
- GRÉPINET, O. & BÉGUIN, P. (1986). Sequence of the cellulase gene of *Clostridium thermocellum* coding for endoglucanase B. *Nucleic Acids Research* **14**, 1791–1799.
- GRÉPINET, O., CHEBROU, M.-C. & BÉGUIN, P. (1988). Nucleotide sequence and deletion analysis of the xylanase gene (*xynZ*) of *Clostridium thermocellum*. *Journal of Bacteriology* **170**, 4582–4588.
- HALL, J. & GILBERT, H. J. (1988). The nucleotide sequence of a carboxymethylcellulase gene from *Pseudomonas fluorescens* subsp. *cellulosa*. *Molecular and General Genetics* **213**, 112–117.
- HALL, J., HAZLEWOOD, G. P., BARKER, P. J. & GILBERT, H. J. (1988). Conserved reiterated domains in *Clostridium thermocellum* endoglucanases are not essential for catalytic activity. *Gene* **69**, 29–38.
- HALL, J., HAZLEWOOD, G. P., HUSKISSON, N. S., DURRANT, A. J. & GILBERT, H. J. (1989). Conserved serine-rich sequences in xylanase and cellulase from *Pseudomonas fluorescens* subsp. *cellulosa*: internal signal sequence and unusual protein processing. *Molecular Microbiology* **3**, 1211–1219.
- HAZLEWOOD, G. P., ROMANIEC, M. P. M., DAVIDSON, K., GRÉPINET, O., BÉGUIN, P., MILLET, J., RAYNAUD, O. & AUBERT, J.-P. (1988). A catalogue of *Clostridium thermocellum* endoglucanase, β -glucosidase and xylanase genes cloned in *Escherichia coli*. *FEMS Microbiology Letters* **51**, 231–236.
- HAZLEWOOD, G. P., DAVIDSON, K., LAURIE, J. I., ROMANIEC, M. P. M. & GILBERT, H. J. (1990). Cloning and sequencing of the *celA* gene encoding endoglucanase A of *Butyrivibrio fibrisolvens* strain A46. *Journal of General Microbiology* **136**, 2089–2097.
- HAZLEWOOD, G. P., DAVIDSON, K., LAURIE, J. I., HUSKISSON, N. S. & GILBERT, H. J. (1993). Gene sequence and properties of CelI, a family E endoglucanase from *Clostridium thermocellum*. *Journal of General Microbiology* **139**, 307–316.
- HENRISSAT, B., CLAEYSSSENS, M., TOMME, P., LEMESLE, L. & MORNON, J.-P. (1989). Cellulase families revealed by hydrophobic cluster analysis. *Gene* **81**, 83–95.
- JAUROS, S., RÜCKNAGEL, K. P., SCHWARZ, W. H., KRATZSCH, P., BRONNENMEIER, K. & STAUNDENBAUER, W. L. (1990). Sequence analysis of the *Clostridium stercoarum* *celZ* gene encoding a thermoactive cellulase (Avicelase I). Identification of catalytic and cellulose binding domains. *Molecular and General Genetics* **223**, 258–267.
- JUY, M., AMIT, A. G., ALZARI, P. M., POLJAK, R. J., CLAEYSSSENS, M., BÉGUIN, P. & AUBERT, J.-P. (1992). Three dimensional structure of a thermostable bacterial cellulase. *Nature, London* **357**, 89–91.
- KELLETT, L. E., POOLE, D. M., FERREIRA, L. M. A., DURRANT, A. J., HAZLEWOOD, G. P. & GILBERT, H. J. (1990). Xylanase B and an arabinofuranosidase from *Pseudomonas fluorescens* subsp. *cellulosa* contain identical cellulose-binding domains and are encoded by adjacent genes. *Biochemical Journal* **272**, 369–376.
- LAMED, R. & BAYER, E. A. (1988). The cellulosome of *Clostridium thermocellum*. *Advances in Applied Microbiology* **33**, 1–46.
- LAMED, R., NAIMARK, J., MORGENSTERN, E. & BAYER, E. A. (1987).

- Specialized cell surface structures in cellulolytic bacteria. *Journal of Bacteriology* **169**, 3792–3800.
- LEE, S. F. & FORSBERG, C. W. (1987). Purification and characterization of an α -L-arabinofuranosidase from *Clostridium acetobutylicum*. *Canadian Journal of Microbiology* **33**, 1011–1016.
- McKAY, R. M., LO, O., WILLICK, G., ZUCKER, M., DAIRD, S., DOVE, M., MORANELLI, F. & SELIGY, V. (1986). Structure of a *Bacillus subtilis* endo- β -1,4-glucanase gene. *Nucleic Acids Research* **14**, 9159–9170.
- MANDELS, M. (1985). Application of cellulases. *Biochemical Society Transactions* **13**, 414–416.
- MAYER, F., COUGHLAN, M. P., MARI, Y. & LJUNGDAHL, L. G. (1987). Macromolecular organization of the cellulolytic enzyme complex of *Clostridium thermocellum* as revealed by electron microscopy. *Applied and Environmental Microbiology* **53**, 785–792.
- MEINKE, A., BRAUN, C., GILKES, N. R., KILBURN, D. G., MILLER, R. C., JR & WARREN, R. A. J. (1991). Unusual sequence organization in CenB, an inverting endoglucanase from *Cellulomonas fimi*. *Journal of Bacteriology* **171**, 308–314.
- MORAG, E., HALEVY, I., BAYER, E. A. & LAMED, R. (1991). Isolation and properties of a major cellobiohydrolase from the cellulosome of *Clostridium thermocellum*. *Journal of Bacteriology* **179**, 25–31.
- O'NEIL, G. P., GOH, S. H., WARREN, R. A. J., KILBURN, D. G. & MILLER, R. C. (1986). Structure of the gene encoding the exoglucanase of *Cellulomonas fimi*. *Gene* **44**, 325–330.
- PILZ, I., SCHWARZ, E., KILBURN, D. G., MILLER, R. C., JR, WARREN, R. A. J. & GILKES, N. R. (1990). The tertiary structure of a bacterial cellulase determined by small-angle X-ray scattering analysis. *Biochemical Journal* **271**, 277–280.
- POOLE, D. M., HAZLEWOOD, G. P., LAURIE, J. I., BARKER, P. J. & GILBERT, H. J. (1990). Nucleotide sequence of the *Ruminococcus albus* SY3 endoglucanase genes *celA* and *celB*. *Molecular and General Genetics* **223**, 217–223.
- POOLE, D. M., MORAG, E., LAMED, R., BAYER, E. A., HAZLEWOOD, G. P. & GILBERT, H. J. (1992). Identification of the cellulose binding domain of cellulosome subunit S1 from *Clostridium thermocellum*. *FEMS Microbiology Letters* (in the Press).
- POUTANEN, K., TENKANEN, M., KORTE, H. & PULS, J. (1991). Accessory enzymes involved in the hydrolysis of xylans. In *Enzymes in Biomass Conversion*, ACS Symposium series 460, pp. 426–436. Edited by G. F. Leatham and M. E. Himmel. Washington, DC: American Chemical Society.
- RIXON, J. E., FERREIRA, L. M. A., DURRANT, A. J., LAURIE, J. I., HAZLEWOOD, G. P. & GILBERT, H. J. (1992). Characterization of the gene *celD* and its encoded product 1,4- β -D-glucan glucohydrolase D from *Pseudomonas fluorescens* subsp. *cellulosa*. *Biochemical Journal* **285**, 947–955.
- ROBSON, L. M. & CHAMBLISS, G. H. (1989). Cellulases of bacterial origin. *Enzyme and Microbial Technology* **11**, 626–644.
- ROUVINEN, J., BERGFORS, T., TEERI, T., KNOWLES, J. K. C. & JONES, T. A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* **249**, 380–386.
- SAUL, D. J., WILLIAMS, L. C., GRAYLING, R. W., CHAMLEY, L. W., LOVE, D. R. & BERGQUIST, P. L. (1990). *CelB*, a gene coding for a bifunctional cellulase from the extreme thermophile *Caldocellum saccharolyticum*. *Applied and Environmental Microbiology* **56**, 3117–3124.
- SEIBOTH, B., MESSNER, R., GRUBER, F. & KUBICEK, C. P. (1992). Disruption of the *Trichoderma reesei* *cbh2* gene coding for cellobiohydrolase II leads to a delay in the triggering of cellulase formation by cellulose. *Journal of General Microbiology* **138**, 1259–1264.
- SHEN, H., SCHMUCK, M., PILZ, I., GILKES, N. R., KILBURN, D. G., MILLER, R. C., JR & WARREN, R. A. J. (1991). Deletion of the linker connecting the catalytic and cellulose-binding domains of endoglucanase A (CenA) of *Cellulomonas fimi* alters its conformation and catalytic activity. *Journal of Biological Chemistry* **266**, 11335–11340.
- SHOSEYOV, O. & DOI, R. (1990). Essential 170 kDa subunit for degradation of crystalline cellulose by *Clostridium cellulovorans* cellulase. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 2192–2195.
- SHOSEYOV, O., TAKAGI, M., GOLDSTEIN, M. A. & DOI, R. H. (1992). Primary sequence analysis of *Clostridium cellulovorans* cellulose binding protein A. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 3483–3487.
- STUTZENBERGER, F. (1990). Bacterial cellulases. In *Microbial Enzymes and Biotechnology*, pp. 37–70. Edited by W. M. Fogarty & C. T. Kelly. London & New York: Elsevier Applied Science Publishers.
- TOKATLIDIS, K., SALAMITOU, S., BÉGUIN, P., DHURJATI, P. & AUBERT, J.-P. (1991). Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. *FEBS Letters* **291**, 185–188.
- WATANABE, T., KASAHARA, N., AIDA, K. & TANAKA, H. (1992). Three N-terminal domains of β -1,3-glucanase A1 are involved in binding to insoluble β -1,3-glucan. *Journal of Bacteriology* **174**, 186–190.
- WEST, C. A., ELZANOWSKI, A., YEH, L.-S. & BARKER, W. (1989). Homologues of catalytic domains of *Cellulomonas* glucanases found in fungal and *Bacillus* glucosidases. *FEMS Microbiology Letters* **59**, 167–172.
- WHISTLER, R. L. & RICHARDS, E. L. (1970). Hemicelluloses. In *The Carbohydrates – Chemistry and Biochemistry*, 2nd edn, vol. 2A, pp. 447–469. Edited by W. Pigman & D. Horton. New York: Academic Press.
- WILSON, D. B. (1988). Cellulases of *Thermomonospora fusca*. *Methods in Enzymology* **160**, 314–323.
- WONG, K. K. Y., TAN, L. U. L. & SADDLER, J. N. (1988). Multiplicity of β -1,4-xylanases in microorganisms: functions and applications. *Microbiological Reviews* **52**, 305–317.
- WOOD, T. M. (1989). Mechanisms of cellulose degradation by enzymes from aerobic and anaerobic fungi. In *Enzyme Systems for Lignocellulose Degradation*, pp. 17–35. Edited by M. P. Coughlan. London & New York: Elsevier Applied Science Publishers.
- WU, J. H. D., ORME-JOHNSON, W. H. & DEMAINE, A. L. (1988). Two components of an extracellular protein aggregate of *Clostridium thermocellum* together degrade crystalline cellulose. *Biochemistry* **27**, 1703–1709.
- ZHANG, J.-X. & FLINT, H. J. (1992). A bifunctional xylanase encoded by the *xynA* gene of the rumen cellulolytic bacterium *Ruminococcus flavefaciens* 17 comprises two dissimilar domains linked by an asparagine/glutamine-rich sequence. *Molecular Microbiology* **6**, 1013–1023.