CHARACTERIZATION OF ENDOGLUCANASE CELA FROM THE RUMEN BACTERIUM *CLOSTRIDIUM LONGISPORUM*

by

Volker Mittendorf

A dissertation submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy in the Faculty of Science, University of Cape Town.

Cape Town, August 1994

CERTIFICATE OF SUPERVISOR

In terms of paragraph GP 8 in "general rules for the Degree of Doctor of Philosophy (PhD), I, as supervisor of the candidate Volker Mittendorf, certify that I approve of the incorporation in this dissertation of material that has already been submitted or accepted for publication.

Signed by candidate

Professor J.A. Thomson
Department of Microbiology
University of Cape Town

Dedicated to the women in my life,
My mother, past and future loves and the sea.

CONTENTS

ACKNOWLEDGMENTS

I would like to thank the following people from lab 207 and 208 for their camaraderie, help, discussions and encouragement over the past few years, Di James, Andrew Hackland, Jono Lane, Khotso Mokhele, Molapo Qhobela, Rachel Saunders, Eddy Omolo, Gerardo Herrera, Sandy Lennox and Genevieve Wilson. I am grateful for the enthusiasm and support of my supervisor, Professor Jennifer Thomson. I appreciated the freedom she granted for self-development in my studies as well as extramural activities. The help from the departmental assistants, especially Charlie Hendrickse and Keyam Diedericks, was greatly appreciated.

I thank Prof. M.B. von Wechmar and the animal unit for the production of the anti-CelA antiserum and Prof. W.F. Brandt for the N-terminal amino acid sequence determinations.

I am grateful for financial assistance from AECI Ltd. and from the Foundation for Research Development.

ABSTRACT

Cellulose is the most abundant organic compound on earth and offers great potential as a source of renewable energy and other chemicals. Cellulases are being studied to elucidate the enzymatic degradation of cellulose. We are interested in the molecular mechanisms of microbial degradation of cellulose in the rumen. The long-term aim is the potential genetic modification of lignocellulolytic activities in the rumen. *Clostridium longisporum* was obtained from Varel (1989) because oxygen-resistant endospores might be suitable "vectors" for the introduction of genetically modified enzyme systems into the rumen via animal feeds. It is a sporadically occurring rumen bacterium and its role in ruminal cellulolysis is unclear. The aim of this project was the initial characterization of cellulases produced by *C. longisporum*.

The *celA* gene was obtained by screening a library of *C. longisporum* genomic DNA in *Escherichia coli* for clones expressing CMCase activity. Approximately 38 CMCase-positive clones were obtained and the plasmid pCM4 was isolated from the clone expressing the highest activity. Southern analysis indicated that another plasmid, pCM64, contained a larger insert including the insert of pCM4. A total of 3620 bp were sequenced and a 1548–bp open reading frame, termed *celA*, was found. This gene showed homology with other endo-ß-1,4-glucanases from family 5 (Henrissat & Bairoch, 1993). Plasmid pCM64 was found to contain the whole *celA* gene encoding endoglucanase CelA, while pCM4 has a 5'–truncated gene, termed *celA*Δ5', which encodes a fusion protein, CelAΔN', that was initiated from an ATG codon in the vector. Sequence analysis of *celA* revealed the presence of a type I cellulose-binding domain (Béguin & Aubert, 1994) at the COOH–terminus of CelA.

The transcriptional start site of *celA* was mapped using primer extension and was found to be the same in *C. longisporum* and in *E. coli* expressing the cloned *celA* gene. A consensus *E. coli* –10 promoter region could be identified (AATAAT), but not a –35 promoter region. Two direct repeats (TATTGAATTTAT) separated by 15 nucleotides flank the region where the consensus –35 promoter regions would have been. The size of the *celA* mRNA transcript corresponded with the size of the ORF.

CelA was found to be secreted into the periplasm in *E. coli*, but CelAΔN', which lacks a signal peptide, was retained in the intracellular fraction. Both proteins were purified and their N–terminal amino acid sequences were determined and found to correspond with the DNA sequence of their genes. In *E. coli* proteolytic cleavage of CelA at or near a putative linker region resulted in the appearance of two active polypeptides with molecular weights of 57 and 47 kDa. The former was the full-length enzyme, while the latter was the catalytic domain from which the cellulose-binding domain (CBD) had been removed (CelAΔCBD). The intracellularly located CelAΔN' was not subject to proteolytic degradation. The pH and temperature optima of CelA were pH 4.8 and 43°C, respectively. CelA showed highest activity on barley ß-glucan, but also hydrolysed lichenan, CMC and xylan.

Northern blots indicated that *celA* was transcriptionally induced in *C. longisporum* cultures grown on barley ß-glucan, but no *celA* mRNA transcripts could be detected during growth on cellobiose. The induction of CelA and other extracellular glucanases was visible on zymograms, but significant quantities of these activities were present in uninduced cultures grown on cellobiose. This indicated a low constitutive expression of *celA*. ß-Glucan and xylan zymograms showed the expression of at least five major glucanases, termed CelA to CelE, in the culture supernatant and two, CelB and CelC, were chosen for further investigation.

CHAPTER I

GENERAL INTRODUCTION:

THE ENZYMATIC DEGRADATION OF CELLULOSE

TABLE OF CONTENTS

I hope I have not done injustice to some people by plagiarizing their works incorrectly.

1.1 Introduction

Cellulolytic microorganisms produce an arsenal of enzymes whose primary function it is to hydrolyze ß-1,4-glycosidic linkages in the major plant structural polysaccharides, cellulose and xylan. These enzymes, called cellulases and xylanases, are being investigated by many laboratories worldwide in an effort to gain insight into the intricacies of their molecular structures and enzymatic mechanisms and the regulation of their biosynthesis. This research will hopefully one day enable the effective utilization of the most abundant renewable resource for energy and chemicals on earth, namely cellulose. The ultimate goal is the total saccharification of cellulose into glucose with the coupled fermentation of the glucose into solvents and fuels, but this is not yet economically viable (Béguin & Aubert, 1994). Another as yet unattainable long-term goal is the improvement of fibre digestion in the rumen. Other "present-day" applications of cellulases will be discussed below. This review will concentrate on recent work done on the enzymology, biosynthesis and regulation of cellulases of rumen bacteria, which are responsible for fibre digestion in the rumen. The general enzymology of cellulases will be discussed broadly.

1.2 Substrate complexity

One of the main obstacles for the industrial utilization of cellulases for the saccharification of cellulose is their relatively low turnover rates. The main reason for this is the complexity of the substrate. The cellulose polymer consists of ß-1,4 linked glucose residues and the chain lengths vary between 100 and 14 000 residues. Chains are stacked parallel to each other and intramolecular hydrogen bonding leads to the formation of microfibrils containing highly ordered crystalline regions interspersed with less ordered amorphous regions (Béguin & Aubert, 1994). Microfibrils are insoluble, hydrophobic and show considerable resistance to enzymatic degradation by purified cellulase systems depending on the degree of crystallinity. The crystallinity of the cellulose does however not seem to limit the degradation of the substrate in the rumen and other limiting factors which decrease the accessibility of the cellulose to the enzymes seem to be more important. Primary plant cell walls contain randomly arranged microfibrils embedded in a matrix of

cell wall components. Secondary cell walls contain sheets of microfibrils arranged parallel to each other. Several such sheets are embedded in different orientations in a matrix of hemicellulose and lignin. Lignocellulose is the term used to describe this complex amalgam of plant structural polysaccharides.

The structure of hemicellulose and the xylanases needed for its degradation have been reviewed by Chesson & Forsberg (1988) and Thomson (1993). Glucuronoarabinoxylans are the main component of the hemicellulose in the primary cell walls of monocotyledonous plants and in the secondary cell walls of both mono- and dicots. They consists mainly of ß-1,4 linked units of xylopyranosyl carrying side-chains of glucuronic acid units and short branched chains of α-1,3 and 1-5 linked arabinofuranose residues. The degree of polymerization and the nature and frequency of the side chains vary in different plants, especially between mono- and dicots. Hemicellulose is generally weaker than cellulose and less resistant to hydrolysis. Pectic substances can also be regarded as hemicellulose. These polysaccharides consist of α-1,4 linked galacturonic acid units esterified at their carboxylic acid groups with methanol or substituted at carbon-2 with acetyl groups. Pectic substances are found in all plants, but especially in dicots, where they accumulate in the middle lamella (region between adjacent cell walls, which does not contain cellulose). Pectinolytic degradation can lead to the maceration of plant tissues.

The lignins are complex polymers of variously substituted phenylpropane units, which can be cross-linked with the hemicellulose (Chesson & Forsberg, 1988). Lignins are more resistant to enzymatic digestion than cellulose and hemicellulose, especially under anaerobic conditions (Hungate, 1988). The enzymatic degradation of lignin generally involves free-radical oxidative mechanisms and usually takes place in aerobic environments. The white rot fungus *Phanerochaete chrysosporium* is one of the few microorganisms able to degrade lignin completely by producing a series of high-potential peroxidase enzymes (some of which produce smaller and more diffusible one-electron oxidants such as Mn^{III} complexes) which act purely as one-electron oxidants (Sinnott, 1990). Some microorganisms can gain access to cellulose through partial delignification of the substrate (Béguin & Aubert, 1994). Lignin is formed when its major precursors, p-coumaryl, coniferyl and sinapyl alcohols, are released into the cell wall and polymerized in a random fashion *in situ* to form a three-dimensional macromolecule (Chesson & Forsberg, 1988). This results in the formation of at least 20 different types of linkages. The lignin decreases the porosity of the substrate and limits the enzymatic degradability of cell walls

by preventing the diffusion of the cellulolytic and hemicellulolytic enzymes into the cell walls. This limits the enzymatic attack to the surface of feed particles heavily impregnated with lignin. Another factor which limits the accessibility of the cellulose to the enzymes is the presence of epicuticular waxes and the cuticle. Access by cellulolytic microorganisms is limited to broken edges of feed particles or naturally occurring openings like stomata and lenticels.

1.3 Cellulolytic microorganisms

Many cellulolytic microorganisms are found in a wide variety of environments in which cellulosic materials accumulate. These microorganisms, mainly bacteria and fungi, produce cellulases and hemicellulases to degrade the complex substrate. Only a few microorganisms can achieve the complete degradation of native cellulose, but many can produce incomplete cellulolytic enzyme systems. The interactions between microorganisms possessing different cellulase systems are important for the degradation of plant materials by mixed consortia existing in e.g. the rumen or in compost. Plants also produce cellulases which are mainly associated with developmental processes e.g. maturation of fruits and abscission of leaves.

The aerobic degradation of cellulose in topsoil has been studied in a number of fungi, e.g. *Phanerochaete chrysosporium* and *Trichoderma reesei*, and bacteria, e.g. *Cellulomonas*, *Pseudomonas*, *Thermomonospora* and *Microbispora* (Béguin & Aubert, 1994). Lignin degradation takes place mainly under aerobic conditions. Anaerobic degradation of cellulose coupled with the fermentation of the breakdown products has been studied mainly in the bacterium *Clostridium thermocellum*, a number of rumen bacteria, e.g. *Fibrobacter*, *Butyrivibrio* and *Ruminococcus* spp., and in the rumen fungi *Neocallimastix frontalis* and *N. patriciarum* (Béguin & Aubert, 1994).

1.4 Classification of cellulases and xylanases

Cellulases are classified into three major groups depending on their substrate specificities: ß–1,4-endoglucanases cleave internal ß-1,4-glycosidic bonds; cellobiohydrolases, also

called exo-cellulases, release cellobiose from the nonreducing end of cellulose; and ß-glucosidases hydrolyze cellobiose to glucose (Gilbert & Hazlewood, 1993). This general classification, however, does not do justice to the cellulases, because it ignores the multiplicity of cellulases. Multidomain enzymes and also the relatively broad substrate specificities of some enzymes pose a big problem . The new classification of glycosyl hydrolases, which includes all cellulases, has recently been reviewed to incorporate the latest developments (Henrissat & Bairoch, 1993; Workshop on nomenclature of cellulases, 1994). Up to date this new classification system contains about 600 sequences of glycosyl hydrolases grouped into 48 families based on sequence similarities of catalytic domains. Hydrophobic cluster analysis was used for the comparison of the amino acid sequences of the enzymes; this means that the proteins were compared by their secondary structures (Henrissat *et al.*, 1989; Gilkes *et al.*, 1991; Henrissat, 1991). The three dimensional folding patterns are better conserved than the primary sequences of proteins, therefore members of a familiy are expected to have the same fold. The catalytic residues are conserved within these families and the stereochemistry of the reactions catalyzed, i.e. retaining or inverting mechanism, is consistent. Cellulases and xylanases are ordered into 11 families. Some of these families contain enzymes originating from a range of organisms spanning different kingdoms (Gilbert & Hazelwood, 1993).

Xylanases, or hemicellulases, are required by most cellulolytic microorganisms to gain access to the cellulose in the plant cell walls. Several bacteria, e.g. *C. thermocellum* and *Fibrobacter succinogenes* (formerly *Bacteroides*), produce significant xylanase activity, but do not utilize the xylan or the degradation products for growth (Stewart & Bryant, 1988; Morag *et al.*, 1990; see section 2.4.1.3). Since hemicellulose contains a wider variety of linkages than cellulose, a large number of hemicellulases with different substrate specificities can be found: these include ß-D-xylanases, ß-D-galactanases, ß-D-mannanases, acetylesterases, α-arabino-furanosidases, α-glucuronidases and ß-xylosidases (Thomson, 1993). The hemicellulases, just like the cellulases, are classified into a number of families of glycosyl hydrolases based on sequence similarities (Henrissat & Bairoch, 1993).

1.5 Fungal cellulases

Bacterial and fungal cellulases share many characteristics with regard to molecular structure and enzymatic mechanisms. Fungal systems have been investigated for a longer time, but recently bacterial cellulases have been investigated more intensely, because they have been reported to have higher specific activities (Johnson *et al.*, 1982). No recent comparative studies are however available and it is for instance not certain whether bacterial cellulases are more active than those from anaerobic fungi like *Neocallimastix* or *Piromyces*. Since most of this review deals with bacterial cellulases, a brief description of fungal cellulolytic systems will be given. In comparison to bacterial cellulases, which are intracellular, cell associated, or secreted and can occur individually or as multi-enzyme complexes, cellulases from aerobic fungi are mostly secreted and occur individually, i.e. they form non-aggregating systems (see below), while cellulases from anaerobic fungi also aggregate. Fungal cellulases have received a great deal of attention, because they are secreted in large quantities and up to 20 g/l have been obtained from *Trichoderma reesei* (Bailey & Nevalainen, 1981; Divne *et al.*, 1994). Cellulases from *T. reesei* and the white rot fungus *Phanerochaete chrysosporium* have received the most attention. Some others are from *Fusarium solani, Penicillium funiculosum, Talaromyces emersonii* and the rumen fungus *Neocallimastix frontalis* (Coughlan & Ljungdahl, 1988; Wood *et al.*, 1988). The main cellulolytic components of aerobic fungi are endoglucanases, cellobiohydrolases (CBH) and ß-glucosidases. Other enzymes found in some fungi are glucohydrolases, cellobiose oxidases and cellobiose dehydrogenases (Wood, 1992). Two genes encoding endoglucanases have been cloned from *T. reesei*. The cellobiohydrolases of *T. reesei* and *Penicillium pinophilum* were found to exist in mainly two forms, CBH I and CBH II, constituting approximately 60 % and 20 % of the extracellular protein, respectively. Both CBHs exist in several iso-forms and consist of two domains, a catalytic domain separated by a linker region from the cellulose-binding domain. They have slightly different but overlapping substrate specificities and hydrolysis mechanisms, but the debate about which enzyme has more of an endo- or exo-type action still goes on. The determination of the three-dimensional structures of the catalytic domains from CBH I and CBH II has helped to answer some of these questions (see below).

1.6 Non-aggregating and aggregating cellulase systems

1.6.1 Non-aggregating systems

In these systems the cellulolytic enzymes are secreted and do not form large multi-enzyme complexes. Fungal cellulase systems generally follow this pattern, and some bacteria are speculated to have analogous systems i.e. *Pseudomonas flourescens* subsp. *cellulosa*, *Clostridium stercorarium* and *Cellulomonas fimi* (Béguin *et al.*, 1992; Hazlewood *et al.*, 1992; Meinke *et al.*, 1994). A simplified model of cellulose degradation in such systems has been generally accepted, in which endo-cellulases, exo-cellulases (cellobiohydrolases) and ß-glucosidases act in synergy. The endo-acting enzymes hydrolyze bonds in the amorphous regions of the cellulose which leads to the formation of numerous non-reducing ends. The exo-acting enzymes act on the newly created non-reducing ends and continue to release cellobiose from the crystalline regions of the cellulose. The ß-glucosidases are important because they remove the cellobiose which progressively inhibits the cellobiohydrolases (Gilbert & Hazlewood, 1993; Béguin & Aubert, 1994). This simplified model does not take into account the variable endo- and exo-glucanase character of many cellulases, which varies depending on the substrate and the reaction conditions. It also does not explain the phenomena of endo-endo or exo-exo synergism, i.e. the degradation of crystalline cellulose by the cooperative action of two endo- or two exo-glucanases (Coughlan & Ljungdahl, 1988; Gilbert & Hazlewood, 1993). One possible explanation for the apparent synergism between two cellobiohydrolases could be competitive adsorption or the formation of more efficient binary complexes between CBH I and CBH II which could result in an increased turnover of the enzymes. Simple multi-enzyme complexes have been observed in *T. reesei* which could account for these synergistic effects (Béguin & Aubert, 1994). Another explanation could be the different stereochemical conformations presented to the enzymes by the cellulose molecule; the removal of successive cellobiose units from one cellulose molecule would expose on the neighboring chain a target site requiring a different stereochemical specificity (Wood, 1992). It also seems that in some cases the purified CBH enzymes used for exo-exo synergism studies were still contaminated with small quantities of endoglucanase activities which led to some anomalous results.

1.6.2 Aggregating cellulase systems

A number of bacteria synthesize a variety of cellulolytic and structural proteins, which are secreted and assembled into a multi-enzyme complex on the cell surface. These complexes play important roles in the attachment to and the degradation of plant material. Such complexes have been observed in a variety of anaerobic bacteria, including *C. thermocellum*, *C. cellulovorans*, *C. papyrosolvens* C7, *C. cellulolyticum*, *R. albus*, *Fibrobacter succinogenes*, *Bacteroides cellulosolvens*, *Butyrivibrio fibrisolvens* H17c and in the rumen fungi *Neocallimastix frontalis* and *N. patriciarum* (Fierobe *et al.*, 1991; Lin & Thomson, 1991; Shoseyov *et al.*, 1992; Béguin & Aubert, 1994; Pohlschröder *et al.*, 1994; Zhou *et al.*, 1994).

The cellulosome of *C. thermocellum* is the best characterized example of this kind of system (Felix & Ljungdahl, 1993; Béguin & Aubert, 1994). Electron microscope studies revealed cell-associated cellulosomes and polycellulosomes having molecular masses of 2-6,5 MDa and 50-100 MDa, respectively. These complexes are thought to be assembled inside "skin-like" covers on the cell surfaces. Upon contact with the substrate this covering attaches the bacteria with the cellulosomes to the substrate. The products of cellulolysis are channeled through these "contact channels" to the cells where they seem to be taken up by active transport. The cellobiose and cellodextrins are thought to be metabolized by cytoplasmic cellobiose phosphorylases and cellodextrin phosphorylases (Felix & Ljungdahl, 1993). Periplasmic and cytoplasmic ß-glucosidases are present, but they are not part of the cellulosome nor are they found in the culture supernatant (Gräbnitz *et al.*, 1991; Katayeva *et al.*, 1992). During later growth phases the bacteria detach themselves from the substrate, but the cellulosomes and polycellulosomes stay attached and continue the cellulolysis. Eventually the complexes disintegrate into separate polypeptides suspended in the culture supernatant, which can still hydrolyze soluble, but not crystalline, substrates (Felix & Ljungdahl, 1993).

The cellulosomes consist of 14 to 26 polypeptides as revealed by SDS-PAGE. Up to date the genes encoding 15 endoglucanases, two xylanases, two ß-glucosidases and one lichenase have been cloned from *C. thermocellum* NCBI 10682. The nucleotide sequences for nine endoglucanase genes, one xylanase gene, one lichenase gene and the *celS* cellobiohydrolase (CBH) have been determined (Béguin & Aubert, 1994). The genes seem to be scattered throughout the bacterial genome and function as monocystronic transcription units (Béguin *et al.*, 1988). The cellulosome contains mainly endoglucanases,

but another major cellulosome component identified and sequenced recently is the CelS CBH (Wang *et al.*, 1993a).

CipA (for cellulosome integrating protein, formerly S_L or S1) is a cellulose-binding protein to which the other catalytic components of the cellulosome bind (Hall *et al.*, 1988; Tokatlidis, 1991; Fujino *et al.*, 1992; Salamitou *et al.*, 1992). The DNA sequence of *cipA* revealed nine internal repeated elements of about 500 nucleotides each (Gerngross *et al.*, 1993). These nine repeated elements of CipA interact with the conserved 22 amino acid residue repeats present in the various cellulases which are part of the cellulosome. CipA also contains a type 3 cellulose-binding domain, thus CipA serves as scaffolding protein and cellulose binding factor in the organization of the cellulosome. CipA also contains one of the conserved 22 residue repeats at the COOH terminus which were speculated to be important for the concatenation of CipA subunits or in the anchoring of the cellulosome to the cell surface (Fujino *et al.*, 1993; Beguin & Aubert, 1994). Fujino *et al.* (1993) found an ORF3 3' from the *cipA* gene which encoded a region similar to the reiterated segments of CipA as well as elements which are also present in the S-layer proteins of *Bacillus brevis* and *Acetogenium kivui*. It was speculated that the function of cell bound protein ORF3p was to attach the cellulosome to the cell surface by binding to the 22 residue repeat at the COOH terminus of CipA. How the cellulosomes are assembled and whether they have a defined stoichiometry and topology is unknown. Since 11 cellulases with the 22 residue repeats are already known, but there are only 9 "docking" sites on CipA, it seems as if the cellulosomes can have at least moderately variable designs (Béguin & Aubert, 1994). The exact mechanism of cellulolysis is also unknown and the relative importance of the endoglucanases and the CelS cellobiohydrolase is debated.

1.7 Functional domains of cellulases

Sequence analysis of cellulases and xylanases has revealed the modular structure of these enzymes. The different domains have been reviewed by Gilkes *et al.* (1991a) and Béguin & Aubert (1994) and will not be discussed in great detail in this review. Catalytic domains connected to cellulose-binding domains (CBDs) by "flexible" linker regions are commonly found. Other domains assist in binding of the cellulases to multi-enzyme complexes, or in the attachment to the bacterial cell surface. The extents and functions of these domains have in many cases been defined by partial proteolysis and by the cloning and expression

of fragments of cellulase genes (Tomme *et al.*, 1988; Gilkes *et al.*, 1988, 1989; Jauris *et al.*, 1990; Ferreira *et al.*, 1990, 1991; Couthino *et al.*, 1992; Maglione *et al.*, 1992; Ramalingam *et al.*, 1992; Goldstein *et al.*, 1993; Ong *et al.*, 1993)

1.7.1 Catalytic domains

To hydrolyze the different linkages found in lignocellulose a variety of catalytic domains with different secondary structures have evolved. Cellulases have been classified into 11 families initially on the basis of sequence homologies of the amino acid sequences of their catalytic domains and secondly by hydrophobic cluster analysis, which reveals similarities in apparent secondary structure between proteins showing very low sequence homology (Henrissat, 1989; Henrissat & Bairoch, 1993). Whether this classification accurately reflects the number of different three-dimensional designs of catalytic domains is debatable. Cellulases belonging to different families might share common structural features or have similar structures.

1.7.2 Cellulose-binding domains

These domains have been reviewed extensively by Gilkes *et al.* (1991a) and Béguin & Aubert (1994). CBDs have been grouped into four families. Type 1 CBDs have been found in a variety of bacterial cellulases, including xylanases, an arabionfuranosidase and a chitinase. These CBDs have a length of about 100 residues and contain four conserved tryptophan residues which were shown to be important for the adsorption to the cellulose substrate, but the exact binding mechanism is not yet clear (Poole *et al.*, 1993; Din *et al.*, 1994). Type 2 CBDs have only been found in fungal cellulases. These CBDs have a length of about 30 residues and the 3-D structure of the *T. reesei* CBH I CBD has been determined by NMR spectroscopy (Kraulis *et al.*, 1989). One side of this wedge-shaped domain is hydrophilic and contains three tyrosine residues which are thought to interact with the cellulose substrate. Type 3 CBDs have a length of about 130 residues and have been found in a variety of bacterial cellulases and in cellulosome scaffolding proteins e.g. CipA from *C. thermocellum*. Type 4 CBDs comprise a new family and have so far only been found in three cellulases (Béguin & Aubert, 1994).

The presence of CBDs in cellulases has in many cases been linked to the ability to degrade crystalline cellulose. Removal of the CBDs usually led to increased activity on soluble cellulosic substrates, like CMC, and decreased activity on crystalline substrates (Béguin & Aubert, 1994). There are however enzymes which possess a CBD but have no activity against crystalline cellulose e.g. the closely related endoglucanases CelE from *C. thermocellum*, CelA from *C. longisporum* and CelA from *R. albus* (CelA grafted onto CBD from *P. fluorescens* subsp. *cellulosa* xylanase C), and cellodextrinase CELC from *P. fluorescens* subsp. *cellulosa* (Durrant *et al.*, 1991; Ferreira *et al.*, 1991; Poole *et al.*, 1991; Mittendorf & Thomson, 1993). It is speculated that the CBDs anchor the cellulases to the substrate and the catalytic cores at the other end of the flexible linker domains are able to hydrolyze several glycosidic bonds in a localized area (Béguin & Aubert, 1994). The diffusion of the cellulases is limited and the hydrolytic activity is concentrated in some places. A more active role for CBDs has also been proposed where the binding of the CBD destabilizes the structure of the crystalline cellulose. The CBD from *C. fimi* CenA has been shown to cause changes in the microstructure of cellulose fibres (Din *et al.*, 1991).

1.7.3 Linkers

Functional domains in cellulases are usually linked by short linker sequences with variable lengths (6 to 59 residues) and little sequence homologies between different enzymes (reviewed by Gilkes *et al.*, 1991a). Some contain a high proportion of hydroxyamino acids or short repeats. Some linkers are rich in aspartate or glutamate residues or both. The main function of these linkers is to spatially separate the various domains of the cellulase and maybe to orientate the catalytic domain relative to the CBD and to the substrate. An analysis of the linker of endoglucanase CenA from *C. fimi* suggested that the linker has an extended, apparently rigid conformation and it was also shown that the deletion of the linker reduced the catalytic activity of the enzyme on a range of substrates (Shen *et al.*, 1991). In many cases proteases cleave cellulases in or near their linkers (Tomme *et al.*, 1988; Gilkes *et al.*, 1988, 1989; Jauris *et al.*, 1990). Glycosylation of exoglucanase Cex in *C. fimi* has been shown to protect the enzyme from proteolytic cleavage between the functional domains (Langford *et al.*, 1987). Subsequently it was shown that only the linker of Cex was glycosylated when the enzyme was produced in *Streptomyces lividans* (Ong *et al.*, 1994).

1.7.4 Multi-catalytic domain proteins

Recently enzymes with more than one catalytic domain have been identified (Béguin & Aubert, 1994). The *celB* gene from the extreme thermophile *Caldocellum saccharolyticum* encodes a large polypeptide with an exoglucanase domain at the amino terminus and an endoglucanase domain at the carboxy terminus separated by linkers and a central domain with unknown function (Saul *et al.*, 1990). The ß-mannanase gene from the same bacterium encodes an enzyme containing a domain with ß-mannanase activity, an endoglucanase domain and two binding domains, all of which are separated by linkers (Gibbs *et al.*, 1992). The *celD* gene (cDNA clone) from the rumen fungus *Neocallimastix patriciarum* encodes three separate domains with endoglucanase, cellobiohydrolase and xylanase activities (Xue *et al.*, 1992a). The *xynA* and *xynD* genes from *R. flavefaciens* 17 encode bifunctional enzymes, *xynA* encoding two xylanase domains from separate families and *xynD* encoding a xylanase and a ß-(1,3-1,4)-glucanase (Zhang & Flint, 1992; Flint *et al.*, 1993). It is unclear how significant the role of these multi-functional enzymes is in the degradation of plant material.

1.7.5 Repeated segments

22-24 Residue repeats: The 22 amino acid residue repeated sequences are necessary for the assembly of cellulases into the cellulosomes (see above). These repeats have so far been found in 12 cellulase genes in *C. thermocellum*, five genes in *C. cellulolyticum* and one gene in *C. cellulovorans*. All of these anaerobic bacteria form multi-enzyme cellulolytic complexes (Béguin & Aubert, 1994).

Domains similar to fibronectin type III domains: These motifs of approximately 100 residues have been found in seven genes in several bacteria including *C. fimi* and *Bacillus lautus*. In the chitinase A1 gene of *B. lautus* these repeats may form part of a chitinase-binding site. The repeats in CenB and CenD of *C. fimi* are speculated to have a similar function or may be involved in protein-protein interactions, but their exact functions are unknown (Gilkes *et al.*, 1991; Meinke *et al.*, 1993; Béguin & Aubert, 1994).

Motifs similar to S-layer proteins: An endoglucanase from *Bacillus* sp. KSM 635 and a xylanase from *Thermoanaerobacter* each contain three copies of this motif which is present in

S–layer proteins from *B. brevis* and *Acetogenium kivui*. The S–layer proteins are components of the cell wall and it is thought that these motifs anchor the proteins in the S–layer. Three ORFs in *C. thermocellum* located 3' of the *cipA* gene were found to contain 3 copies of this motif each. The functions of these ORFs is unknown, but it is speculated that ORF3p serves to anchor the cellulosome in the cell wall (Fujino *et al.*, 1993; Béguin & Aubert, 1994).

1.8 3-D structures and catalytic mechanisms

Three-dimensional structures of several cellulases have been determined recently, e.g. cellobiohydrolase CBH II from *T. reesei* (Rouvinen *et al.*, 1990), CelD from *C. thermocellum* (Juy *et al.*, 1992), endoglucanase E2 from *Thermomonospora fusca* (Spezio *et al.*, 1993), endoglucanase V from *Humicola insolens* (Davies *et al.*, 1993), endo-ß-1,4-xylanase II (Törrönen *et al.*, 1994) and cellobiohydrolase CBH I from *T. reesei* (Divne *et al.*, 1994). Several more are being crystallized for X-ray diffraction analysis at the moment, e.g. Cex from *C. fimi* (Bedakar *et al.*, 1992), endoglucanase I from *H. insolens* (Davies *et al.*, 1992) and CelCCA from *C. cellulolyticum* (Roig *et al.*, 1993). The crystal structures of the catalytic domains have contributed to the understanding of the hydrolytic activities of these cellulases. The active site of CelD has six substrate-binding sites, labelled A to F from the non-reducing end of the substrate, and the presumed catalytic residues Glu 555 and Asp 201 are in the vicinity of the glycosidic bond between subsites D and E (Juy *et al.*. 1992). Site-directed mutagenesis has been used in the case of CelD to confirm the identity of the catalytic residues (Chauvaux *et al.*, 1992). The design of the active site was found to be responsible for the endo- or exo-modes of cleavage, e.g. CBH II and endoglucanase E2 both belong to family 6 and have similar secondary structures, but differ in their active site accessibilities to the substrate (Spezio *et al.*, 1993). In CBH II two loops enclose its active site in a tunnel thus only permitting the substrate to be threaded through the tunnel which leads to the release of cellobiose from the non-reducing end of the chain. In E2 the active site is an open cleft, approximately 11 Å deep running across the entire molecule, to which the substrate polymer binds. One of the loops present in CBH II is missing in E2 while the other one might have some flexibility. This permits the substrate polymer free access to the active site across the whole length of it and results in random endocleavages. The catalytic domain of CBH I has a 40 Å long active site tunnel containing seven glucosyl

binding sites. The reducing end of the cellulose chain is thought to enter at subsite G and cellobiose units are cleaved from the end of the chain at the other end of the tunnel between sites B and C. Due to interactions between the chain and residues lining the active site it is retained in the active site for further catalytic cycles. This explains the distribution of CBH I at the crystalline face of cellulose particles in contrast to CBH II which is mostly found at the non-reducing end of cellulose chains. Aromatic residues, especially tryptophan, are situated in the walls of the active site tunnels of some cellulases (e.g. CBH I and CBH II) and in cellulose-binding domains, where they are thought to interact with the carbohydrate substrate through hydrogen bonding along the planar surfaces of glucosyl-residues (Divne *et al.*, 1994).

Two types of hydrolytic reaction mechanisms can be distinguished, the "retaining" and "inverting" ones. Both mechanisms are nucleophilic substitutions at the saturated carbon of the anomeric center, which can lead to either the retention or the inversion of the chemical groups at the anomeric carbon (Sinnott, 1990). Two residues are involved in the general acid-base mechanism. The first residue acts as a general acid by protonating the oxygen of the glycosidic-bond (the ß-1,4-bond). The second residue acts as a nucleophile by either stabilizing the resulting oxocarbonium ion which is formed in the retaining reaction or by deprotonizing the H_2O to form the OH^- which binds to the anomeric carbon in the inverting mechanism. The retaining mechanism is a two-step reaction while the inverting mechanism only involves one step (Béguin & Aubert, 1994).

1.9 Regulation of cellulase biosynthesis

Two modes of regulation of cellulase synthesis can be distinguished. Repression occurs when more easily metabolizable carbon sources are added to cultures growing on cellulose. Induction occurs when cellulose is the only carbon source available (Béguin & Aubert, 1994). Our knowledge of molecular details of the regulation of cellulases in almost all cellulolytic microorganisms is very scant. Most of the regulatory data consist of comparisons of individual or collective cellulase activities or the quantification and characterization of mRNA transcripts using probes derived from cloned genes. Very few regulation studies have used neutral media, which allow growth without influencing cellulase regulation, making the interpretation of the data difficult (Béguin & Aubert,

1994). Another factor complicating regulation studies has been the lack of a gratuituous inducer of cellulase gene transcription.

The cellular mechanisms which lead to the observed repression or induction effects are mostly unknown, i.e. the nature of the intracellular inducers and the participating regulatory proteins have not been determined. The involvement of the phosphoenolpyruvate-dependent phosphotransferase system (PTS) in the transport of hexoses and pentoses by ruminal bacteria, e.g. *Streptococcus bovis* and *Selenomonas ruminantium*, has been demonstrated (Martin & Russel, 1986, 1987; Strobel, 1993), but it is not clear if and how the PTS system is involved in the regulation of synthesis of cellulases in various cellulolytic microorganisms. It is possible that a PTS-regulated system similar to the one in enteric bacteria, involving cAMP and the cAMP receptor protein (CRP), could be responsible for the transcriptional control of cellulase genes (Postma *et al.*, 1993; Saier & Reizer, 1994). In *Bacillus subtilis* HPr, a central PTS regulatory protein, and CcpA, a putative DNA-binding protein, appear to mediate one form of catabolite repression of various operons encoding both non-PTS and PTS catabolic enzymes and transport proteins (Deutscher *et al.*, 1994; Saier & Reizer, 1994). It was found that in *B. subtilis ccpA* mutants the *bglS* gene, encoding the ß-(1,3-1,4)-endoglucanase, and the *xynB* gene, encoding the ß-xylosidase B, were no longer subject to catabolite repression by glucose (Krüger *et al.*, 1993; Lindner *et al.*, 1994). The cloning and sequencing of a *C. longisporum* operon encoding several PTS proteins, including an enzyme II for aryl-ß-glucoside transport, a phospho-ß-glucosidase and an antiterminator protein, indicates the presence of a PTS system in this bacterium, where it might be implicated in the regulation of the cellulase genes (Gordon Brown, personal communication).

The recognition of the extracellularly located large insoluble substrate and the induction of genes necessary for its degradation and assimilation pose a considerable riddle to researchers. Chemotaxis of cellulolytic microorganism towards the insoluble carbon source, which would probably involve the same cellular mechanisms, is also not well understood (Hsing & Canale-Parola, 1992). The generally accepted mechanism for induction is that cellulose and hemicellulose first undergo limited hydrolysis by cellulases and xylanases produced in low constitutive amounts. The soluble hydrolysis products are taken up by the cell either actively or passively where they can be enzymatically modified to generate the active inducer for transcription of cellulase genes (Béguin & Aubert, 1994). It is also possible that membrane-bound receptor-like binding sites (chemoreceptors) for

certain cellulose breakdown products are involved, similar to the receptors for fungal elicitors (e.g. ß-1,6 and ß-1,3-linked glucans) in plant-microbe interactions (Yoshikawa & Sugimoto, 1993).

Regulation of fungal cellulases has been reviewed recently by Kubicek *et al.* (1993). A model is proposed for *Trichoderma* in which conidial bound constitutively expressed cellobiohydrolases carry out the initial exo-exo attack on the cellulose. The resulting disaccharides (cellobiose and cellobiono-1,5-lactone) are then taken up by the mycelia where they are possibly metabolized to the true inducers, which then induce cellulase synthesis. A ß-linked disaccharide permease is crucial for the transport of cellobiose and other disaccharides into the mycelia. The cellobiose may be metabolized further to form the "true inducer", but ß-glucosidases cleave most of the cellobiose. Therefore it is speculated that other disaccharides, which resemble cellobiose, but are poor substrates for the ß-glucosidase, may be true cellulase inducers. Sophorose (ß-1,2-glucosyl-glucose) has for long been considered the true inducer (Mandels *et al.*, 1962), but there is contrary evidence e.g. it does not induce the full complex of cellulases (Béguin & Aubert, 1994). Sophorose can be formed via transglycosylation by the ß-glucosidase or endoglucanase I (EG I). Cellobiono-1,5-lactone has also been shown to induce cellulases, but it is speculated that this might be an indirect effect due to the inhibition of ß-glucosidase (Szakmary *et al.*, 1991).

T. reesei has two separate ß-glucosidases, one is extracellular and one is intracellular (Kubicek *et al.*, 1993). The function of the sophorose-inducible intracellular enzyme is not clear, but in *Penicillium purpurogenum* the intracellular ß-glucosidase has been implicated in the formation of the inducer gentiobiose from cellobiose (Kurasawa *et al.*, 1992). A similar scenario exists in *C. thermocellum* which produces a periplasmic ß-glucosidase A and a cytoplasmic ß-glucosidase B (Katayeva *et al.* 1992). Both are produced constitutively during growth on glucose, but the synthesis of only ß-glucosidase B is repressed by cellobiose. It is speculated that the function of ß-glucosidase A is to effect the continuous hydrolysis of cellobiose to relieve cellobiohydrolase inhibition, while ß-glucosidase B has to form the true inducer of cellulase activity by transglycosylation.

1.10 Transcription of cellulase genes

Most cellulase genes have been found to be transcribed as independent monocistronic units and stemloop terminator structures have been identified at the 3'-ends of many cellulase genes e.g. *celA*, *celD*, and *celF* from *C. thermocellum* (Béguin *et al.*, 1986; Mishra *et al.*, 1991); *cenA*, *cenB*, *cenC*, *cenD*, *cbhA* and *cex* from *C. fimi* (Greenberg *et al.*, 1987a, 1987b; Moser *et al.*, 1989; Meinke *et al.*, 1993, 1994); and the EgI gene from *R. albus* F-40 (Ohmiya *et al.*, 1989). In *C. fimi* the *cbhA* and *cenD* genes are linked and the inverted repeats at the 3'-end of *cbhA*, which form a putative stemloop terminator structure, overlap typical *C. fimi* promoter sequences at the 5'-end of *cenD*; it is not clear whether these two genes are transcribed together in a dicistronic mRNA. In *C. cellulolyticum* a cellulase gene cluster was found to encode an unidentified open reading frame, ORF1, and three endoglucanase genes, *celCCC*, *celCCG* and *celCCE*. These clustered genes seem to be transcribed in a polycistronic message (Bagnara-Tardif *et al.*, 1992). A lichenase and a xylanase, which are separated by 155 bp, are clustered in *Bacillus polymyxa* (Gonsalbes *et al.*, 1991). In *Phanerochaete chrysosporium* three cellobiohydrolase genes, *cbh1-1*, *cbh1-2* and *cbh1-3*, share homology with each other and form a cluster, but they are regulated separately (Covert *et al.*, 1992). Southern blots of genomic DNA suggested that there are as many as six *cbh1*-like genes in this fungus. The multiplicity of *cbh1*-like genes is unusual and *T. reesei* has been reported to contain only one copy.

Transcription initiation sites of several cellulase genes have been determined. In the *R. albus* EgI gene transcription is initiated 59 bp upstream of the ATG codon in both the parent strain and *E. coli* (Ohmiya *et al.*, 1989). On the other hand *C. thermocellum celA* and *celD* (Béguin *et al.*, 1986; Mishra *et al.*, 1991) and *C. fimi cenA* and *cenB* (Greenberg *et al.*, 1987a, 1987b) were shown to make use of two different promoters depending on the host or on the culture conditions. In *C. fimi* high levels of transcription of *cenB* from the $cenBp_1$ promoter took place in cells grown in the presence of CMC, but $cenBp_2$ was used for the constitutive expression of *cenB* in the presence of glucose or glycerol.

Very little is known about molecular interactions at the promoters of cellulase genes. Inverted 14–bp repeats upstream of a number of cellulase genes from *Thermomonospora fusca* and the actinomycetes *Streptomyces halstedii* and *Streptomyces lividans* showing homology with each other have been reported (Lao *et al.*, 1991; Fernándes-Abalos et al., 1992; Jung et al., 1993). In *T. fusca* a protein that binds to the 14-bp inverted repeat

upstream of *celE* has been identified. The binding activity was only present when cellulase synthesis was induced and therefore resembled an activator protein (Lin & Wilson, 1988). The *R. albus* EgI gene contains an 11–bp inverted repeat 44 bp upstream from its transcription initiation site (Ohmiya *et al.*, 1989). The function of the *cbh2* promoter from *T. reesei* was studied in the original host by making a reporter gene construct utilizing the *E. coli uidA* gene, which encodes a ß-glucuronidase (Stangl *et al.*, 1993). Induction of the reporter gene was observed in the presence of cellulose and sophorose and during sporulation. At least two DNA-protein binding areas were identified and it seemed that some of the DNA-binding proteins were also cellulose-inducible. This study highlights the usefulness of suitable genetic transfer systems to investigate cellulase functions in the original hosts. Unfortunately these genetic tools are as yet unavailable in most cellulolytic microorganisms.

1.11 Applications of cellulases

Current and envisaged applications of lignocellulolytic enzymes have been reviewed recently (Gilbert & Hazlewood, 1993; Béguin & Aubert 1994), therefore only a few examples will be discussed. Most of the applications involve the partial hydrolysis of the substrate, e.g. the pre-treatment of cellulosic feedstocks to improve the nutritional quality and digestibility of animal feeds. The degradation of ß-glucans in poultry feeds decreases digestive problems due to viscosity of the polymer (Walsh *et al.*, 1993). The improvement of silage by increasing microbial digestion of fibre during the silage process is being investigated (Flores, 1991) and it was shown that a genetically modified silage lactic acid bacterium, *Lactobacillus plantarum*, expressing extracellular cellulase genes from *C. thermocellum*, is able to proliferate during the ensilage process (Sharp *et al.*, 1994). The genetic modification of various traits of rumen bacteria, other than their cellulolytic systems, is also being considered, e.g. to increase lactate utilization and to increase levels of lysine and methionine in microbial protein (Orpin, 1988; Wallace, 1992). Biobleaching of paper pulp with hemicellulases, to reduce the amount of chlorine utilized in the normal bleaching process, has led to the investigation of thermostable and alkaline xylanases (Simpson *et al.*, 1991; Arase *et al.*, 1993; Nakamura *et al.*, 1993). A completely different approach to the same problem of reducing the bleaching requirement of wood pulps, is the genetic modification of lignin biosynthetic pathways in plants to decrease their lignin contents (Shibata *et al.*, 1993). The expression of a *R. albus* endoglucanase gene in tobacco

plant cells could also serve similar purposes (Kawazu *et al.*, 1993). Cellulases are being used in the textile industry and in detergents for "bio-stoning" and "bio-polishing" to modifiy the texture of materials (Sato *et al.*, 1993). Recombinant yeast strains expressing cellulases and hemicellulases will soon be a reality in wine making processes and for the production of various biological compounds (Pérez-González *et al.*, 1993). Potential uses for cellulosomes as docking proteins in complexes similar to avidin-biotin complexes have been predicted by Bayer *et al.* (1994). A heterologous endoglucanase from *C. cellulovorans* was cloned and expressed in the solventogenic bacterium *Clostridium acetobutylicum* with the aim of improving its cellulolytic capabilities to make the acetone-butanol fermentation economically viable (Kim *et al.*, 1994). The extensive degradation of the cellulosic content of agricultural and domestic wastes and the coupled fermentation of the products to fuels, e.g. ethanol, will however still require a lot of basic research to improve factors like cheap substrate-pretreatment processes and the improvement of the efficacity of cellulases and the fermentation (Demain & Lynd, 1993; Béguin & Aubert, 1994).

1.12 Lignocellulose degradation in the rumen

Ruminants cannot produce cellulases and have to rely on rumen microorganisms for the fermentation of plant structural carbohydrates. A mutualistic relationship exists between the ruminant and the microbial population in its rumen in which the ruminant provides the optimal growth environment for the microbes in return for volatile fatty acids and microbial cells which represent the energy source and the proteinaceous food for the animal. The major volatile fatty acids produced by the microorganisms are acetic, propionic, butyric and higher acids (Hungate, 1988). The ruminant maintains the anaerobic rumen environment at a temperature of 39–40°C, it stabilizes the pH at about 6.0-6.7 through the copious secretion of $NaHCO_3$ in the saliva, it regularly supplies the rumen with plant tissue and it assists in the fermentation of it by ruminating the cud and by mixing the rumen contents with contractions of the rumen wall (Hungate, 1988). The rumen is essentially a blind sac linked to the alimentary tract and its pre-peptic location (preceding the stomach) makes the host completely dependent on the rumen bacteria because they ferment all ingested food. This is in contrast to a post-peptic location (an enlarged caecum usually) which enables the host to absorb any digestive products prior to the microbial fermentation (Hungate, 1988). The compartmentalization of the rumen into five distinct regions is important for the fermentation. Different populations of the rumen

microflora are found in these compartments associated with feed particles, the rumen wall and the liquid phase of the rumen contents (Stewart & Bryant, 1988). The colonization of the ingested plant materials by the various rumen microbes precedes the enzymatic degradation and assimilation of the carbohydrates and different strategies can be observed (Chesson & Forsberg, 1988).

The rumen microflora and fauna include a large variety of permanent or transiently represented anaerobic or facultatively anaerobic bacteria, fungi and protozoa. Establishment of the rumen microorganisms takes place when the mother grooms the young animal and in the process transfers microbes present in its mouth to the offspring. Salivation on the feed or pasture provides another mechanism to establish the microorganisms in the gut of the young animals. Diurnal and seasonal fluctuations in rumen microbes, depending on the feeding cycle and seasonal factors due to diet, are observed (Dehority & Orpin, 1988).

1.13 Cellulolytic rumen microorganisms

1.13.1 Rumen bacteria

The rumen bacteria have been reviewed extensively by Stewart & Bryant (1988), therefore only a few selected species will be mentioned here. Recently the growth rates of some of the principal rumen bacteria were compared and it was found that strains most abundant in the rumen grew more rapidly than the less abundant strains (Van Gylswyk *et al.*, 1992). *Prevotella ruminicola*-like strains (formerly *Bacteroides ruminicola*) and strains of *Butyrivibrio fibrisolvens* had the fastest growth rates, other bacteria were strains of *Ruminococcus albus*, *R. flavefaciens*, *Selenomonas ruminantium*, *Megasphaera elsdenii*, *Eubacterium cellulosolvens* and a butyrate-forming coccus. It was concluded that, besides energy and nitrogen sources, other growth factors (e.g. vitamins present in yeast extract used to supplement the axenic cultures) played an important role in the relative predominance of different bacterial species in the rumen.

Substrate preferences and different strategies of substrate utilization involving various catabolite regulatory mechanisms enable rumen bacteria to survive in separate or overlapping niches in the rumen (Russel & Baldwin, 1978). The utilization of 2-

deoxyribose, a precusor in nucleic acid synthesis, by a relatively low proportion of rumen bacteria, especially *S. ruminantium*, illustrates the point of selective niches (Rasmussen, 1993). *R. flavefaciens* utilizes cellobiose but not glucose as growth substrate (Helaszek & White, 1991), while *R. albus* preferrentially utilizes cellobiose but was able to grow on glucose (Thurston *et al.*, 1993). Similarly *P. ruminicola* is not able to regulate glucose transport and utilization (Russel, 1992). The secretion of cellulolytic, hemicellulolytic and amylolytic enzymes into the rumen fluid leads to the accumulation of breakdown products in the culture fluid which are available to other species. This leads to synergistic growth effects in the degradation of substrates during co-cultivation experiments with combinations of pure cultures (Osborne & Dehority, 1989; Cotta, 1992).

Butyrivibrio fibrisolvens solubilizes hemicellulose in plant cell walls more extensively than cellulose (Stewart & Bryant, 1988). Several genes encoding xylanolytic and cellulolytic enzymes have been cloned recently, including genes encoding a 73–kDa xylanase, a cellodextrinase (*ced1*), an endoglucanase (*end1*) and a ß-glucosidase from *B. fibrisolvens* H17c (Berger *et al.*, 1989, 1990; Lin *et al.*, 1990; Lin & Thomson, 1991a), and the endoglucanases *celA* gene from *B. fibrisolvens* A46 (Hazlewood *et al.*, 1990b). A cytoplasmic arabinofuranosidase was purified from *B. fibrisolvens* GS113 and characterized (Hespell & O'Bryan, 1992). The induction of a number of xylanolytic enzymes by xylan, xylo-oligosaccharides and xylobiose was observed (Williams & Withers, 1992). Two large multi-enzyme complexes similar to the *C. thermocellum* cellulosomes were found in the extracelluar fractions of a *B. fibrisolvens* H17c culture. They contained xylanases and endoglucanases and were called "xylanosomes" (Lin & Thomson, 1991b). The extracellular polysaccharide production by *B. fibrisolvens* was investigated and the chemical composition and structure of the polysaccharide was determined (Wachenheim & Patterson, 1992; Anderson *et al.*, 1993).

Fibrobacter succinogenes (formerly *Bacteroides succinogenes*) is one of the most predominant bacterial species isolated from the rumen (Stewart & Bryant, 1988). One reason for this is its ability to degrade crystalline cellulose readily. The cellulase and hemicellulase activity is found mainly extracellularly. Its cellulases have been investigated extensively and various enzymes were purified and characterized e.g. a cellodextrinase (Huang & Forsberg, 1987) and endoxylanases 1 and 2 (Matte & Forsberg, 1992). It also possesses an a-glucuronidase (Smith & Forsberg, 1991) and at least 3 endoglucanases (McGavin *et al.*, 1990). Four xylanase genes have been cloned in *E. coli* (Malburg *et al.*, 1993). A pH-

sensitive sodium symport mechanism has been implicated for the transport of glucose into the cells, which requires an electrical or a sodium gradient (Chow & Russel, 1992).

Prevotella ruminicola (formerly *Bacteroides ruminicola*) has a high growth rate as observed in axenic culture (Van Gylswyk *et al.*, 1992) and it is presumed to play an important role in proteolysis and the uptake and fermentation of peptides in the rumen (Stewart & Bryant, 1988). Pentoses, which are breakdown products of hemicellulose, are metabolized preferentially over cellobiose and sucrose and *P. ruminicola* seems to possess some pentose/sodium symport mechanisms for the transport of xylose and arabinose (Strobel, 1993b).

Ruminococcus flavefaciens and *R. albus* are considered to be some of the most cellulolytic rumen bacteria (Stewart & Bryant, 1988). Adherence of the cells to the cellulose substrate and the presence of protuberances on the cell surface have been correlated with the ability to degrade crystalline cellulose (Morris & Cole, 1987; Stewart *et al.*, 1990). The nutrient 3–phenylpropanoic acid seems to improve the affinity of *R. albus* for cellulose and thereby enhances its degradation, but the causes of this effect are unknown (Stewart & Bryant, 1988; Morrison *et al.*, 1990). *R. flavefaciens* cellulases have been investigated extensively and several enzymes, including two endoglucanase complexes containing a number of separate active polypeptides, an exo-ß-1,4-glucanase and an endo-ß-1,3-glucanase were purified from the extracellular culture fluid and analyzed (Gardner *et al.*, 1987; Doerner & White, 1990; Erfle & Teather, 1991). Several genes encoding cellulases from *R. flavefaciens* were cloned in *E. coli* and analyzed, including a cellodextrinase gene (*celA*), two xylanase genes (*xynA* and *xynB*) and a bifunctional xylanase gene (*xynD*), which encodes two different xylanase catalytic domains (Barros & Thomson, 1987; Flint *et al.*, 1991, 1993; Brown *et al.*, 1993). The transcriptional regulation of five cloned genes has been demonstrated by northern analysis to depend on the growth substrate (Doerner *et al.*, 1992; Wang *et al.*, 1993).

Cellulolytic enzymes from *Ruminococcus albus*, including high-molecular weight cell bound cellulase, a cellobiosidase and a ß-glucosidase, have been described (Stewart & Bryant, 1988). Gene cloning resulted in the analysis of a ß-glucosidase gene (Honda *et al.*, 1988) and several endoglucanase genes (Howard & White, 1988; Ware *et al.*1989; Ohmiya *et al.*, 1989, 1991; Poole *et al.*, 1990; Deguchi *et al.*, 1991).

Selenomonas ruminantium occurs in large numbers in the rumen of animals fed on cereal grains (Stewart & Bryant, 1988). It was shown to phosphorylate glucose by utilizing a phosphoenolpyruvate (PEP)-dependent system and competition experiments indicated that separate phosphotransferase (PTS) enzymes II were present for glucose and sucrose (Martin & Russel, 1986, 1988). The utilization and transport of pentose sugars also seemed to be regulated by the PTS system (Strobel, 1993). *S. ruminantium* is one of the few rumen bacteria able to utilize 2-deoxyribose, a product of nucleic acid degradation, as growth substrate (Rasmussen, 1993).

Streptococcus bovis is one of the major amylolytic rumen bacteria and grows rapidly in the rumen of animals fed on an excess of starch i.e. a rich cereal diet (Stewart & Bryant, 1988; Cotta, 1992). The rapid fermentation of the starch and production of lactate can lead to the development of lactic acidosis. As a result of the lowered pH the fibre digestion by other rumen microbes is decreased. *S. bovis* phosphorylates cellobiose, glucose, mannose and sucrose with a PEP-dependent PTS system (Martin & Russel, 1986, 1987).

Methanogenic bacteria are not cellulolytic, but they influence fibre digestion in the rumen by utilizing the H_2 produced during the rumen fermentation to reduce CO_2 to CH_4 and H_2O. The fermentative conversion of fatty acids to CH_4 does not occur in the rumen, because it does not yield as much energy as the conversion of hexoses to fatty acids and therefore takes place very slowly, which would lead to the washout of the microbial species in the rumen. The maintenance of a low partial pressure of H_2 by methanogenesis has a strong effect on the production of H_2 and other products (fatty acids) by the fermentative rumen bacteria. The interspecies H_2 transfer plays an important role in regulating the relative amounts of acetate and propionate. (Wolin & Miller, 1988). The production of CH_4 represents a loss of energy to the ruminant because it is linked with decreased production of propionate. Inhibition of methanogenesis by certain chemicals can be used to manipulate rumen fermentation (Van Nevel & Demeyer, 1988). Monensin is an ionophore antibiotic which decreases methane production and leads to increased propionate production.

1.13.2 Rumen protozoa

The importance of rumen protozoa to the ruminants is unclear because protozoa-free or defaunated ruminants can be established i.e. they are not essential to the host (Dehority & Orpin, 1988; Williams & Coleman, 1988). The protozoans obtain their energy from the fermentation of plant structural and storage polysaccharides and it seems that they have hemicellulolytic and pectinolytic activities. They prey on bacteria to obtain nitrogenous precursors for protein and nucleic acid biosynthesis and lipid precursors and this can reduce bacterial populations significantly (Wolin & Miller, 1988). Protozoal predation was found to be the main reason for the rapid loss of recombinant and unmodified strains of *Lactobacillus plantarum* introduced into the rumen (Sharp *et al.*, 1994). The aim of the study was to monitor the persistence of genetically modified silage lactic acid bacteria in the rumen. Protozoan concentrations in domestic ruminants fed on roughage diets range from 10 to $50x10^4$ per ml, and from 50 to $150x10^4$ per ml for animals fed concentrate type rations (Dehority & Orpin, 1988).

1.13.3 Rumen fungi

Neocallimastix frontalis, N. patriciarum, Piromonas communis and *Sphaeromonas communis* are anaerobic saprophytic fungi found in the rumen (Orpin & Joblin, 1988). Zoospores are part of their life cycle and it was observed that the concentration of ciliated zoospores increased rapidly after feeding. This was due to the release of the zoospores from a reproductive body (sporangium) on the vegetative phase of the fungi which was found to be stimulated by haeme-containing compounds in the food. The flagellates lose their motility after about one hour and develop into the vegetative phase on the food particles (Dehority & Orpin, 1988). They secrete endo- and exo-cellulases into the culture medium to degrade plant material. The fungal mycelia seem to be able to penetrate and break up plant fibres and thus gain better access to the substrate than the bacteria. It was speculated that the proteolytic activities of these fungi play an important role in the penetration of the plant material. Initial attack of the substrates takes place at broken edges and stomata. Even though they exhibit strong cellulolytic activities the net effects of the fungi in the rumen are difficult to determine. It was found that in sheep treated with biocides against fungi the rumen bacteria were more effective at degrading resistant plant fibre (Orpin & Joblin, 1988).

Wood *et al.* (1988) described some of the cellulolytic activities of *N. frontalis*. It produces a multi-enzyme complex with a molecular weight of between 1 and 2 MDa. Sequence analysis of the genes *celB* and *xynA*, encoding endoglucanase CelB and xylanase XYLA, from *N. patriciarum* revealed repeated sequences of 30 residues, which are speculated to serve the same function as the *C. thermocellum* repeats in securing the enzymes to the scaffolding protein (Gilbert *et al.*, 1992; Zhou *et al.*, 1994). Other cellulase genes obtained from a cDNA expression library of *N. patriciarum* are *celA*, *celC* and *celD*, encoding cellobiohydrolase A, endoglucanase B and a multidomain protein containing an endoglucanase, cellobiohydrolase and a xylanase domain, respectively (Xue *et al.*, 1992a, 1992b). XYLA is also a multidomain enzyme with two homologous xylanase domains. Northern analysis showed transcriptional induction of *celA*, *celB* and *celC* during growth on cellulose, but *celD* was transcribed at similar levels during growth on glucose and cellulose. Catabolite regulated expression of cellulolytic activity was also observed in a *Piromyces*-like ruminal fungus (Morrison *et al.*, 1990b). Inhibition of *N. frontalis* cellulolytic activity, but not of fungal growth, by non-cellulolytic proteins secreted by *R. albus* and *R. flavefaciens* has been observed, but the importance of this in the rumen environment is not known (Stewart *et al.*, 1992; Bernalier *et al.*, 1993).

1.14 Genetic transfer in rumen bacteria

There is considerable interest in the genetic modification of cellulolytic enzyme systems in the rumen (Orpin, 1988; Wallace, 1992), although whether this is feasible is debatable. Suitable genetic transfer systems have to be developed for this purpose, but also as tools in the analysis of gene function in rumen bacteria by classical mutation and complementation studies. A number of plasmids have been isolated by several groups out of various strains of *B. fibrisolvens*, *S. ruminantium* and *Ruminococcus*, and the construction of shuttle vectors has been attempted (Hazlewood & Teather, 1988). Two small cryptic plasmids, pJDB21 and pRJF1, from *S. ruminantium* and *Butyrivibrio*, respectively, have been sequenced (Hefford *et al.*, 1993; Zhang & Brooker, 1993). Shuttle vectors based on a *P. ruminicola* 9.6–kb cryptic plasmid and an *E. coli* vector were constructed and transferred into various *Bacteroides* strains and into *P. ruminicola* by conjugation or by electroporation (Béchet *et al.*, 1993). Electroporation was used to transform *P. ruminicola*, *Bacteroides uniformis*, *R. albus*, *B. fibrisolvens* and *S. bovis* (Thomson & Flint, 1989; Cocconcelli *et al.*, 1992; Whitehead,

1992). Flint *et al.* (1988) demonstrated the plasmid-associated transfer of tetracycline resistance between two strains of *P. ruminicola.* The transfer of a naturally occuring conjugative plasmid from *P. ruminicola* between bacteria from the human colon and ruminal bacteria was shown (Shoemaker *et al.*, 1992). A *P. ruminicola* xylanase gene on plasmid pVAL-RX was highly expressed in colonic *Bacteroides* species and was stably integrated into the *Bacteroides thetaiotaomicron* chromosome by using the suicide vector pVAL-7 (Whitehead *et al.*, 1991). Tn916 and pAMß1 were conjugatively transferred between *Enterococcus faecalis, B. fibrisolvens, S. bovis* and *R. albus* (Hespell & Whitehead, 1991a, 1991b, Aminov *et al.*, 1993). Similarly the shuttle vector pRDB5 was transferred from the colonic *Bacteroides uniformis* to *P. ruminicola* (Shoemaker *et al.*, 1991). The identification and partial purification of restriction endonucleases from *R. flavefaciens* and *R. albus* and the unravelling of the methylation protective mechanism in the former host are important because they point out certain considerations and restrictions in the development of genetic transfer systems in these bacteria (Morrison *et al.*, 1992a, 1992b).

1.15 *Clostridium longisporum*

We are interested in the characterization and possible modification of cellulose degradation in the rumen. A large obstacle in the modification of the rumen microflora is the extreme oxygen-sensitivity of these bacteria. Aero-tolerant endospores as possible "vectors" for introducing genetically modified cellulase genes into the rumen, especially for use in commercial animal feedstocks, prompted our investigations of *C. longisporum* B6405, which had recently been re-isolated by Varel (1989). He reported the formation of subterminal spores by this bacterium when grown on insoluble carbon sources. *C. longisporum* had originally been isolated and characterized by Hungate (1957) but was subsequently lost from pure culture because of difficult culturing requirements. Similar clostridia had been isolated from water buffaloes (Sinha & Ranganathan, 1983). Isolate B6405, which was obtained from the rumen of a bison, was very similar to the original isolates and was therefore termed *C. longisporum.* It was found to degrade hemicellulose from alfalfa cell walls more extensively than various other rumen bacteria, even though it was not able to grow on larchwood xylan, xylose or arabinose. In a subsequent comparison of the cellulolytic activities of *C. longisporum* and *R. albus* SY3 the latter degraded alfalfa and barley straw more extensively (Varel *et al.*, 1989). The characterization of cellulolytic enzymes from this bacterium was expected to explain some

of the peculiarities of these clostridia, i.e. their sporadic occurrence and their relatively low numbers in the rumen. The work presented in this thesis represents the first cloning and detailed analysis of a cellulase gene from this bacterium. The regulation and expression of this gene in *C. longisporum* was investigated. The total endogenous endoglucanase activity was surveyed briefly in preparation for more detailed analysis of other cellulase genes. The knowledge gained from this will be used to establish whether it will be possible to genetically modify some of its cellulolytic activities for the potential re-introduction into the rumen.

CHAPTER II

CLONING AND SEQUENCING OF THE *celA* GENE OF *C. longisporum*

TABLE OF CONTENTS

2.1 SUMMARY

A genomic library of *Clostridium longisporum* B6405 in the host *Escherichia coli* was screened for endo-ß-glucanases and strains carrying plasmids pCM64 and pCM4 were isolated. The nucleotide sequence of a 3620-bp fragment was found to carry a 1548-bp open reading frame (ORF), termed *celA*, which encodes an endo-1,4-ß-glucanase, CelA, assigned to family 5 (Henrissat & Bairoch, 1993). N-terminal amino acid sequence determination revealed that pCM64 encoded the full-length *celA* gene including a signal sequence, while pCM4 carried a 5'-truncated *celA* gene expressed as an N-terminal fusion protein, CelAΔN', without a signal sequence. This fusion protein was initiated from an ATG codon in the vector. The sequence data revealed that CelA has a catalytic domain (CD) and a cellulose-binding domain (CBD), which are separated by a putative linker domain. The CBD shows homology with other bacterial CBDs from *Cellulomonas fimi* belonging to the type I family of CBDs (Béguin & Aubert, 1994).

The transcriptional start site of *celA* was mapped using primer extension and was found to be the same in *C. longisporum* and in *E. coli* expressing the cloned *celA* gene. A consensus *E. coli* –10 promoter region could be identified (AATAAT), but not a –35 promoter region. Two direct repeats (TATTGAATTTAT) separated by 15 nucleotides flank the region where the consensus –35 promoter regions would have been. The size of the *celA* mRNA transcript corresponded with the size of the ORF. A potential stemloop structure was found 18 nucleotides downstream of the 3' stop codon, which could be responsible for termination of transcription. Another less efficient terminator was found at the 5' end of the cellulose binding domain, which could be responsible for truncated *celA* mRNA transcripts encoding only the catalytic domain of the enzyme.

2.2 INTRODUCTION

The complexity of the substrate lignocellulose is responsible for the relatively large number of cellulases and hemicellulases required for its degradation by microorganisms. Purification of individual cellulases out of the mixture of enzymes found in culture fluids or in cell free extracts and the characterization of such purified fractions has been problematic and could not explain the multiplicity of enzymes fully. Cloning of cellulase genes in heterologous non-cellulolytic hosts has overcome most of these problems. Sequence analysis of cloned cellulase genes has provided a wealth of information regarding the domain structure of these enzymes and has enabled the identification and classification of enzymes with similar structures and functions. Routine procedures for the identification and characterization of cellulases are the screening of genomic libraries, the subsequent sequence analysis of clones expressing specific enzymatic activities (or any other interesting characteristics) and characterization of the purified recombinant enzymes. Unfortunately there can be no certainty whether all the different cellulase genes expressed by a microorganism have been cloned as some genes have been very elusive e.g. "true" exo-glucanases. The possibility that some of these genes cause lethality in the heterologous hosts can not be disproved.

The characterization of cellulases from *C. longisporum* was the aim of this project. The *celA* gene was the first to be cloned and analyzed and it was found to encode an endo-ß-glucanase, CelA, with a modular structure containing a catalytic domain and a cellulose-binding domain. This work has already been published by Mittendorf & Thomson (1993). The transcription initiation site for the expression of *celA* was found to be the same in *C. longisporum* and in *E. coli*.

2.3 MATERIALS AND METHODS

Methods and techniques not described or referenced here were performed as described by Sambrook *et al.* (1989).

2.3.1 Microorganisms, plasmids and culture conditions

C. longisporum B6405 ATCC 49440 was obtained from D.H. Varel (1989) and cultured in an anaerobic cabinet (5 % H_2, 10 % CO_2 and 85 % N_2 gas phase) in non-rumen fluid medium (NRF) (Caldwell & Bryant, 1966) containing 2 % (w/v) cellobiose but without glucose or starch. The NRF medium was buffered with 0.1 % (w/v) $NaHCO_3$ instead of 0.2 % (w/v) Na_2CO_3. Cultures were stored in Hungate tubes at -70°C on NRF medium slants containing 1.2 % (w/v) agar. *E. coli* LK111 (Zabeau & Stanley, 1982), *E. coli* JM105 (Yanisch-Perron *et al.*, 1985), *E. coli* K514 and *E. coli* K514λ (Wood, 1966; Zabeau & Stanley, 1982) were used for cloning purposes. DNA fragments were subcloned into pEcoR251 (Zappe *et al.*, 1986), pUC18, pUC19 (Yanisch-Perron *et al.*, 1985) or the Bluescript vector, pSK(M13+) (Stratagene). *E. coli* BL21 (F-*hsdS gal*) (Studier & Moffat, 1986) was used to prepare RNA. *E. coli* XL1-Blue and the helper phage VCS-M13 were obtained from Stratagene and used for the production of ssDNA for the primer extension experiments.

2.3.2 *C. longisporum* DNA preparation

This organism produces copious quantities of an orange-pigmented exo-polysaccharide that entraps the clostridial cells. For the extraction of DNA the cells were grown in 1.5 l NRF medium for 15 to 20 h and the culture harvested early in the life cycle when the cells were still motile and little of the slime (exo-polysaccharide) had been produced. The cell pellet was resuspended in 10 ml solution A (10 mM Tris.Cl pH 8.0, 25 % (w/v) sucrose) containing 5 mg/ml lysozyme and incubated for 30 min at 37°C. EDTA and SDS were added to final concentrations of 0.1 M and 2 % (w/v) respectively, and the progress of cell lysis, which took 10 to 30 min, monitored microscopically. The cell lysate was extracted once with one volume phenol : chloroform : isoamyl-alcohol (25:24:1), twice with one volume chloroform : isoamyl-alcohol followed by H_2O-saturated ether. The solution was

treated with RNAse and dialyzed for 24 h against TE (10 mM Tris.HCl pH 8.0, 1 mM EDTA).

2.3.3 Construction and screening of the genomic library

The *C. longisporum* chromosomal DNA was partially digested with *Sau*IIIA restriction endonuclease and fragments between 5 and 10 kb isolated from a sucrose gradient. These were cloned into the *Bgl*II site of the positive selection vector pEcoR251, transformed into *E. coli* K514 and screened on Luria-Bertani medium (LB) (Sambrook *et al.*, 1989) containing 1 % (w/v) carboxymethyl-cellulose (CMC). The plates were stained with Congo Red (Teather & Wood, 1982) to identify CMC degrading clones.

2.3.4 DNA manipulations and sequencing

CMCase-positive clones pCM4 and pCM64 were isolated, mapped with restriction endonucleases and subcloned according to standard procedures (Sambrook *et al.*, 1989). The subcloning and sequencing strategy is explained in the legends to Figs. 2.1 and 2.2 and the plasmid maps of pCM64, pCM4, pCMI and pCML are given in the Appendix. The complete sequence of 3620 bp containing the *celA* gene and flanking regions was assembled from the sequences obtained from pCMC, pCMJI and pCMJII, the former carrying the 5'-truncated *celA* gene plus the downstream region and the latter two carrying the 5'-part of the *celA* gene and the 5'-upstream region. The DNA was sequenced in both directions using clones carrying overlapping deletions obtained by *Exo*III exonuclease degradation (Sambrook *et al.*, 1989). The GCG software was used to analyze the sequence data (Genetics Computer Group, 1991).

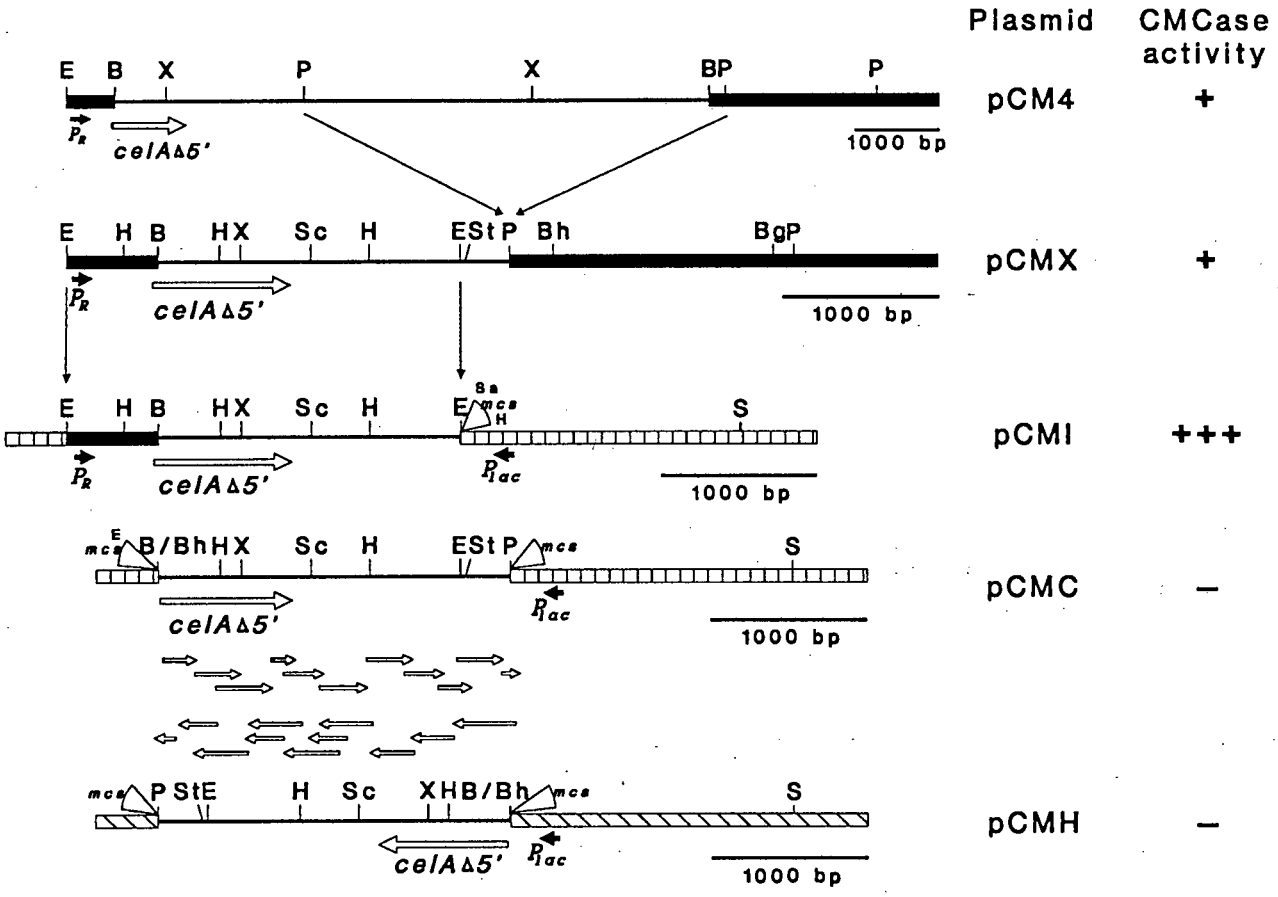


Fig. 2.1. Restriction maps of *celA*Δ5'-carrying plasmids, subcloning and sequencing strategy. pCMX is a 5.1–kb *Pst*I deletion derivative of pCM4. pCMI was obtained by cloning the 2.53-kb *Eco*RI fragment of pCMX, which includes 590 bp of pEcoR251 vector DNA, into pUC19. pCMC and pCMH contain the same 2.26-kb *Bgl*II-*Pst*I insert as pCMX, but in pUC19 and pUC18, respectively. A number of *Exo*III-shortened derivatives of pCMC were sequenced in both orientations as indicated by the arrows below the map of pCMC. The position and directions of the λP_R promoter in pEcoR251 and the P_{lac} promoter in pUC18 and pUC19 are indicated. Restriction sites of selected endonucleases are indicated as follows: B, *Bgl*II; Bg, *Bgl*I; Bh, *Bam*HI; E, *Eco*RI; H, *Hind*III; P, *Pst*I; Sc, *Sca*I; St, *Stu*I; X, *Xba*I. B/Bh is the *Bgl*II-*Bam*HI cloning junction. mcs, multiple cloning site containing other unique restriction sites as indicated by the wedge-shaped segment. CMCase activity is indicated on the right-hand side of the figure. See the Appendix for more detailed plasmid maps of pCM4 and pCMI.

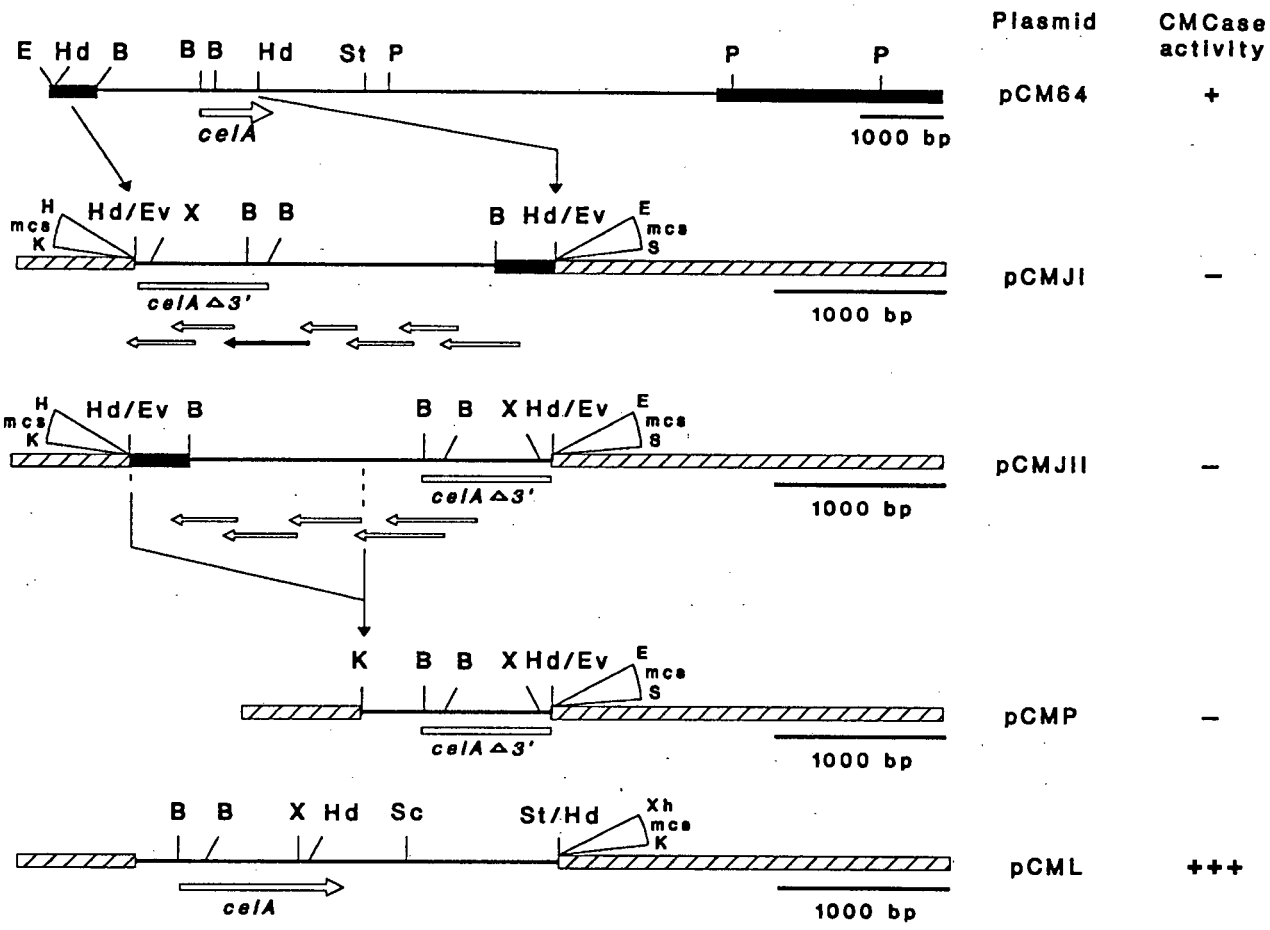


Fig. 2.2. Restriction maps of the full-length *celA* gene on various plasmids and sequencing strategy. pCMJI and pCMJII contain the 2.65 kb *Hind*II fragment of pCM64, cloned in both orientations into the *Eco*RV site of the vector pSK. During these subclonings, 720 bp of the 3' end of *celA* were deleted as indicated by *celA*Δ3' in the diagram. Selected *Exo*III deletion derivatives were sequenced as indicated by the arrows below the maps of pCMJI and pCMJII. pCMP was obtained by *Exo*III shortening of pCMJII. pCML was constructed for the purification of CelA encoded by *celA*. The 3' end of the *celA* gene was spliced onto the 5' end of the *celA*Δ3' gene by cloning the 1.4-kb *Xba*I-*Stu*I fragment of pCMC (see Fig. 2.1) into pCMJI-23 (the *Exo*III-shortened derivative of pCMJI shown with the solid arrow) cut with *Xba*I and *Hind*III (the latter site being in the multiple cloning site, mcs). Restriction sites of selected endonucleases are indicated as follows: B, *Bgl*II; E, *Eco*RI; Ev, *Eco*RV; H, *Hind*III; Hd, *Hind*II; K, *Kpn*I; P, *Pst*I; S, *Sac*I; Sc, *Sca*I; St, *Stu*I; X, *Xba*I; Xh, *Xho*I. Hd/Ev and St/Hd are the *Hind*II-*Eco*RV and the *Stu*I/*Hind*II cloning junctions, respectively. CMCase activity is indicated on the right-hand side of the figure. mcs, multiple cloning site containing other unique restriction sites as indicated by the wedge-shaped segment. See the Appendix for more detailed plasmid maps of pCM64 and pCML.

2.3.5 Southern blot analysis of *C. longisporum* DNA

Chromosomal DNA from *C. longisporum* and "mini-prep" plasmid DNA was cut with restriction enzymes and run on TAE agarose gels according to standard procedures (Sambrook *et al.*, 1989). The gels were capillary blotted with 0.4 M NaOH onto Hybond N+ nylon membranes (Amersham) and probed with non-radioactive DIG-labelled DNA probes using the Boehringer Mannheim DNA labelling and detection kit.

2.3.6 Primer extension to map the transcriptional start site

2.3.6.1 Brief overview of the method

The primer extension protocol was adapted from that of Hu & Davidson (1986), which involves the hybridization of RNA to a ssDNA construct containing the antisense message of the gene of interest. This is followed by the hybridization of a γ-^{32}P-endlabelled primer to the RNA:ssDNA complex 100 to 200 bases upstream of the start codon of the gene of interest. T4 DNA polymerase is used to extend the nascent DNA strand from the primer towards the gene. Due to the lack of 5'→3' exonuclease activity the T4 DNA polymerase drops off the ssDNA template when it encounters a RNA molecule hybridized to the ssDNA template. Thus labelled extension products are formed which terminate at the transcription start site and the exact base at which the extension is terminated can be determined from a sequencing ladder obtained with the same primer.

2.3.6.2 Construction of pCMP and preparation of ssDNA

pCMP was constructed by exonuclease (*Exo*III) deletion shortening of pCMJII (Fig. 2.2). pCMP contained the 3' truncated *celA* gene (only the first 821 bp of the total 1548-bp open reading frame were present; see Fig. 2.5 for the extent of the 3' truncation) together with 231 bp of 5' upstream sequence in the bluescript vector pSK(M13+). pCMP was transformed into *E. coli* XL1-Blue and ssDNA was obtained after infection with the helper phage VCS-M13 using a standard protocol from the supplier (Stratagene). Care was taken during the construction of pCMP to ensure that it would yield ssDNA complementary to the 5' region of the *celA* mRNA.

2.3.6.3 Preparation of total RNA from *C. longisporum* and *E. coli*

Total RNA from *C. longisporum* and *E. coli* cultures was prepared using a slightly adapted protocol from Aiba *et al.* (1981). The cell pellets were resuspended in 0.3 M sucrose, 0.01 M sodium acetate (pH 4.5) and lysed with 2 % (w/v) SDS, 0.01 M sodium acetate (pH 4.5). The cell lysate was extracted three times with hot TE-buffered phenol (65°C) followed by an ethanol precipitation. The pellets were resuspended in Milli-Q H_2O (Millipore H_2O purification system) and treated with DNAse I, which was followed by another phenol extraction and ethanol precipitation. RNA was stored in Milli-Q H_2O. To prevent RNA degradation glassware was heat treated, plastic ware was new or treated with hypochlorite and only RNAse-free chemicals were used.

2.3.6.4 Primer extension protocol

pCMP ssDNA (250 to 600 ng) was hybridized to total RNA from *C. longisporum* or *E. coli* (5 to 15 µg). The volume of the hybridization reaction was increased with RNA-sterile H_2O prior to the ethanol precipitation to prevent the precipitation of salts. γ-^{32}P-endlabelled T7 sequencing primer (5' AATACGACTCACTATAG 3') was annealed to the RNA:ssDNA pellets in T4 DNA polymerase buffer. The synthesis reaction was carried out by adding T4 DNA polymerase and dNTPs. The reaction was stopped after 10 min by the addition of sequencing stop buffer (Sequenase kit) and the samples were denatured at 85-90°C for 3 min. Aliquots of the reactions were loaded on a sequencing gel together with sequencing ladders obtained from the same phagemid construct (pCMP) and the same primer using the Sequenase sequencing kit.

2.4 RESULTS

2.4.1 Growth characteristics of *C. longisporum*

Since this strictly anaerobic rumen bacterium is not cultured easily at all times, a short description of its growth characteristics will be given.

2.4.1.1 Long-term storage at -70°C and growth on solid medium

A fresh overnight *C. longisporum* culture was used to inoculate NRF slants in Hungate tubes. These tubes were incubated at 37°C overnight and, if a faint orange-pigmented growth was visible on them, stored at -70°C. A viable culture was obtained from a slant that had been stored for three years. Upon retrieval from -70°C storage the slants were thawed and one drop of liquid was used to inoculate NRF medium. Growth was usually visible after 16 h. The cells were streaked out on solid NRF medium and after two days dark-orange pigmented colonies with spreading tendencies were observed. Cells remained viable on solid medium for up to two weeks and single colonies were used to inoculate starter cultures.

2.4.1.2 Growth curves, cell motility and pH of the medium

Small batch cultures of *C. longisporum* grown in NRF medium containing cellobiose show rapid growth. The pH of the medium affected cell growth considerably. Maximum cell density obtained after 24 h was 3.5×10^7 cfu/ml and the pH at that sampling time was about pH 5.5. The pH of the medium was initially 7.1, but during the course of the fermentation it dropped steadily. Motile cells were observed during log phase until the pH was about 6.75 and the cell density was 5.3×10^6 cfu/ml. The generation times also reflected the state of the cells, which depended on the pH of the medium. The initial generation time was 1.1 h/generation, later on it was 3 h/generation. The mean generation time from 6 h to 24 h was 1.5 h/generation. After 24 h cells looked thin and sickly and the first signs of cell lysis were observed. The growth rate on cellobiose was relatively high and different growth rates can be expected on other carbon sources. The digestibility of the carbohydrates has a large influence on the generation time of the

bacteria, and as such is an important factor in the maintenance of the bacteria in the continuous flow system of the rumen where washout can occur (Hungate, 1988). During the late growth stage a lot of extracellular polysaccharides (EP) were produced and the orange color of the culture intensified. For this reason only "young" cultures were used for the extraction of DNA and for other purposes.

2.4.1.3 Utilization of other carbon sources

C. longisporum was able to grow on the following carbon sources: cellobiose, ß-glucan, glucose, salicin, sucrose. Some growth was observed with fructose, but none with cellulose (ball-milled filter paper or Avicel), pectin and xylan. Varel (1989) investigated the utilization of various carbon sources by *C. longisporum*, but some minor discrepancies were found, e.g. pectin and cellulose were reported to be utilized. The inability to grown on xylose or xylan is interesting because several xylanases are produced by *C. longisporum* (see sections 3.4.5 and 4.4.4). This has also been observed with other bacteria, e.g. *C. thermocellum*, which produce significant xylanase activity, but do not utilize the xylan or its degradation products for growth (Morag *et al.*, 1990).

2.4.2 Construction and screening of the genomic library

Approximately 12 000 clones with an average insert size of 5 kb were obtained for the genomic library. Screening for CMCase activity produced 40 clones which were placed into 6 groups according to similarities of endonuclease restriction patterns. pCM4 (Fig. 2.1) exhibited the strongest CMCase phenotype and was chosen for further studies. Southern blot analysis revealed that two other clones, pCM4b and pCM4c, were identical to pCM4, and that pCM64 (Fig. 2.3 A) encoded the same cellulase gene on a larger insert, but the other clones were different (refer to section 4.4.4 - at least 12 of these clones were found to encode the same gene termed *celB*). The *Pst*I digest of pCM64 (Fig. 2.3 A) shows seven restriction fragments, which is in contrast to the restriction map of pCM64 (Appendix) which indicates only three *Pst*I restriction sites. The reason for this is that the sample of pCM64 mini-prep DNA used for the restriction digest contained two different plasmids which were separated during later manipulations. Southern blot analysis also confirmed that the insert DNA of pCM4 originated from *C. longisporum* (Fig. 2.3 B). *C. longisporum* chromosomal DNA was digested with three different restriction enzymes and the single bands that were obtained on the Southern blot in each case indicate, that the *celA* gene is

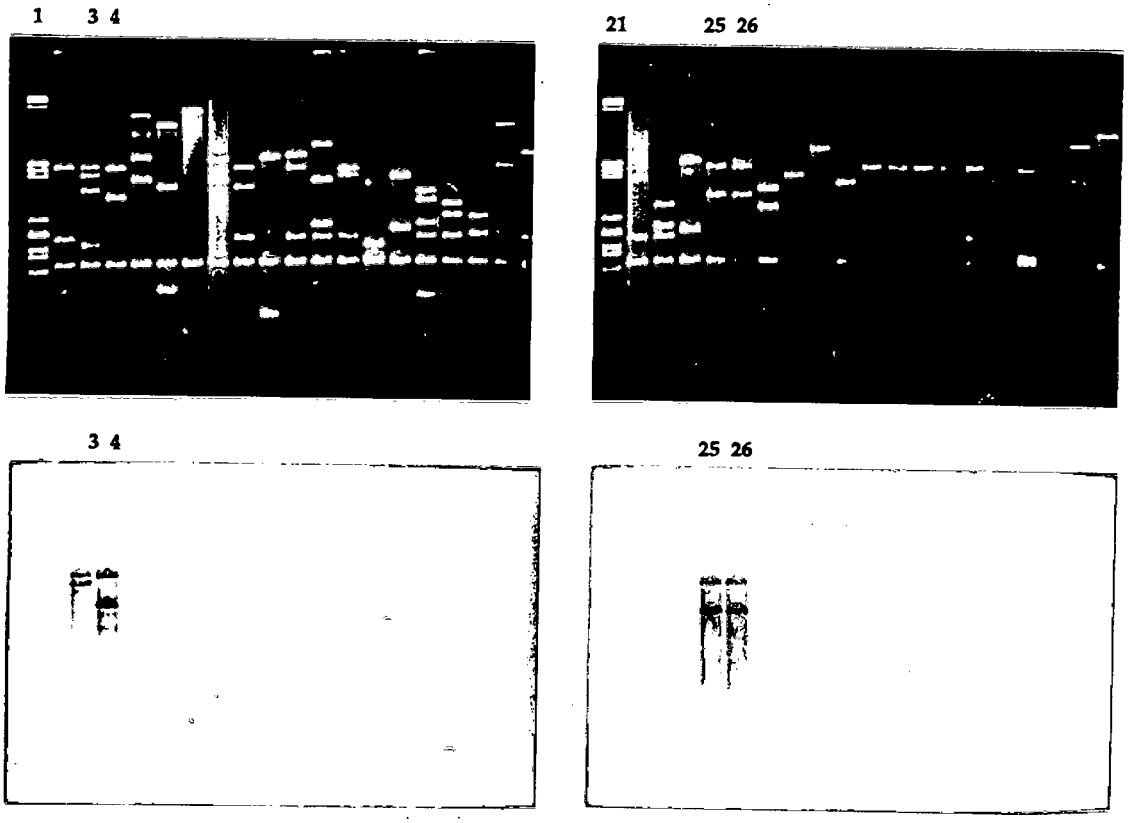


Fig. 2.3 (A) Southern blot of 38 CMCase-positive clones isolated from the genomic library. "Mini-prep" DNA of these clones was digested with *Pst*I, run on an agarose gel and transferred with 0.4 M NaOH in a capillary blot onto Hybond–N+. The probe was a DIG-labelled 4440–bp *Xba*I restriction fragment from pCM4. Clones in selected lanes are: 3, pCM64; 4, pCM82 (≡pCM4); 25, pCM4b; 26, pCM4c. Lanes 1 and 21 contain the molecular size markers, λ DNA digested with *Pst*I.

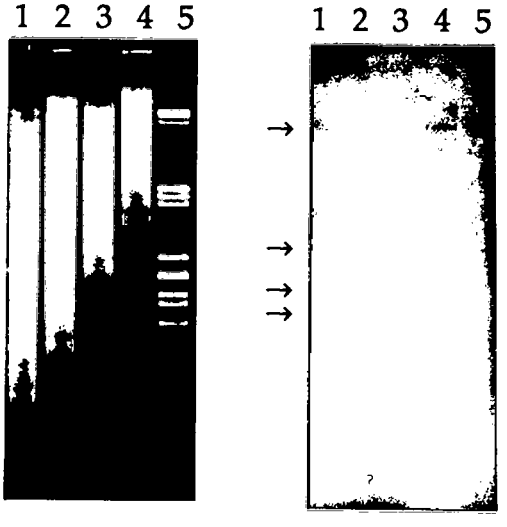


Fig. 2.3 (B) Southern blot of *C. longisporum* genomic DNA. The probe was a DIG-labelled 975–bp *Bgl*II+*Sca*I restriction fragment from pCML. Lanes 1 to 4 each contain 10 µg genomic DNA digested with *Eco*RI+*Bgl*II, *Pst*I+*Bgl*II, *Eco*RI and *Pst*I, respectively. Lane 5 contains the molecular size markers, λ DNA digested with *Pst*I. Arrows indicate faint bands on the Southern blot.

present as a single copy on the bacterial genome. This is in correspondence with the general finding that the multiplicity of cellulases in prokaryotes is not due to extensive gene duplication within the bacteria (Gilbert & Hazlewood, 1993).

2.4.3 Cloning and sequencing of the 5'-truncated *celA'* and the full-length *celA* gene

A number of overlapping fragments of the 7.15 kb *C. longisporum* insert DNA of pCM4 were subcloned into pUC19, but none had CMCase activity (results not shown). Only pCMX, a deletion derivative of pCM4 (Fig. 2.1), was found to express CMCase activity. pCMC and pCMH contained the same *C. longisporum* insert DNA as pCMX, but did not express CMCase activity. Initially it was thought that the external λ promoter P_R was required for the expression of *celA* and could not be replaced by P_{lac} on pCMC or pCMH. However later sequence determination of the first 15 N-terminal amino acid residues of the enzyme purified from pCMI (see section 3.3.4) showed that is was an N-terminal fusion protein, termed CelAΔN'. Expression of the gene, termed *celA*Δ5', was found to be initiated from an in-frame ATG codon in the vector DNA (Fig. 2.4 A). The first seven amino acids of the mature protein were derived from the vector, pEcoR251, and the next eight corresponded to the DNA sequence of the *celA*Δ5' gene on pCMC. This fusion protein lacks a signal sequence and the start methionine was not present in the mature protein. A similar fusion protein was not formed in pCMC and pCMH, due to the lack of a suitable ATG start codon in-frame with *celA*Δ5'.

This finding led us to investigate clone pCM64 (Fig. 2.2) which contained the same gene but had weaker CMCase activity as seen on CMC plates (results not shown). Restriction endonuclease analysis of pCM64 and sequence analysis of pCMJI and pCMJII showed that pCM64 contained the full-length *celA* gene together with 1358 bp of 5'-upstream sequence (Fig. 2.2 and 2.4 B). A contiguous sequence of 3620 bp of *C. longisporum* DNA (GenBank accession number L02868) was obtained by assembling the sequence obtained from pCMC (*celA*Δ5' and the downstream sequence), and from pCMJI and pCMJII (*celA*Δ3' and the upstream sequence); the nucleotide sequence and the derived amino acid sequence are shown in Fig. 2.5. The *celA* ORF is 1548 bp long and stretches from nucleotides 1246 to 2794 in the sequence. Homology comparisons with other cellulases revealed that CelA belongs to family 5 (Henrissat & Bairoch, 1993). This used to be family A4 according to

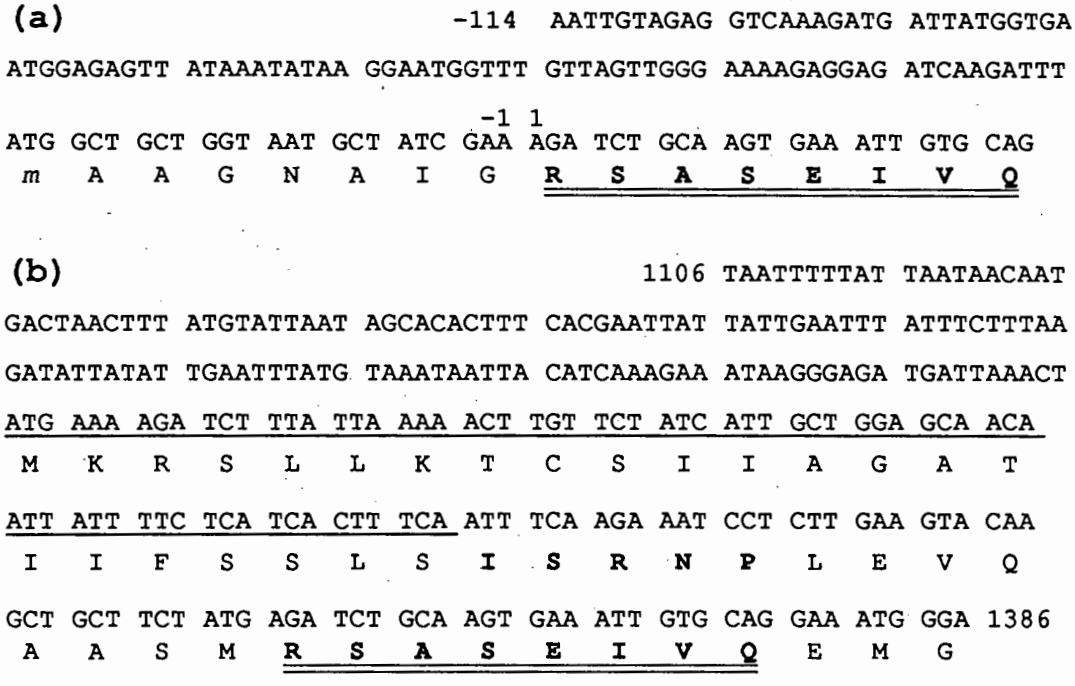


Fig. 2.4. (a) Nucleotide sequence of the 5'-end of *celAΔ5'* and the experimentally determined N-terminal amino acid sequence of CelAΔN'. The methionine expressed by the start codon was not present in the mature protein and is indicated with a small letter *m*. The numbering is based on the beginning of the *C. longisporum* DNA insert in pCM4; the upstream sequence is derived from the vector pEcoR251. **(b)** Nucleotide sequence of the 5'-end of *celA* and the derived amino acid sequence. The sequence of the first five residues of the mature CelA protein (bold letters) was confirmed by N-terminal amino acid sequencing. The underlined residues constitute the signal peptide. The double-underlined amino acid sequence indicates the start of the overlap of *celA* and *celAΔ5'*. Numbering corresponds to GenBank entry LO2868.

```
   1 TACTAACACATATAGAAAGCATTTCAGAATTCTTCTTAAACCACAATCTTCAAATCTTTTCAAAAATACCATTATCATCAGGGTTTACCATGTCTACACT
 101 TCCAGTAATGATTTATATATGTATAATAAATACTGGTATAGGTTATGCATGCTATTTTAAGGCTCTAGAAAAAACTTCAGCAACTAAGACATCTTTAGTA
 201 TTTTTCTTCAAACCAATTTTAGCTCCAATACTAGCTTTAATAATCTTAAAAGAATCTTTATCTTTTAATATGCTTATAGGTATATTATTAATATTGTGTG
 301 GTTCAATGACATCAATACTTCCATCACTTAAAAATAATAAAAATATTAAGCTTAAAAAGGCAGCAATAAAATAAAAGGGTATGTAATTTTGAACAAAATA
 401 GACTGTTCAAAACTACATGCCCTTTAAAATATATAGCTTTCTAGCTTAAGCACCTTTCCAAAATCTGATACAATATCAAAGCTCATCTTTTTTCCATTAA
 501 GCTTTTTTATTAACTGTAAACTTGGTATATTTAAAATATCAGCTGCAAAAATAAGTCCTTCTATCTTATAATTTTCTTTTAAAAAATCTTTTATAAGAAG
 601 TAAAGCTTCATAAGCATATCCCTTTTTTCTTTCATCTTCACATATCCAAATTCCAACTTCCGGACATTCATCCTTCAAATTATAAAGCTCTATACTTCCA
 701 ACAAACTTATCTTCTTCATTCCACAAGCTAAGCATAGTATTATCACCCTTTATACTTAAATTTATAAATTCCTGCAAAAAACTTTCAGCTTCTTTTATAG
 801 ACTTAAAAGAATTTGCAAGCTGGTATTTTACTACCTCTTCAGTAAAGTTTTCATAATATAAAGCTGCCTCTTCAATTTTTGTAGAAGTTATTCTAAGCCT
 901 TTTTCCTTTTATTTCAATTGTATTATTCATAAAAAACCTTCTTAAATATTCTTAATGTATTCTTTAACTGATATAACCTTTTAATCTGAAATTCTTCATA
1001 GCAATACAAATCAACTATACTGTTCATTATAGAATAAATATAAAAATTATAAAAGAAAAATTTATGTTATTTTACATGCTGGTAATGCTAAAACTGGTAG
1101 TATTCTAATTTTTATTAATAACAATGACTAACTTTATGTATTAATAGCACACTTTCACGAATTATTATTGAATTTATTTCTTTAAGATATTATATTGAAT
1201 TTATGTAAATAATTACATCAAAGAAATAAGGGAGATGATTAAACTATGAAAAGATCTTTATTAAAAACTTGTTCTATCATTGCTGGAGCAACAATTATTT
                                                  M  K  R  S  L  L  K  T  C  S  I  I  A  G  A  T  I  I  F
1301 TCTCATCACTTTCAATTTCAAGAAATCCTCTTGAAGTACAAGCTGCTTCTATGAGATCTGCAAGTGAAATTGTGCAGGAAATGGGAGTTGGGTGGAATCT
       S  S  L  S  I  S  R  N  P  L  E  V  Q  A  A  S  M  R  S  A  S  E  I  V  Q  E  M  G  V  G  W  N  L
1401 TGGAAATACTTTAGATGCTAAGATTACTAACCTGTCTTATAATACTTCACCAATATCCTTTGAAACAGGCTGGGGAAATCCTGTTACTACAAAGGCTATG
       G  N  T  L  D  A  K  I  T  N  L  S  Y  N  T  S  P  I  S  F  E  T  G  W  G  N  P  V  T  T  K  A  M
1501 ATTGATAAAATCAAAAATGCTGGATTTAAAACCATAAGAATACCAACTACTTGGGGAGAACACTTAGATGGTAATAACAAACTTAATGAAGAATGGGTAA
       I  D  K  I  K  N  A  G  F  K  T  I  R  I  P  T  T  W  G  E  H  L  D  G  N  N  K  L  N  E  E  W  V  K
1601 AGAGAGTTAAAGAAGTTGTTGATTATTGTATAGCAGATGATCTTTATGTTATCTTAAATACTCATCATGAAGGAAACTGGGTTATTCCAACTTACGCTAA
         R  V  K  E  V  V  D  Y  C  I  A  D  D  L  Y  V  I  L  N  T  H  H  E  G  N  W  V  I  P  T  Y  A  K
1701 AGAATCTTCAGTAACTCCAAAACTAAAAACTCTTTGGACTCAAATATCTGAAGCTTTCAAAGATTATGATGATCATTTAATATTTGAAACTCTTAACGAA
         E  S  S  V  T  P  K  L  K  T  L  W  T  Q  I  S  E  A  F  K  D  Y  D  D  H  L  I  F  E  T  L  N  E
1801 CCAAGACTTGAAGGAACTCCTTATGAATGGACAGGTGGTACAAGTGAATCTCGTGATGTTGTAAATAAATATAACGCAGCTGCTTTAGAATCTATAAGAA
       P  R  L  E  G  T  P  Y  E  W  T  G  G  T  S  E  S  R  D  V  V  N  K  Y  N  A  A  A  L  E  S  I  R  K
1901 AAACTGGTGGTAATAACTTATCTAGAGCTGTTATGATGCCAACTTATGCAGCTTCTGGTTCATCTACTACAATGAATGATTTTAAAGTTCCAGATGATAA
         T  G  G  N  N  L  S  R  A  V  M  M  P  T  Y  A  A  S  G  S  S  T  T  M  N  D  F  K  V  P  D  D  K
2001 AAATGTAATTGCATCTGTTCATGCATATTCTCCTTACTTCTTTGCTATGGACACAAGCAGCAATTCAGTTAACACATGGGGAAGTTCTTATGACAAATAT
       N  V  I  A  S  V  H  A  Y  S  P  Y  F  F  A  M  D  T  S  S  N  S  V  N  T  W  G  S  S  Y  D  K  Y
2101 TCTTTAGATGTAGAATTAGATTCTTACTTAAATACTTTTAAATCTAAAGGAGTTCCTGTTGTTATTGGTGAATTTGGTTCTATAAATAAAAATAATACTT
       S  L  D  V  E  L  D  S  Y  L  N  T  F  K  S  K  G  V  P  V  V  I  G  E  F  G  S  I  N  K  N  N  T  S
2201 CTTCAAGAGCAGAGCTTGCTGAATATTATGTTACAGCTGCTCAAAAGAGAGGTATTCCTTGTGTATGGTGGGATAACAACTATGCTGAAACTAATAAGGG
         S  R  A  E  L  A  E  Y  Y  V  T  A  A  Q  K  R  G  I  P  C  V  W  W  D  N  N  Y  A  E  T  N  K  G
2301 TGAAACTTTTGGATTACTAAATAGAAGTACTTTAAACTGGTACTTTAGTGATATTAAAGATGCTTTAATTAGAGGATATAAAAATGTACATCCTGAAGCT
       E  T  F  G  L  L  N  R  S  T  L  N  W  Y  F  S  D  I  K  D  A  L  I  R  G  Y  K  N  V  H  P  E  A
2401 ACTGAAGATGATAAACCAAGCACAGATGTAACCAACCCTGATTCAGGCAATACTAAACCAGACAGTGGTAATACTAATCCTGGTACAGAAACTACTACTC
       T  E  D  D  K  P  S  T  D  V  T  N  P  D  S  G  N  T  K  P  D  S  G  N  T  N  P  G  T  E  T  T  T  P
2501 CAACAGATAATGAAAAGATTTCAATAACTTCAAAAATTAATGATTGGGGCGGTGCTTACCAAGCTGATTTCACTTTAAAAAATAACACTTCAAGTGACAT
       T  D  N  E  K  I  S  I  T  S  K  I  N  D  W  G  G  A  Y  Q  A  D  F  T  L  K  N  N  T  S  S  D  I
2601 TAACAATTGGTCTTTTAAGATAAAGAAAAATGATATTGTGTTTACTAATTACTGGGATGTTAAGATAACTGAAGAAAATGGATATTATGTTGTTACACCT
       N  N  W  S  F  K  I  K  K  N  D  I  V  F  T  N  Y  W  D  V  K  I  T  E  E  N  G  Y  Y  V  V  T  P
2701 CAAGCTTGGAAAACAACTATCCTCGCAAATTCTTCTATAGTTATTTCTATTCAAGGCACTGGAAAAGTAATTAGTAATTTTGAGTATAAATTTGACTAGA
       Q  A  W  K  T  T  I  L  A  N  S  S  I  V  I  S  I  Q  G  T  G  K  V  I  S  N  F  E  Y  K  F  D  *
2801 TAACCTAAATAAAATTTAAGCCTAAACAATATAACTTCTATCTTGTTTAGGCTTAAATTATACTACTCGTCCATCATATTACGGTTATTATATGAGTACA
2901 CATTATAATCTCTCCAAATAAGTTCAAGATTTGTAAATGTATCTGAAATACTTTCCTTTTCTCTCATTCCTACTGCTATTATAAGTGCATTTATTACTGA
3001 TAAAGGTGCTACTAATGAATCAACAAATGAAGCCATATTACTTTGAGCTATTAACGTATAATCTGCATGTGAAGCAAGTGGAGATAATAAACTATCTGTA
3101 AGTGCTAAAACCTTAGCTCCTCTTCTCTTAGAAAAATCTAAGGCATCTATAGTTTTTGTTGCATATCTTGGGAAACCAATCCCTATAACTAAATCTCCTT
3201 CTCCCACATTAATCATTTGTTCAAAAATGTCTGAAATTCCGTAGCTTACTATTCTTACATTTTGAAGAATTATATTCAAATAAAATCCTAAGAATTCAGC
3301 AAGAGCTGTAGAACTTCTAAGGCCTATTATATATATTTTCTTTGCTTCAAAAATTGAATTTAAAACTTGTTCAAAGGTATAAGGGTTTATCTTTTCAAGA
3401 GTTGCTCTTATATTTTCCATATCTGCCTTTAATACTCCCTTTAAAGTATCTCCATCAGAAATAAGGTCCTTTGAAAGCTCTAATCTTTGAACAGTTGTTA
3501 GCTTATTTTTTATTAGCTCTTGAAGAGCTTTTTGTAATTTAGGATAACCTAAAAATCCTAATTCATTTGCAAATCTTACTACTGTAGATTCAGATACACC
3601 TACAGAATCTCCAAGCTTTG
```

Fig. 2.5. Nucleotide sequence of *celA* and adjacent regions. The derived amino acid sequence is presented. The leader sequence is underlined, the putative linker region is double-underlined and the CBD is underlined with a dashed line. The arrow at position 2070 in the nucleotide sequence indicates the extent of *celA* that is retained by *celAΔ3'* on pCMJI and pCMJII. Numbering corresponds to GenBank entry LO2868.

Béguin (1991) until the classification system was changed (refer to section 1.5). The *celE* gene encoding EGE of *C. thermocellum* (Hall *et al.*, 1988) shows 59 % DNA sequence homology in a 1040 bp fragment and 48.4 % amino acid homology in a 374 amino acid fragment with *celA* and CelA, respectively. Sequence alignments with other family 5-endoglucanases indicated that Glu286 of mature CelA (encoded by nucleotides 2170-2173 in Fig. 2.5) is probably the nucleophile. This glutamate, which is conserved in all enzymes in family 5, was reported to be the nucleophile in endoglucanase CelC from *C. thermocellum* (Wang et al., 1993b).

2.4.4 Identification of the transcriptional start site

Figure 2.6 shows the results of two separate primer extension and termination reactions separated by a sequencing ladder and flanked by ddA and ddT termination ladders in lanes 2 and 11, respectively. Lane 1 constitutes a negative control, where RNA isolated from *E. coli* BL21 was hybridized with pCMP ssDNA and the end-labelled primer. Since no *celA* mRNA:ssDNA complex could form, the T4 DNA polymerase synthesized long extension products from the annealed primer i.e. no extension termination occurred in the vicinity of the 5' end of the *celA* ORF. Lanes 3 and 10 show termination bands formed by the T4 DNA polymerase on a pCMP ssDNA template annealed with *E. coli* BL21 (pCML) RNA. Since this strain carried the *celA* gene on the high copy number plasmid vector pSK(M13+) (Fig. 2.2), large quantities of the *celA* mRNA transcripts were present and hybridized to the complementary *celA* on the pCMP ssDNA. The termination band in lane 10 as indicated by the arrow is very faint because very little pCMP ssDNA was present in the reaction. Lanes 4 and 9 show termination bands formed by the T4 DNA polymerase on the pCMP ssDNA template annealed with *C. longisporum* RNA. The *C. longisporum* culture was grown on NRF medium containing barley ß-glucan to induce the *celA* gene (see induction studies under section 4.4.1). The termination bands visible in lanes 9 and 10 clearly correspond with the adjacent T bands in lanes 8 and 11. Taking into consideration the skewness of the sequencing ladder, the termination bands in lanes 3 and 4 correspond with the bands in lanes 9 and 10. Minor termination bands are visible in lanes 3,4 and 9, but these were ignored as they did not constitute such definite signals as the major one. These bands indicate the length of the extension products which terminated one base upstream of the first nucleotide of the mRNA molecules hybridized to the ssDNA template. Thus the C nucleotide 27 bases upstream of the ATG start codon of the *celA* gene was identified as the first nucleotide of the *celA* mRNA. Another important finding was that the mRNA transcripts in *C. longisporum* and in *E. coli* started at the same base.

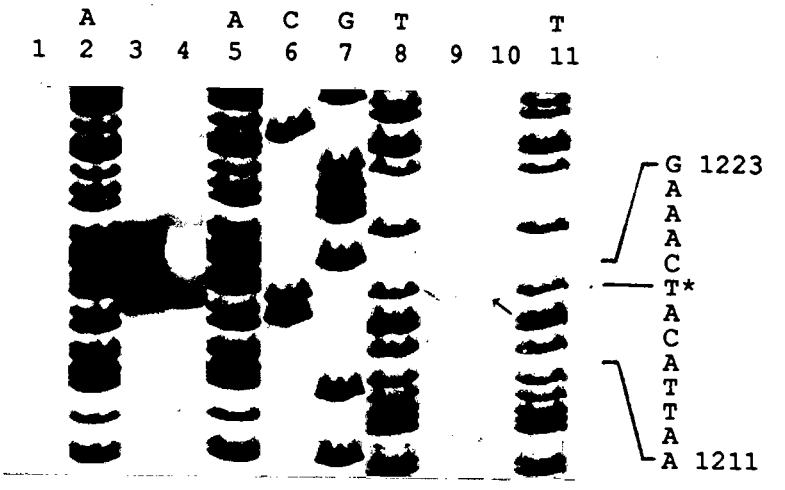


Fig. 2.6 Primer extension to map the transcriptional start site. Lanes 2, 5-8 and 11 show the sequencing ladder obtained from double-stranded DNA dideoxy termination sequencing. Lanes 1, 3, 4, 9 and 10 show the primer extension and termination reactions which contained mRNA from *E. coli* or from *C. longisporum*. Lane 1 shows the *E. coli* negative control; lanes 3 and 10 contained *E. coli* BL21 (pCML) mRNA and lanes 4 and 9 contained *C. longisporum* mRNA. The arrow indicates the position of a very faint band. The sequence of the region of interest is shown on the right and the T* corresponds with the main termination band in lanes 3, 4, 9 and 10.

2.4.5 Identification of the -10 signal, a direct repeat and the ribosome binding site

The exact determination of the transcriptional start site enabled the identification of an AATAAT –10 RNA polymerase recognition sequence which closely resembled the *E. coli* σ^{70} consensus sequence TATAAT (Fig. 2.7). No *E. coli* –35 region could be identified, but a 12–bp direct repeat was found flanking this region. The TATTGAATTTAT direct repeats are separated by 15 bases. Analysis of the 1218 bp of *C. longisporum* DNA upstream of *celA* revealed that this is the only significant repeat present. 11 bases downstream of the start of the mRNA transcript and 12 bases upstream of the ATG start codon a possible Shine-Dalgarno ribosome binding site, AGGGAGA, was identified.

```
                                                    1106 TAATTTTTAT TAATAACAAT
GACTAACTTT ATGTATTAAT AGCACACTTT CACGAATTAT TATTGAATTT ATTTCTTTAA
                          -10          1          S-D
GATATTATAT TGAATTTATG TAAATAATTA CATCAAAGAA ATAAGGGAGA TGATTAAACT

ATG AAA AGA TCT TTA TTA AAA ACT TGT TCT ATC ATT GCT GGA GCA ACA
 M   K   R   S   L   L   K   T   C   S   I   I   A   G   A   T
ATT ATT TTC TCA TCA CTT TCA ATT TCA AGA AAT CCT CTT GAA GTA 1338
 I   I   F   S   S   L   S   I   S   R   N   P   L   E   V
```

Fig. 2.7. Nucleotide sequence of the 5'-end of *celA*. The direct repeats are underlined, the –10 promoter region is bold and in italics. The transcription start site is in bold and numbered 1. The Shine-Dalgarno site is overlined. The translated amino acid sequence is shown. The sequence numbering corresponds with GenBank entry L02868.

2.4.6 Identification of transcription termination signals

Fifteen bases downstream of the stop codon of the *celA* gene a stemloop structure was identified (Fig. 2.8 A) which probably constitutes a transcription termination signal. Judging from the size of the full-length *celA* transcripts on the northern blot (see section 4.4.1 and Fig. 4.1) it is evident that *celA* forms a monocistronic transcription unit since the *celA* ORF is 1548 bp long. Most cellulase genes have been found to be transcribed in monocistronic units (Béguin & Aubert, 1994) and stemloop structures have been identified at the 3' ends of many cellulase genes (section 1.11; Béguin *et al.*, 1986; Greenberg *et al.*, 1987a; Moser *et al.*, 1989).

Another smaller stemloop structure was identified within the *celA* gene at the 5' end of the CBD (Fig 2.8 B and 2.8 C). Even though the free energy of this stemloop is much less than that of the stemloop at the 3' end of *celA*, it appears to be active in *C. longisporum*. The evidence for this is the faster migrating faint band on the Northern blot in Fig. 4.1. The two bands differ in size by about 280 (+/-25) bases as determined from a standard curve using the RNA size markers on the blot. The calculated sizes of the full-length *celA* mRNA and the truncated transcript are 1645 and 1385 bases, respectively, the size difference between them being 260 bases assuming that transcription started at the site as determined below and stopped within 5 bases from the end of the terminators. The "tailoring" of the *celA* mRNA to produce only the catalytic domain could be an alternative and metabolically less taxing mechanism to the proteolytic cleavage between the catalytic and cellulose-binding domains, which has been described for a number of ß-glucanases (Gilkes *et al.*, 1988, 1989, 1991; Hall *et al.*, 1989; Jauris *et al.*, 1990; Fierobe *et al.*, 1991; Hansen *et al.*, 1992). More experimental evidence would be needed to substantiate this speculation. Neither stemloop terminators in *C. longisporum* seem to be of the rho-independant type, since the region of dyad symmetry is not followed by a run of U bases (Watson *et al.*, 1987). The stemloop terminator of the *C. thermocellum celA* gene does have such a run of U bases (Béguin *et al.*, 1986), but *cenA*, *cex* and *cenC* of *C. fimi* do not (Greenberg *et al.*, 1987; Moser *et al.*, 1989).

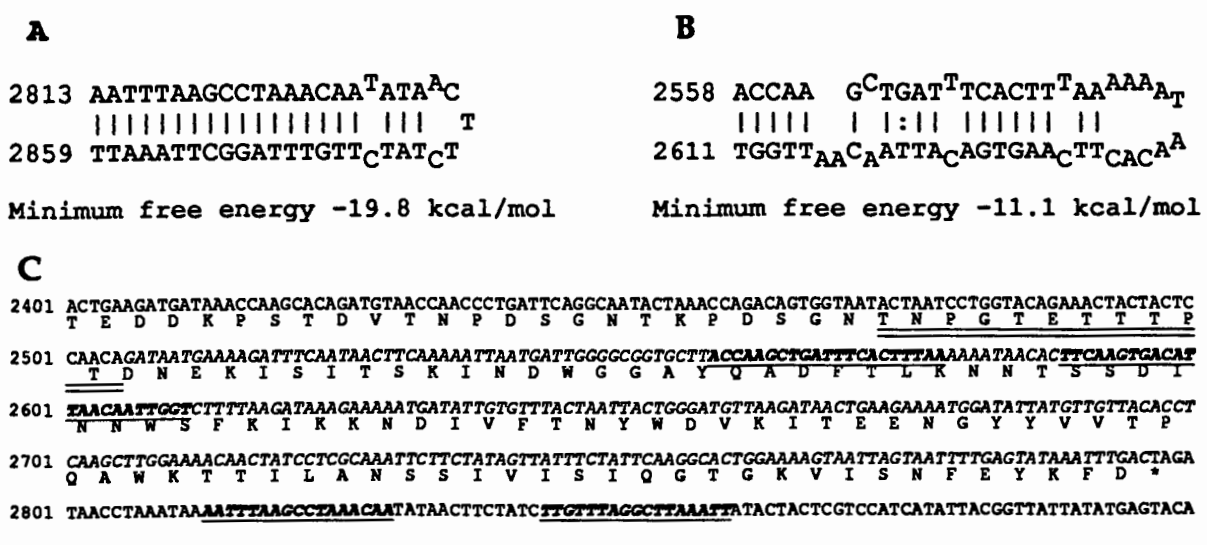


Fig. 2.8. Putative transcription termination signals (A) at the 3' end of *celA* and (B) within the CBD of *celA*. The minimum free energy was calculated with the FOLD software (Zuker & Stiegler, 1981; Genetics Computer Group, 1991). **(C)** Nucleotide sequence of the 3' end of *celA*. The two transcription termination signals are underlined and in bold. The CBD is italicized and the putative linker region is double-underlined. The derived amino acid sequence is shown. The sequence numbering corresponds with GenBank entry L02868.

2.5 DISCUSSION

2.5.1 Growth of *C. longisporum* and production of an antibacterial agent

Growth of *C. longisporum* in batch culture in liquid NRF medium is characterized by the onset of cell death and the complete loss of viability when the pH of the medium gets too low (≤ pH 5.5) due to the accumulation of acidic fermentation products. In the rumen environment the pH is maintained at a level of 6.0-6.7 by the alkaline saliva copiously secreted into the mouth of the ruminant during feeding and at a lower rate during other periods. In this way much larger concentrations of rumen bacteria are achieved in the rumen than can be obtained in batch culture (Hungate, 1988). The detrimental effects of low pH on the growth of rumen bacteria are well documented (Russel & Dombrowski, 1980; Miyazaki *et al.*, 1992). When ruminants are fed high starch diets, certain amylolytic and lactate-utilizing rumen bacteria like *Selenomonas ruminantium* and *Streptococcus bovis* flourish, which leads to the increased production of lactic acid which in turn causes a decrease in the pH of the rumen fluid. This effect is known as lactic acidosis in ruminants. It causes a decrease in fibre digestion in the rumen due to the washout of the main cellulolytic ruminal bacteria (Dehority & Orpin, 1988; Stewart & Bryant, 1988).

During a growth curve experiment in which the culture had become contaminated with an unidentified bacterium, it was noticed that *C. longisporum* produced an antibacterial agent which inhibited the growth of the contaminant on plates and in liquid medium. Initially the contaminant was present in large numbers, but *C. longisporum* growth occurred normally and eventually the culture reached its normal maximum cell concentration, while the contaminant disappeared. On the plates fairly large zones of clearing around the *C. longisporum* colonies were visible against a lawn of contaminants; when the *C. longisporum* colonies were only as big as pinheads, the zones of clearing were about 1 cm in diameter. No further work was done to investigate this phenomenon, but it might an interesting topic for another project. The production of bacteriocin-like substances by rumen staphylococci has been investigated by Lauková & Mareková (1993) who found that the inhibitory agents showed a wide range of activity against Gram-positive and -negative indicator organisms from different sources. The bacteriocin-like substances were stable and sensitive to trypsin, susceptible to chloroform and heat-sensitive. Bacteriocin production by rumen bacteria may represent an adaptation to a dynamic and competitive

growth environment. A different competitive mechanism was found in *Ruminococcus albus* and *R. flavefaciens* cultures which produced proteins that inhibited the activity of cellulases of the ruminal fungus *Neocallimastix frontalis*, but did not affect fungal growth (Stewart *et al.*, 1992; Bernalier *et al.*, 1993).

C. longisporum produced a great deal of extracellular polysaccharides (EP) during late log and early stationary growth phases. The production of EP by another rumen bacterium, *Butyrivibrio fibrisolvens*, has been investigated. Anderson *et al.* (1993) determined the chemical composition and structure of the capsular polysaccharide of *B. fibrisolvens* X6C61, while Wachenheim & Patterson (1992) determined the nutritional requirements and conditions for EP production by *B. fibrisolvens* nyx. They found that conditions which improved growth usually improved EP production, but EP was not produced in the lag or stationary phase. With *C. longisporum* it was observed that insoluble cellulose particles became entrapped in the EP and this author speculates that the rumen bacteria secrete the EP for that specific purpose. The buoyant mat of microbial mass and feed particles on the liquid surface of the rumen content is probably consolidated by EP to facilitate close interaction between bacteria and the feed particles and to prevent the excessive removal of microbial mass out of the rumen into the omasum (Stewart & Bryant, 1988).

2.5.2 Sequence analysis of *celA*

celA is the first gene from *C. longisporum* B6405 to be cloned and analyzed since the organism has only recently been reisolated from the rumen of a bison by Varel (1989). The gene was sequenced and found to belong to the large and well characterized family 5 of endoglucanases [EC 3.2.1.4] as reviewed by Henrissat & Bairoch (1993). To date this family contains about 60 cellulases, all of which are proposed to hydrolyze their substrates with the retention of configuration at the anomeric carbon. The DNA sequence data was confirmed by the determination of the N-terminal amino acid sequence of purified CelA and CelAΔN'. The derived amino acid sequence of CelA shows high homology with EGE from *C. thermocellum* and other endo-1,4-ß-glucanases from ruminal bacteria, *B. fibrisolvens* H17c, *R. albus* F-40 and *R. albus* SY3, that are closely clustered on the same branch in the unrooted phylogenetic tree of family 5 (Family A, Gilkes et al., 1991a; Workshop on nomenclature of cellulase, 1994). Endoglucanases BrEND from *Prevotella ruminicola* and CelB from the rumen fungus *Neocallimastix patriciarum* also show good homology with all

of these endoglucanases (Vercoe & Gregg, 1992; Zhou *et al.*, 1994). The lack of introns in the *celB* gene, which was obtained through cDNA cloning, and the homology with other rumen bacterial endoglucanases points to the possibility that the ancestral gene was transferred between rumen colonizers and that the horizontal transfer from prokaryotes to the eukaryotic fungus took place. Similarly it was also speculated that the *xynA* gene from *N. patriciarum*, encoding xylanase XYLA, had been transferred horizontally between rumen prokaryotes and lower eukaryotes (Gilbert *et al.*, 1992).

2.5.3 Transcription start site of the *celA* mRNA transcript

The primer extension experiments revealed that *C. longisporum* and *E. coli* use the same transcription initiation site 27 bp upstream of the initiation codon. It was not determined whether *C. longisporum* utilized separate transcription initiation sites for the induction of *celA* during growth on ß-glucan and for the low constitutive expression of *celA* during growth on cellobiose (see induction studies in section 4.4.1 and 4.5.1). The main reason for this was the inability to detect *celA* mRNA on northern blots during growth on cellobiose. In the *R. albus* EgI gene transcription is initiated 59 bp upstream of the ATG codon in both the parent strain and *E. coli* (Ohmiya *et al.*, 1989), but *celA* from *Prevotella ruminicola* utilizes different transcription initiation sites (Vercoe & Gregg, 1991). Both of these endoglucanases belong to the same family of glycosyl hydrolases as CelA from *C. longisporum*. *C. thermocellum celA* and *celD* (Béguin *et al.*, 1986; Mishra *et al.*, 1991) and *C. fimi cenA* and *cenB* (Greenberg *et al.*, 1987a, 1987b) were shown to make use of two different promoters depending on the host or on the culture conditions.

Utilization of the same transcription initiation sites by *C. longisporum* and *E. coli* leads to the speculation that perhaps some regulatory proteins from *E. coli* are able to bind to the upstream regulatory regions of *celA*. No experiments were done to investigate the possible regulation of *celA* in *E. coli*, but titration of regulatory proteins might cause the severe plasmid instability problems which were experienced with the multicopy plasmid constructs pCMI and pCML (results not shown). Problems with plasmid instability and cell lethality have also been experienced when the endoglucanase *celB* gene from *R. albus* SY3 was cloned into a pUC high-copy number vector (Poole *et al.*, 1990) and also with *xynA*, *xynB*, *celA* and *celB* from *P. fluorescens* subsp. *cellulosa* (Ferreira *et al.*, 1991).

2.5.4 Regulatory features of *celA*

The determination of the transcription start site led to the identification of a –10 promoter binding site and the identification of direct repeats upstream of the *celA* gene. The *R. albus* EgI gene contains an 11-bp inverted repeat 44 bp upstream from its transcription initiation site (Ohmiya *et al.*, 1989). Indirect repeats upstream of a number of fungal cellulase genes showing homology with each other have been reported, but no sequence similarities could be found with the *C. longisporum* and the *R. albus* repeats (Fernándes-Abalos et al., 1992; Jung et al., 1993). A protein has been identified in *Thermomonospora fusca* cell extracts which binds to the 9-bp inverted repeats upstream of *celE* and has been found to resemble an activator protein involved in cellulase induction (Lin & Wilson, 1988).

CHAPTER III

PURIFICATION AND CHARACTERIZATION OF ENDOGLUCANASE CelA

TABLE OF CONTENTS

3.1 SUMMARY

Two proteins, CelA and CelAΔN', were investigated. The full-length *celA* gene on pCML encoded CelA which contained a signal sequence, while the 5'-truncated *celAΔ5'* gene on pCMI encoded an N-terminal fusion protein, CelAΔN', without a signal sequence. Both proteins were purified and determination of their N-terminal amino acid sequences confirmed the DNA sequences of the genes encoding them. CelA was secreted into the periplasm in *E. coli*. In *E. coli* proteolytic cleavage of CelA at or near a putative linker region resulted in the appearance of two active polypeptides of Mr 57 and 47 kDa. The former was the full-length enzyme, while the latter was the catalytic domain from which the cellulose-binding domain (CBD) had been removed (CelAΔCBD). In *E. coli* BL21, which lacks the OmpT and La proteases, the proteolytic degradation of recombinant CelA was greatly reduced. The intracellularly located CelAΔN' was not subject to proteolytic degradation. The pH and temperature optima of CelA were pH 4.8 and 43°C, respectively. CelA had activity on barley ß-glucan, lichenan, CMC and xylan. It showed preferential activity on the larger cello-oligosaccharides (cellohexaose and cellopentaose), while cellotetraose was the smallest substrate degraded completely.

3.2 INTRODUCTION

The purification of cellulases and determination of their substrate specificities, reaction mechanisms and three-dimensional structures has helped to reveal some of the molecular aspects of the biological degradation of lignocellulose. Comparison of the amino acid sequences and 3-D structures of representative enzymes from the different families of glycosyl hydrolases (Henrissat & Bairoch, 1993) have led to the identification of the active site and catalytic residues as well as residues participating in the binding of the substrate (Juy *et al.*, 1992; Divan *et al.*, 1994). Some external factors which influence enzyme activity have been elucidated e.g. in the case of the *T. reesei* endo-ß-xylanase changes in the conformation of the active site were observed in response to the external pH, which explained why the enzyme was more active at pH 5.0 than at 6.5 (Törrönen *et al.*, 1994). Divalent cations, e.g. Ca^{2+}, are important for the stabilization of secondary structures of some cellulases, e.g. three Ca^{2+} binding sites were found in the crystal structure of CelD from *C. thermocellum* (Juy *et al.*, 1992). In the case of the large family 5 of ß-hydrolases

(Henrissat & Bairoch, 1993), to which CelA from *C. longisporum* belongs (section 2.4.3) no 3–D structures are available yet, but CELCCA from *C. cellulolyticum* has been crystallized (Roig *et al.*, 1993).

In this chapter the purification and characterization of endo-ß-glucanase CelA from *C. longisporum* is described (Mittendorf & Thomson, 1993). Two proteins, CelA and CelAΔN', were purified out of the periplasm and the intracellular fraction of *E. coli*, respectively. CelA was found to contain a cellulose-binding domain, which was proteolytically cleaved off in *E. coli*. The effects of pH, temperature and divalent cations on CelA activity were determined.

3.3 MATERIALS AND METHODS

3.3.1 Enzyme localization studies

E. coli BL21 (F- *hsdS gal*) (Studier & Moffat, 1986) was used for the enzyme localization studies. *E. coli* BL21 carrying pCMI and pCML were inoculated from single colonies and grown in LB containing 200 µg/ml ampicillin (Ap). IPTG was added to a final concentration of 0.5 mM during the logarithmic phase of growth to induce the lac operon, since ß-galactosidase was used as a marker enzyme. The cultures were harvested at OD_{600} = 1.0–1.4. Due to the instability of both plasmids, serial dilutions of the cultures were plated on CMC-containing LB medium with and without antibiotic selection to ascertain that all cells contained the plasmids and were expressing the *celA* gene. The cells were harvested and fractionated by the osmotic shock procedure (Nossal & Heppel, 1966). The resulting cellular fractions were assayed for CMCase, ß-galactosidase and alkaline phosphatase activity.

3.3.2 Enzyme assays

CMCase activity was measured using the DNS method as recommended by the Commission on Biotechnology (Ghose, 1987), but the volumes were scaled down to permit the use of Eppendorf tubes. Assays were done at 43°C in 50 mM sodium acetate buffer

pH 5.0. The purified CelA and CelAΔN' samples were diluted in 50 µg/ml BSA to stabilize the enzymes. ß-galactosidase and alkaline phosphatase activities were assayed according to Miller (1972) and Brickman & Beckwith (1975), respectively.

3.3.3 Purification of CelA

E. coli BL21 (pCML) was grown in 4 l LB medium containing 150 µg/ml Ap to a final $OD_{600} = 1.1$. A periplasmic extract, obtained by subjecting the cells to osmotic shock (Nossal & Heppel, 1966), was applied to a Q-Sepharose anion exchange column (Pharmacia) and the CMCase activity eluted in a linear NaCl gradient of 50 mM to 200 mM NaCl in 10 mM Tris.HCl pH 7.5. CMCase-containing fractions were pooled and analyzed by SDS-PAGE.

3.3.4 Purification of intracellular CelAΔN'

E. coli K514 (pCMI) was grown in 2 l LB medium containing 150 µg/ml Ap to a density of $OD_{600} = 1.1$. The cells were harvested, resuspended in 15 ml 50 mM NaCl, disrupted in a French pressure cell and centrifuged (39 000g for 20 min at 4°C) to obtain the cell free extract (CFE). The CFE was subjected to a pH precipitation step by the dropwise addition of 0.1 N HCl to a pH of 4.60. The suspension was centrifuged as above, and the pH of the supernatant increased to pH 7.0 by the addition of NaOH. The pH-precipitated extract was applied to a Q-Sepharose Fast Flow anion exchange column (as above). CMCase-containing fractions were pooled and applied to a Bio-gel P200 gel filtration column (BioRad). Fractions containing CMCase activity were analyzed on SDS-PAGE.

3.3.5 N-terminal amino acid sequencing

The N-terminal amino acid sequences of CelA and CelAΔN' were determined on a gas-liquid solid phase sequencer constructed as outlined by Hewick *et al.* (1981) and Brandt *et al.* (1984). The converted phenylthiohydantoin amino acids were identified by an

isocratic on-line HPLC system on a 3 x 250 mm 3 µ Lichrospher C18 (Bishoff) column (Lottspeich, 1985).

3.3.6 Cellulose binding studies

A periplasmic extract of *E. coli* LK111 (pCML) was prepared and applied to a Q-Sepharose anion exchange column as above. CMCase-containing fractions were pooled and used for the Avicel-binding experiments as described by Ferreira *et al.* (1990) with minor modifications. A 5 ml sample was mixed with 5 ml of 5 % (w/v) Avicel suspended in 0.1 M Tris.HCl pH 7.5 and incubated on a shaker at 0°C for 1 h. The suspension was filtered through a small column and the eluate containing the unbound proteins was kept, while the Avicel was washed twice with 10 ml 0.1 M Tris.HCl pH 7.5. The bound proteins were eluted with 5 % (w/v) SDS at room temperature. Samples were analyzed by SDS-PAGE.

3.3.7 Substrate specificities

The activity of purified CelA on barley ß-glucan, lichenan, laminarin and oat spelts xylan (all obtained from Sigma) was determined according to the CMCase assay described above, except that the substrates were used at the following concentrations: 1 % (w/v) ß-glucan, 0.5 % (w/v) lichenan, 1 % (w/v) laminarin and 1 % (w/v) xylan. In all the assays, except the xylanase assay, the release of reducing sugars was expressed as glucose equivalents using a glucose standard curve. For the xylanase assay a xylose standard curve was used.

3.3.8 HPLC analysis of cellodextrinase degradation

The breakdown products of cellobiose, -triose, -tetraose, -pentaose (Sigma) and cellohexaose (Merck) incubated with purified CelA were analyzed. The reactions were performed at 43°C and contained 8 mg/ml substrate, 30 mM NaCl, 8.1 mM sodium acetate buffer pH 5.0 and 0.5 units/ml of CelA. Aliquots (15 µl) of the reactions were analyzed at

specified time intervals on a Beckman HPLC system with a μBondapak C18 column (Waters). HPLC-grade H_2O was used as solvent and the cellodextrins were eluted in a flow rate gradient increasing from 0.5 to 4 ml/min.

3.3.9 Divalent cation inhibition study

Purified CelA was dialyzed for 24 h with 3 buffer changes at 4°C against 2 l of 50 mM NaCl, 10 mM Tris.HCl pH 7.5 containing 0.1 mM EDTA to scavenge metal cations. CMCase activity was assayed as above in reactions that contained 200 μM unchelated metal cations.

3.3.10 SDS-PAGE and zymograms

SDS-PAGE was performed as described by Laemmli (1970). The 10 % (w/v) gels were stained with PAGE Blue 83 (BDH). Molecular weight standards were obtained from Pharmacia. Zymograms were performed as described by Béguin (1983) and modified by Sharma & Sandhu (1986). CMC (0.1 % (w/v)) was incorporated in the polyacrylamide gels. SDS and mercaptoethanol were added to the samples, which were not boiled prior to loading. After electrophoresis the gels were washed four times with 500 ml 10 mM LiCl for 15 min followed by four washes with H_2O to remove the SDS. After an over-night incubation at 43°C in 50 mM sodium acetate buffer pH 5.0 the CMCase activity was visualized by staining with Congo Red.

3.3.11 Protein assays

The BIO-RAD protein assay was used for the quantitation of proteins. BSA fraction V (Boehringer Mannheim) was used as a standard.

3.4 RESULTS

3.4.1 Enzyme localization of CelAΔN' and CelA

Fifty eight % of CelA was exported to the periplasm in *E. coli* BL21, while 94 % of CelAΔN', lacking the signal sequence, was retained in the cytoplasm (Table 3.1). This corresponded with information gained from the DNA sequences of the *celA* and *celA*Δ5' genes, which indicated that the former contained a signal sequence, while the latter did not (refer to section 2.4.3). N-terminal amino acid sequence determination of the purified CelA protein showed that the first 26 residues encoded by *celA* constitute a typical signal sequence which was cleaved off. The sequence of the first 5 residues of the mature CelA protein was determined and found to be Ile-Ser-Arg-Asn-Pro (Fig. 2.4 in previous chapter).

3.4.2 Purification of CelAΔN' and CelA and proteolytic cleavage of CelA

To aid purification, the copy numbers of the *celA*Δ5' and *celA* genes in *E. coli* were increased by subcloning them from the low-copy number vector pEcoR251 to the high-copy number vectors pUC19 and pSK to obtain pCMI and pCML, respectively (Figs. 2.1 and 2.2; plasmid maps in Appendix I). The intracellular CelAΔN' was purified from cell free extract (Table 3.2 A, Fig. 3.1 A), while CelA was purified from the periplasmic fraction (Table 3.2 B, Fig. 3.1 B). Unlike the intracellular CelAΔN', the periplasmic CelA was prone to proteolytic truncation when purified out of *E. coli* strains LK111 or K514. When CelA-containing fractions from the ion-exchange column were subjected to zymogram analysis, two CMCase bands with apparent mobilities of 57 and 47 kDa were detected (Fig. 3.2 A). The larger band corresponds to the full-length CelA protein while the smaller band, as will be shown later, is the product of proteolytic cleavage at or near the putative linker region (as indicated in Fig. 2.5) that leads to the removal of the cellulose-binding domain to form CelAΔCBD. To avoid this proteolytic degradation, *E. coli* BL21, which lacks the OmpT protease, localized to the outer membrane, and the La protease (product of the *lon* gene) (Chung & Goldberg, 1981; Grodberg & Dunn, 1988), was used for the purification of full-length CelA which was less contaminated with the truncated derivatives (compare lane 8, Fig. 3.1 B with lane 6, Fig. 3.2 B).

Table 3.1. Localization of CelAΔN' and CelA in *E. coli* BL21

Cellular fraction	ß-galactosidase*† (Units.ml^{-1}.min^{-1})		Alkaline phosphatase*†¶ (Units.ml^{-1}.min^{-1})		CMCase* (Units.ml^{-1})	
	pCMI‡	pCML‡	pCMI	pCML	pCMI	pCML
Culture supernatant	5.1 (0.11)	40 (0.31)	ND	0.011 (0.55)	ND	ND
First-stage osmotic shock supernatant	7.1 (0.15)	13 (0.10)	0.0483 (2.8)	0.0817 (4.1)	ND	ND
Periplasmic fraction	117 (2.5)	381 (3.0)	1.44 (84.3)	1.76 (87.9)	0.0397 (5.7)	0.2413 (58)
Intracellular fraction	4475 (97.2)	12400 (96.6)	0.221 (12.9)	0.15 (7.5)	0.6563 (94.3)	0.1723 (42)

* Enzyme activities expressed as % of total activity recovered are given in brackets.

† ß-galactosidase and alkaline phosphatase were used as intracellular and periplasmic marker enzymes, respectively.

¶ The alkaline phosphatase activities were very low because they were repressed by the high phosphate levels in the LB medium on which the cultures were grown. The assays were incubated at 37°C for 120 min to compensate for the low activity.

‡ pCMI and pCML encode CelAΔN' and CelA, respectively.

ND, not detectable.

Table 3.2. Purification of CelAΔN' from *E.coli* K514(pCMI) (A) and CelA from *E.coli* BL21(pCML) (B)

Purification step	Activity (CMCase U/ml*)	Specific activity (U/mg*)	Total activity (Units*)	Yield (%)	Purification (fold)
(A)					
Cell free extract	68	2.5	1134	100	1.0
pH-precipitated extract	39	4.3	535	47	1.7
Q-Sepharose anion-exchange (pooled fractions)	4.2	47	236	21	19
Bio-gel P-200 gel filtration	12	188	125	11	74
(B)					
Periplasmic extract	17.5	54.3	2830	100	1.0
Q-Sepharose anion-exchange - pool I	7.40	130	302	10.7	2.4
- pool II	8.56	147	624	22.0	46.8 2.7
- pool III	6.10	158	400	14.1	2.9

* 1 Unit = 1 µmol.min-1 of glucose equivalents liberated.

Three pools of CelA were obtained from fractions eluted from the anion-exchange column (Fig. 3.1 B). Pool I, which contained no CelAΔCBD, corresponded to the the major part of the protein peak, while pools II and III were collected before and after the main peak, respectively, and contained small amounts of CelAΔCBD. Pools II and III have a slightly higher specific activity than pool I (Table 3.2 B), because the truncated catalytic domain (CelAΔCBD) has a higher activity on CMC than the full-length enzyme. This could also be seen on the zymogram of CelA and CelAΔCBD (Fig. 3.2 A). Higher specific activity of the truncated catalytic domain on soluble substrates has also been reported for other cellulases prone to proteolytic cleavage (Durrant *et al.*, 1991; Fierobe *et al.*, 1991; Hansen *et al.*, 1992).

3.4.3 Cellulose binding studies

Cellulose-binding assays were performed and the full-length CelA (57 kDa) was found to bind to Avicel (Fig. 3.2 B, lane 8) and had to be eluted with 5 % (w/v) SDS. The proteolytic degradation product, CelAΔCBD (47 kDa), did not bind to the Avicel (lane 7). The faint CelAΔCBD band in lane 8 might be due to insufficient washing of the Avicel column, but it is also possible that CelAΔCBD has another weaker cellulose-binding region. CelA could be eluted from the Avicel column using 1 % (w/v) SDS. The least stringent non-denaturing conditions to elute CelA from the Avicel were found to be 1 % (w/v) urea at room temperature. Cold destilled H_2O did not elute CelA from the Avicel, in contrast to EGE from *C. thermocellum* (Durrant *et al.*, 1991), and guanidinium chloride caused only limited desorption (Gilkes *et al.*, 1988).

3.4.4 pH and temperature optima and stability of CelA

pH 4.8 was optimal for CMCase activity and 50 % CMCase activity was obtained at pH 4.3 and 6.8 (Fig. 3.3). It is interesting to note that the pH of the rumen content, which is 6.0-6.7, is not optimal for CelA activity. The temperature optimum was 43°C and corresponded closely to the temperature stability of CelA, which was determined by measuring CMCase activity after 2 hours of incubation of the enzyme at selected temperatures. CelA was found to be stable at 45°C, but higher temperatures reduced the enzyme activity rapidly (Fig. 3.4). The optimal temperature for CelA activity correspondes closely with the temperature of the rumen environment. .

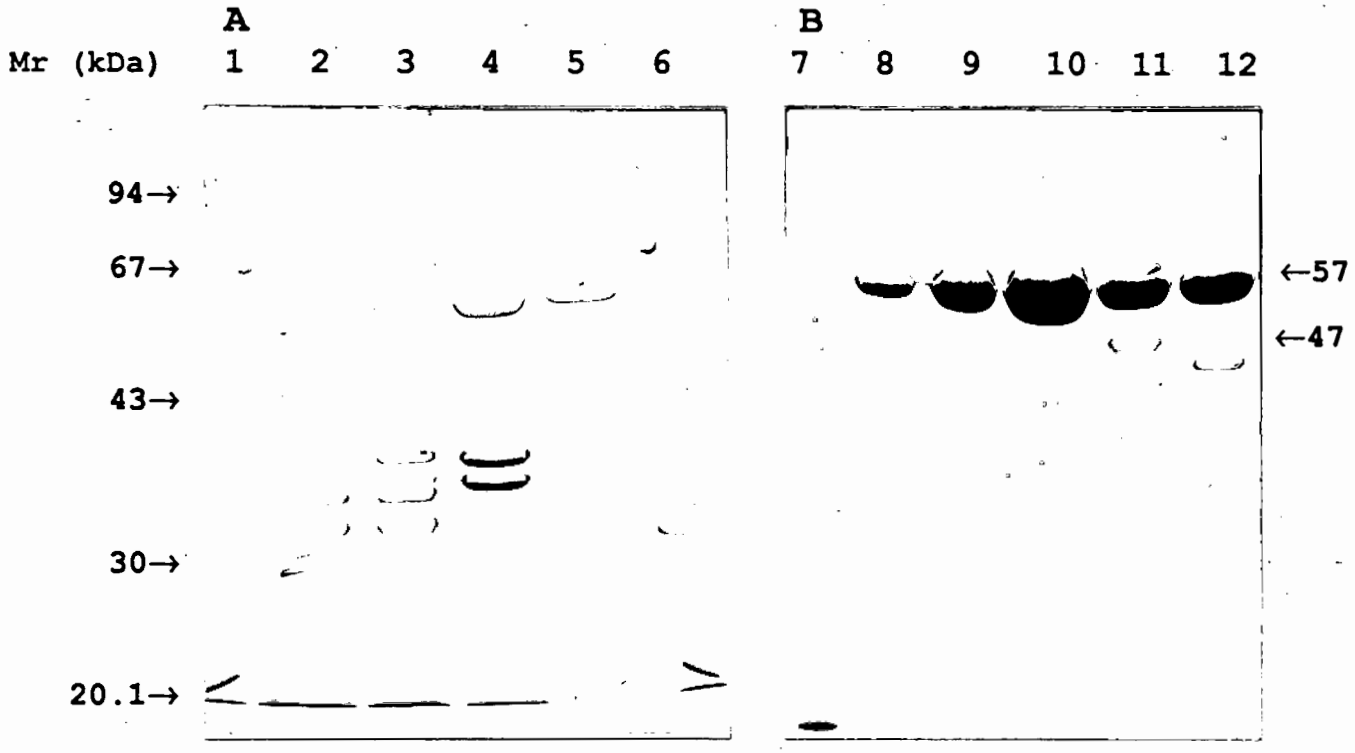


Fig. 3.1 SDS-PAGE of the purification of (A) CelAΔN' from *E. coli* K514 (pCMI) and (B) CelA from *E. coli* BL21 (pCML). Lanes: 1, 6 and 7, molecular size markers; 2, cell-free extract (80 μg); 3, pH–precipitated extract (65 μg); 4, pooled fractions from the ion-exchange column (20 μg); 5, pooled fractions from the gel-filtration column (3 μg); 8, 9 and 10, pool I fractions from the ion-exchange column (5.7 μg, 17 μg and 56.8 μg, respectively); 11, pool II (58.4 μg); 12, pool III (38.7 μg).

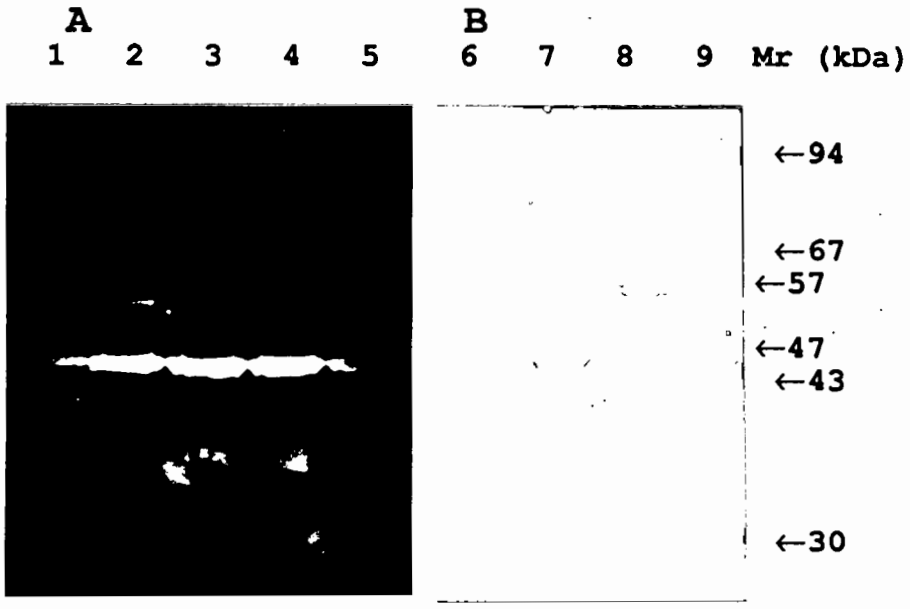


Fig. 3.2 Zymogram and SDS-PAGE analysis of the cellulose-binding study. These samples originated from *E. coli* LK111 (pCML). **(A)** Zymograms of fractions containing CelA and CelAΔCBD. Lanes 1-5 represent fractions eluting from the column corresponding to the centre of the peak of CMCase activity. **(B)** SDS-PAGE of the CelA and CelAΔCBD cellulose-binding study. Lanes: 6, CelA and CelAΔCBD from the ion-exchange column; 7, CelAΔCBD which did not bind to the Avicel; 8, CelA eluted from the Avicel column with 5 % (w/v) SDS; 9, molecular size markers.

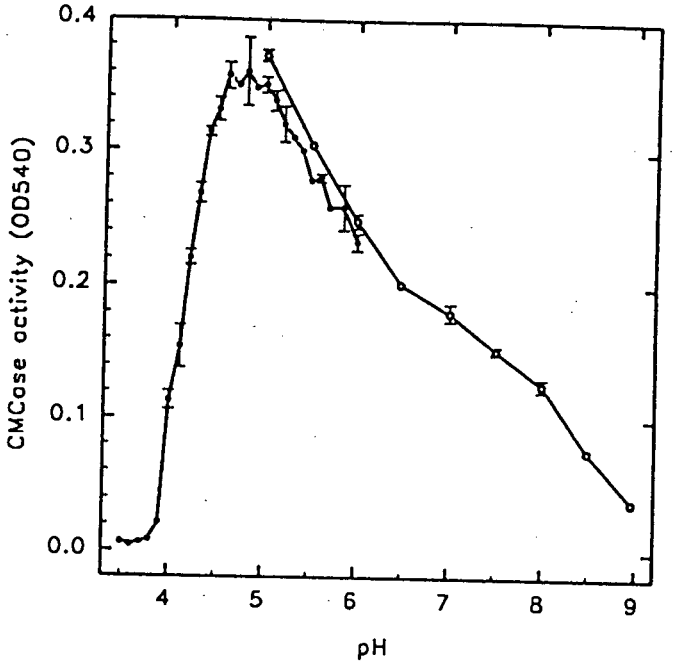


Fig. 3.3 pH optimum of CelA. CMCase activity was assayed in 50 mM sodium acetate buffer (• •) and in 50 mM phosphate citrate buffer (∘ ∘). The release of reducing sugars was measured using the DNS method and expressed as OD_{540nm} (Ghose, T.K., 1987). Measurements were done in duplicate.

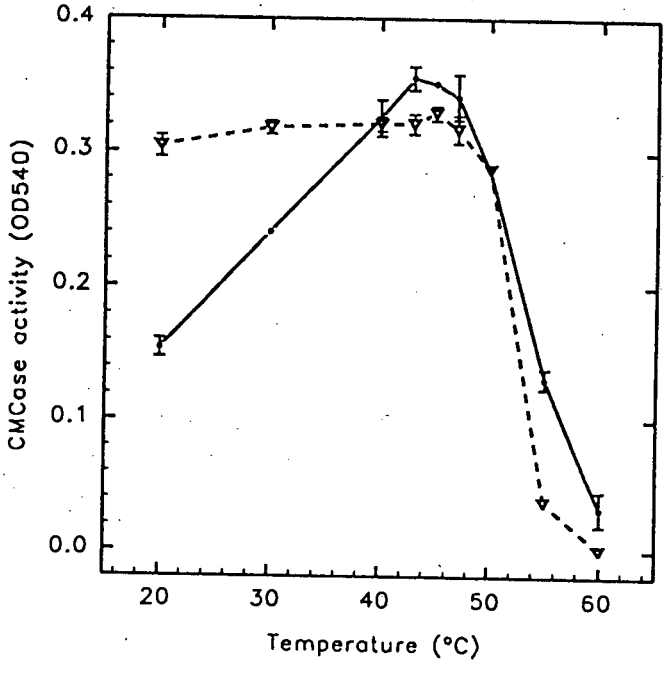


Fig. 3.4 Temperature optimum (• •) and stability (▽ ▽) of CelA. The temperature optimum was determined by assaying the CMCase activity after incubation of the enzyme with the substrate for 30 min at specified temperatures. The release of reducing sugars was measured using the DNS method and expressed as OD_{540nm} (Ghose, T.K., 1987). The temperature stability was determined by assaying the CMCase activity under standard conditions after an initial incubation of the enzyme at specified temperatures for 120 min. Measurements were done in triplicate.

3.4.5 Substrate specificities

CelA had the highest activity on barley ß-glucan, followed by lichenan and CMC (Table 3.3). Avicel and ball-milled filter paper (Whatman No. 1) were not hydrolyzed at all (data not shown). ß-Glucan is a soluble ß-linked polymer containing ß-1,4-linked segments interspersed with ß-1,3 linkages. CelA was not able to hydrolyze 1,3-ß-D bonds in laminarin, but it had low activity on xylan. The ratio of CMCase activity to xylanase activity of CelA is 50:1, while similar ratios of 10:1 and 100:1 have been reported for EGE and EGH from *C. thermocellum*, respectively (Béguin, 1990). CelA xylanase activity was evident on a xylan-containing zymogram (refer to section 4.4.4 and Fig. 4.4). Methylcellulose (0.04 % w/v) was found to inhibit CelA activity by 55 %, but 1 mM glucose and 1 mM cellobiose did not inhibit CelA (data not shown).

Table 3.3. Substrate specificities of CelA

Substrate	Linkages	Specific activity (units/mg) *
CMC	(1,4)-ß-D	188
Barley ß-glucan	(1,3)(1,4)-ß-D	1297
Lichenan	(1,3)(1,4)-ß-D	310
Laminarin	(1,3)-ß-D	0
Oat spelts xylan	(1,4)-ß-D	3.5

* 1 Unit = 1 µmol.min^{-1} of glucose equivalents liberated.

3.4.6 HPLC analysis of cellodextrin degradation

CelA had no activity on cellobiose (G2), but cellotriose (G3) was degraded partially to cellobiose and glucose (G1) after 24 h. Cellotetraose (G4) was degraded completely to G2 (85 %), G3 and G1 after 16 h. Cellopentaose (G5) was rapidly converted to G3 and G2 in 2 h (results not shown). The degradation of cellohexaose (G6) proceeded very rapidly (Fig. 3.5). G5 and G4 were formed as intermediate breakdown products during the degradation of G6, but the final products were G3, G2 and G1. CelA clearly demonstrates a preference for larger cellodextrins.

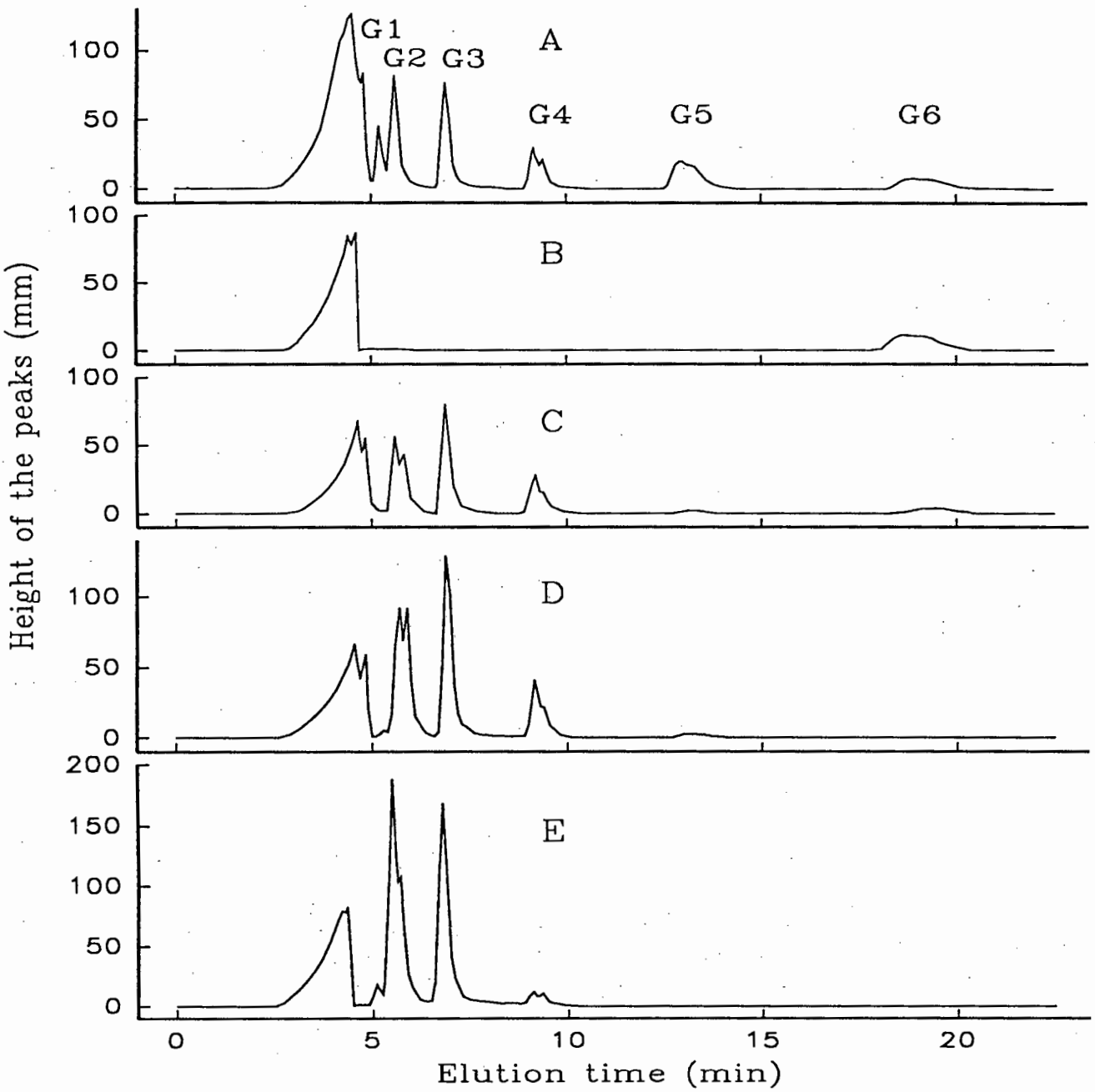


Fig. 3.5 HPLC analysis of cellohexaose (G6) degradation by CelA. A: G1 - G6 mixture resolved by the HPLC column. **B:** No enzyme control of G6 after incubation at 43°C for 170 min. **C, D and E:** G6 + CelA after 7.5 min, 15 min and 145 min incubation at 43°C, respectively. Reaction conditions and elution conditions were as set out in section 3.3.8.

3.4.7 Effects of divalent cations on CelA

Mn^{2+} and Hg^{2+} inhibited CelA, while Ca^{2+}, Ni^{2+}, Zn^{2+}, Mg^{2+} and Cd^{2+} had no significant effect (Table 3.4). Co^{2+} enhanced CelA activity by 44 %. The enhancing effect of Co^{2+} was also observed with the non-dialyzed enzyme preparation and 10 µM Co^{2+} was sufficient to enhance the activity by 20-30 % (results not shown). The addition of 1 mM 2-mercapto-ethanol or 0.5 mM DTT did not have an effect on the CMCase activity, thus decreasing the possibility that the Co^{2+} affects the redox potential (results not shown). The activity of a cellobiohydrolase (exo-glucanase) from the cellulosome of *C. thermocellum* and a ß-xylosidase from the rumen fungus *Neocallimastix frontalis* was also increased in the presence of Co^{2+} (Hebraud & Fevre, 1990; Morag *et al.*, 1991). It was speculated that the stimulatory effect of various divalent cations on cellulase activity was due to the stabilization of the enzymes. The exo-ß-glucanase from *Ruminococcus flavefaciens* FD-1 was moderately inhibited by Co^{2+} (Gardner *et al.*, 1987). Erfle & Teather (1991) investigated the divalent cation susceptibility of a 1,3-ß-D-glucanase from *R. flavefaciens* OR18, but none were found to have a stimulatory effect, while various others inhibited the enzyme.

Table 3.4. Effects of divalent cations on purified CelA[a]

Divalent cation	CMCase activity (% of control)
Control	100[b]
Mn^{2+}	28
Hg^{2+}	29
Ca^{2+}	85
Ni^{2+}	88
Zn^{2+}	89
Mg^{2+}	90
Cd^{2+}	90
Co^{2+}	144

[a] The values presented are averages of three assays. All metals were provided as chloride salts at 0.2 mM.

[b] 100% CMCase activity was equivalent to 9.87×10^{-3} CMCase units.

3.5 DISCUSSION

3.5.1 Molecular mass of CelA

The molecular mass of CelA of 57 kDa as determined by SDS-PAGE was found to correspond with the calculated molecular mass (57.5 kDa) of the gene product of the entire *celA* gene. This does not however take into account the cleavage of the signal sequence, which should increase the mobility of the mature protein on SDS-PAGE. The DNA sequence data was confirmed by the determination of the N-terminal amino acid sequence of purified CelA and CelAΔN'.

3.5.2 Secretion of CelA in *E. coli*

EGE of *C. thermocellum*, which is 48 % homologous to CelA, was found to be secreted into the periplasm of *E. coli* even when the signal sequence was removed (Hazlewood *et al.*, 1990). This was also found for the xylanase, XYLA, from *P. fluorescens* subspecies *cellulosa* (Hall *et al.*, 1989). In contrast the export of CelA into the periplasm of *E. coli* was found to be dependent on the presence of the N-terminal signal peptide. Even when the signal peptide is present translocation of CelA to the periplasm is incomplete (58 %). This could be ascribed to inefficient recognition of the *C. longisporum* signal peptide by the *E. coli* secretion system.

3.5.3 Modular structure of CelA

CelA was found to have a modular structure consisting of a N-terminal catalytic domain (CD) of 408 residues (including the signal sequence) and the C-terminal CBD of 97 residues joined by a putative linker region of 11 residues containing 2 proline and 6 threonine residues. In their review of the modular structure of ß-glycanases, Gilkes *et al.* (1991a) point out that the linker regions of many ß-glucanases are rich in proline and hydroxyamino acid residues. The deletion of 134 residues from the C-terminal end of the CD of CelA causes the loss of its hydrolytic activity, as *celA*Δ3' on pCMJI did not produce an active endoglucanase (see Fig. 2.5 for the extent of the deletion). The cellulose binding

study suggests that the CelA CD might contain another weak cellulose-binding region, because a small but significant amount of CelAΔCBD was still bound to the Avicel after two washes and eluted together with the full-length CelA in 5 % SDS. This small quantity of "bound" CelAΔCBD could however also be due to inadequate washing of the Avicel column.

The CBD of CelA shows amino acid sequence homology with other bacterial CBDs belonging to the type I family of CBDs (Béguin & Aubert, 1994). The CelA CBD does not however contain the two conserved cysteine residues which are present in other CBDs (except the one from *B. fibrisolvens*) of this family. These two residues have been proposed to form extended disulfide bonds in *C. fimi*-type CBDs (Gilkes *et al.*, 1991b).

The duplicated 24 amino acid segments present in most *C. thermocellum* cellulases and also in *C. cellulolyticum* EGA (Béguin, 1990) are absent in CelA. Since these repeats are thought to be required for the binding of the cellulases to the S_L protein, now termed CipA, of the cellulosome (Tokatlidis *et al.*, 1991; Fujino *et al.*, 1992; Salamitou *et al.*, 1992; Béguin & Aubert, 1994), it is possible that *C. longisporum* does not have such a cellulosome structure, but further experiments are required to determine this. Endoglucanase CelB from *Neocallimastix patriciarum*, which has high homology with CelA from *C. longisporum* (see section 2.5.2), has a non-catalytic domain containing short tandem repeats (Zhou *et al.*, 1994). Similar repeats are also present in xylanase A from the same organism (Gilbert *et al.*, 1992) and it is speculated that these might be neccessary for attachment to a scaffolding protein of an enzyme complex similar to the *C. thermocellum* cellulosome. An alternative model for bacterial cellulases, where the individual cellulases are not involved in binding to a multi-protein cellulolytic complex, has been proposed for *P. fluorescens* subsp. *cellulosa* and *C. fimi* (Béguin *et al.*, 1992; Hazlewood *et al.*, 1992; Meinke *et al.*, 1994). The absence of such repeats in CelA from *C. longisporum* does however not rule out the possibility that some multi-enzyme complex is formed. Some of the large inducible proteins of about 80-90 kDa and >120 kDa in the supernatant of *C. longisporum* (see section 4.4 A and Fig. 4.2.1) seem to lack activity against CMC, ß-glucan and xylan and could possibly form such complexes. More experiments will have to be done to obtain a clearer picture.

3.5.4 Proteolytic cleavage between the catalytic and cellulose-binding domains

Proteolytic cleavage between the CD and CBD has been described for a number of ß-glucanases (Gilkes *et al.*, 1988, 1989; Hall *et al.*, 1989; Jauris *et al.*, 1990; Fierobe *et al.*, 1991, 1993; Hansen *et al.*, 1992). In the case of *Cellulomonas fimi* and *Streptomyces reticuli* specific proteases, which were responsible for the processing of CenA, Cex and the Avicelase, were identified in the original hosts and characterized biochemically (Gilkes *et al.* 1988; Moormann *et al.* 1993). The difference in molecular masses of CelA and CelAΔCBD (calculated from the sequence data and determined by SDS-PAGE analysis) confirms that the proteolytic cleavage of CelA takes place in or near the proposed linker region, as is the case for many of these modular-type ß-glycanases (Gilkes *et al.*, 1991a). The use of *E. coli* BL21, which lacks the outer membrane-associated OmpT protease and the *lon* protease (Grodberg & Dunn, 1988), greatly reduced the cleavage of CelA, which was experienced with *E. coli* strains LK111 and K514, during the purification. It was interesting to note that the intracellularly situated CelAΔN' was not prone to proteolytic truncation in any of these *E. coli* strains. The OmpT protease has been shown to be responsible for the processing of various recombinant and homologous proteins in *E. coli*. It was found to cleave the proteins between two consecutive positively charged amino acids (Arg-Arg, Lys-Lys, Arg-Lys, or Lys-Arg) (Hanke *et al.*, 1992; Henderson *et al.*, 1994). A possible OmpT target sequence containing a pair of Lys residues is encoded by bases 2621-2627 within the CBD-encoding region. Neither the C-terminus of CelAΔCBD or the N-terminus of the proteolytically truncated CBD-fragment have however been sequenced, therefore the exact site where cleavage takes place in *E. coli* is unknown.

CHAPTER IV

EXPRESSION AND REGULATION OF *celA* IN *C. longisporum*

TABLE OF CONTENTS

4.1 SUMMARY

Northern blot analysis of RNA from *C. longisporum* revealed induction of transcription of the *celA* gene when barley ß-glucan was used as carbon source while no *celA* mRNA was detected after growth on cellobiose. Western blots, using rabbit antiserum raised against CelA protein purified from *E. coli*, revealed the extracellular location of CelA in *C. longisporum* and confirmed the transcriptional induction during growth on ß-glucan. Significant quantities of CelA were detected in the culture supernatant during growth on cellobiose, although no mRNA could be detected. This indicated a low constitutive expression of *celA*. Zymograms showed variable regulation of these glucanases in cultures grown on barley ß-glucan and cellobiose. ß-Glucan and xylan zymograms showed the expression of at least five major glucanases, termed CelA to CelE, in the culture supernatant.

4.2 INTRODUCTION

The regulation of cellulase production in the competitive and fluctuating rumen environment is important for the survival of the rumen microflora. A large number of cellulolytic and hemicellulolytic enzymes are produced by rumen microorganisms. These enzyme systems have been found to be regulated, and induction by their substrates, or degradation products thereof, has been reported e.g. in *B. fibrisolvens* (Williams & Withers, 1992), *R. flavefaciens* (Flint *et al.*, 1991; Doerner *et al.*, 1992; Wang *et al.*, 1993), a *Piromyces*-like ruminal fungus (Morrison *et al.*, 1990b), *Neocallimastix frontalis* (Barichievich & Calza, 1990) and *N. patriciarum* (Xue *et al.*, 1992a, 1992b). Catabolite repression by more readily fermentable substrates is used to down-regulate the enzyme levels (Hazlewood & Teather, 1988; Bèguin & Aubert, 1994). Constitutive expression of total endoglucanase activity and of selected enzymes has been reported for *Fibrobacter succinogenes* (Huang & Forsberg, 1990; McGavin *et al.*, 1990), but some enzymes seem to be regulated individually. A number of cellulase genes from rumen bacteria and fungi have been cloned (see section 1.13), but their transcriptional regulation in the original host has been investigated only in a few cases, e.g. *celA, celB, celC* and *celD* from *R. flavefaciens* (Doerner *et al.*, 1992; Wang *et al.*, 1993) and *celA, celB, celC* and *celD* from *N. patriciarum* (Xue *et al.*, 1992a, 1992b).

The transcriptional regulation of the *celA* gene from *C. longisporum* was investigated. CelA and other cellulases were found to be expressed at low constitutive levels during growth on cellobiose, but induction was observed during growth on barley ß-glucan.

4.3 MATERIALS AND METHODS

4.3.1 Northern blots

RNA samples were prepared as described under section 2.3.6.3. RNA preparations were mixed with RNA denaturing tracking dye containing deionized formamide and formaldehyde and incubated at 65°C for 15 min (Fourney *et al.*,1988). One µl ethidium bromide (1 mg/ml) was added to each sample, which were then run on a 0.8 % (w/v) agarose gel with TBE electrophoresis buffer. The RNA was transferred onto a Hybond-N+ charged nylon membrane (Amersham) using 0.05 M NaOH for the alkaline transfer in a capillary blot. The membrane was probed with a non-radioactive digoxigenin-labelled (DIG) DNA fragment according to the manufacturer's protocol (Boehringer Mannheim). The 1930-bp *Eco*RI restriction fragment of pCMC (Fig. 2.1) was gel-purified and DIG-labelled using the DIG DNA Labelling and Detection kit from Boehringer Mannheim. Probe bound to the membrane was detected with a chemiluminescent reaction utilizing AMPPD. The RNA molecular standards were obtained from Gibco BRL.

4.3.2 Western blots

CelA was purified out of *E. coli* BL21 (pCML) as described under section 3.3.3. A 0.5 ml saline aliquot containing 28 µg CelA was mixed with an equal volume of incomplete Freund's adjuvant and injected intramuscularly into a rabbit twice per week for three weeks; the first injection was intravenous without the adjuvant. The rabbit was bled once per week from the tenth day onwards for six weeks and serum samples were stored separately. Standard protocols were used for the western blots (Sambrook *et al.*, 1989). Protein samples were run on SDS-PAGE gels and electroblotted onto a nitrocellulose membrane. The membrane was probed with the CelA antiserum (1:7500 dilution) and anti-rabbit IgG alkaline phosphatase conjugate (Sigma Immuno Chemicals) was used as

secondary antibody. Bound antibodies were detected calorimetrically using NBT and X-phosphate (nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate).

4.3.3 SDS-PAGE and zymograms

SDS-PAGE and zymograms containing 0.1 % (w/v) CMC, barley ß-glucan or xylan were performed as described in section 3.3.10.

4.4 RESULTS

4.4.1 Induction of the *celA* gene in *C. longisporum*

The transcriptional induction of the *celA* gene was investigated by probing a northern blot containing various RNA samples with a *celA* probe (Fig. 4.1). *C. longisporum* was cultured on NRF medium containing either cellobiose or barley ß-glucan as carbon source. A third culture was grown on cellobiose supplemented with 1 µg/ml sophorose which was reported to act as an inducer of cellulases in *C. fimi* (Stewart & Leatherwood, 1976) and certain fungi e.g. *Trichoderma viride* (Mandels *et al.*, 1962). Once the cultures had reached late log-phase total RNA was prepared. (Late-log cultures were harvested, because there was too little cell material in mid-exponential cultures.) Total RNA was also prepared from *E. coli* BL21 (pCML), which contains the *celA* gene on a high copy number vector; this was used as a positive control. The northern blot clearly indicated that transcription of the *celA* gene was induced when ß-glucan was utilized as carbon source (Fig. 4.1, lane 2), but not when cellobiose was provided as carbon source or by the presence of sophorose (lanes 4 and 3, respectively). The faint bands visible in lane 1 could be due to non-specific hybridization of the probe to certain RNAs present in the standards. Two bands are visible in lane 2, a dark band and a more rapidly migrating fainter band, which were found to represent mRNA transcripts with lengths of 1650 and 1370 bases, respectively. These bands differ in size by about 280 (+/-25) bases. The upper band represents full-length *celA* mRNA transcripts, while the lower band represents truncated *celA* transcripts, which are speculated to have arisen possibly from transcription termination at a stem-loop terminator located at the 5' end of the CBD encoding region (refer to section 2.4.6).

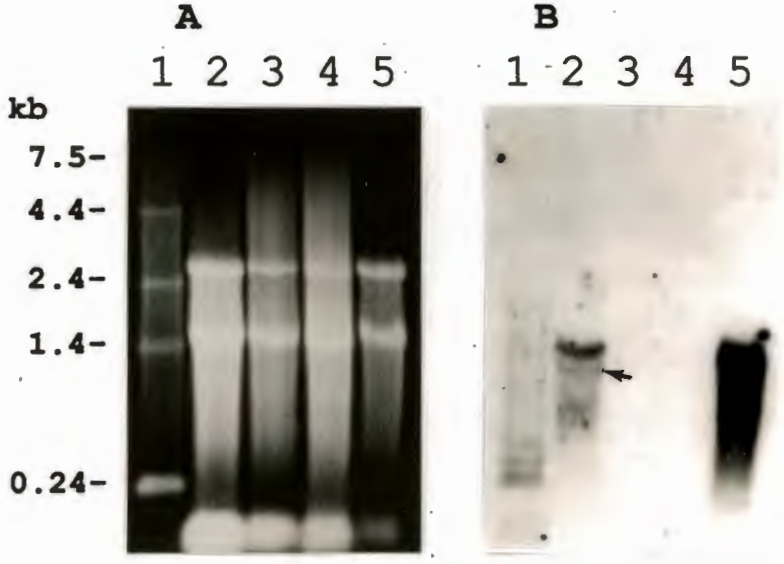


Fig. 4.1 Northern blot showing transcriptional induction of *celA*. The gel photo is shown on the left **(A)** and the autoradiograph of the blotted membrane is on the right **(B)**. Lane 1 contained RNA size markers; lanes 2, 3 and 4 show *C. longisporum* RNA from cultures grown on ß-glucan, cellobiose with sophorose and cellobiose, respectively. Lane 5 contained *E. coli* BL21 (pCML) RNA. Lanes 2 to 5 contained 37, 20, 20 and 11 µg of RNA, respectively. The arrow indicates the faint band which is formed due to transcription termination within the *celA* gene.

4.4.2 Localization of CelA in *C. longisporum*

The western blot (Fig. 4.2 B) shows that CelA was predominantly present in the culture supernatant in *C. longisporum* This result was predicted by the presence of a signal sequence at the 5' end of the *celA* gene (sections 2.4.3 and 3.4.1). The positive control in lane 1 contains purified CelA protein and shows two bands due to proteolytic cleavage of the full-length CelA protein in the linker region separating the catalytic domain and the CBD (section 3.4.2). The 55-kDa band represents the full-length CelA, while the 44-kDa band represents the CelAΔCBD i.e. the catalytic domain without the CBD. These molecular weights differ slightly from the values reported in section 3.4.2 (57 and 47 kDa, respectively), but this can be ascribed to differences in experimental conditions. The lower band in lane 2 (Fig. 4.2 B) is smaller than the 44-kDa band in lane 1. There are three possible explanations for this band. It could be a proteolytic degradation derivative of CelA that is processed differently in *C. longisporum* than in *E. coli*, or it could be another protein antigenically related to CelA. The protein could also have been produced by the truncated *celA* mRNA transcript encoding only the catalytic domain of CelA (see sections 2.4.6 and 4.4.1), but the expected molecular mass of such a protein would have been 47.8 kDa, i.e. the band would have migrated more slowly. All three possibilities are in keeping with the observation that the CelA antiserum seems to recognize the entire CelA protein more readily than the catalytic domain on its own, since the 55-kDa band (full-length CelA) in lane 1 stained more strongly than the 44-kDa band (CelAΔCBD).

4.4.3 Induction of CelA as seen on the western blot and zymogram

Both Figs. 4.2 B and C show that the quantity of CelA and other CMCases in the induced *C. longisporum* culture was considerably higher than in the uninduced culture. It is however important to notice that significant quantities of CelA and other CMCases were present in the culture supernatant of the uninduced *C. longisporum* culture. This is in contrast to the transcriptional induction study where no *celA* mRNA was detected in the uninduced culture on the northern blot (Fig. 4.1). These data indicate that the low levels of constitutively expressed mRNA were not detectable.

By comparing lane 2 with lane 4 in Figs. 4.2 A and C two observations can be made. The proteins present in the supernatant of the induced and uninduced cultures differ and two bands with molecular weights of about 80 to 90 kDa show a marked induction (Fig. 4.2 A). A conspicuous band of activity of ca. 35 kDa which shows no induction is visible on the zymogram (Fig. 4.2 C). A combination of inducible and constitutively expressed cellulases has been reported in other rumen bacteria, namely *R. flavefaciens* and *F. succinogenes* (Huang & Forsberg, 1990; Doerner *et al.*, 1992), and in *C. fimi* (Greenberg *et al.*, 1987a and 1987b; Moser *et al.*, 1989) and *C. thermocellum* (Mishra *et al.*, 1991).

4.4.4 Expression of other cellulases and xylanases in *C. longisporum*

As mentioned in section 2.4.2 38 CMCase-positive clones were isolated by screening the genomic library. Cell free extracts (CFEs) of *E. coli* cultures containing these plasmids were analyzed on CMC zymograms and 12 clones were found to express CMCases with identical molecular masses, termed CelB (results not shown). By running purified CelA in lanes 1 and 2 of the ß-glucan-containing zymogram (Fig. 4.3) and the CFE of *E. coli* K514 (pCL12) containing CelB in lanes 6 and 7 it was possible to resolve the complex set of activity bands present in the supernatant of the *C. longisporum* culture (lanes 3-5). Four to five distinct species of ß-glucanases were identified and termed CelA to CelE. By comparing Figures 4.2 C and 4.3 it becomes evident that CelB is an inducible ß-glucanase while CelC seems to be produced constitutively. Further work on the regulation and characterization of these major species of ß-glucanases is in progress.

The xylan zymogram in Fig. 4.4 can be "reconciled" with the ß-glucan activity gel in Fig. 4.3 at the position of the CelA activity bands. CelA has previously been shown to have xylanase activity (section 3.4.5). CelC and the putative CelE seem to be similar to CelA in that they show activity on both ß-glucan and xylan. It is evident from Figs. 4.2 B and 4.4 that most of the endoglucanase and xylanase activity is found in the extracellular fluid. Two truncated CelA activity bands are visible in Figs. 4.3 and 4.4. which is in contrast to the CMC-zymograms in Fig. 4.2 C and Fig. 3.2 A, where only one degradation product of CelA is visible i.e. CelAΔCBD. This partial degradation of CelA could be attributed to the age of the specific CelA protein sample which had been stored at 4°C as a diluted solution for longer than 2 weeks.

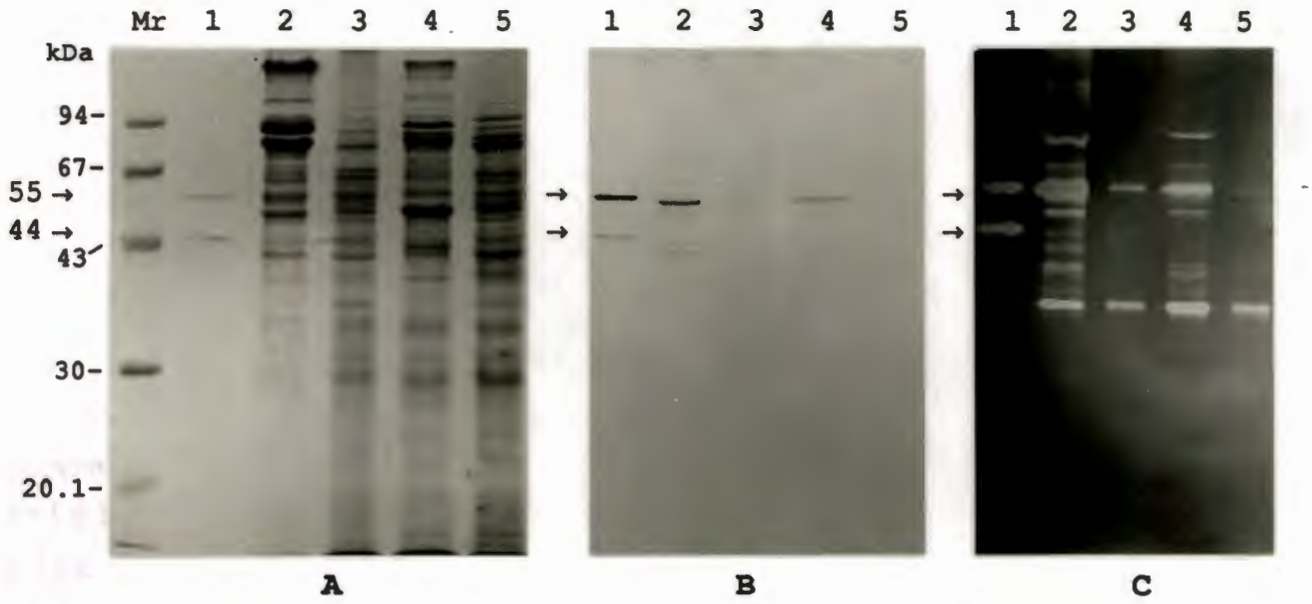


Fig. 4.2 Induction and localization of CelA in *C. longisporum* as shown on (A) SDS-PAGE, (B) Western blot and (C) CMC-containing zymogram. The SDS-PAGE gel shows the molecular weight markers. CelA purified from *E. coli* was used as positive control in lane 1. The 55- and 44-kDa bands indicated with the arrows show the full-length CelA and the proteolytically truncated CelAΔCBD, respectively. Lane 2 contains the extracellular proteins, which were concentrated by ultrafiltration, and lane 3 the cell-associated proteins of a *C. longisporum* culture grown on ß-glucan, while the extracellular proteins and the cell-associated proteins shown in lanes 4 and 5, respectively, were obtained from an uninduced culture grown on cellobiose. 2.3, 0.7 and 0.12 µg of purified CelA were loaded in lane 1 of A, B and C, respectively. 50 µg of protein was loaded in lanes 2 to 5 of A and B, and 5 µg in C, respectively.

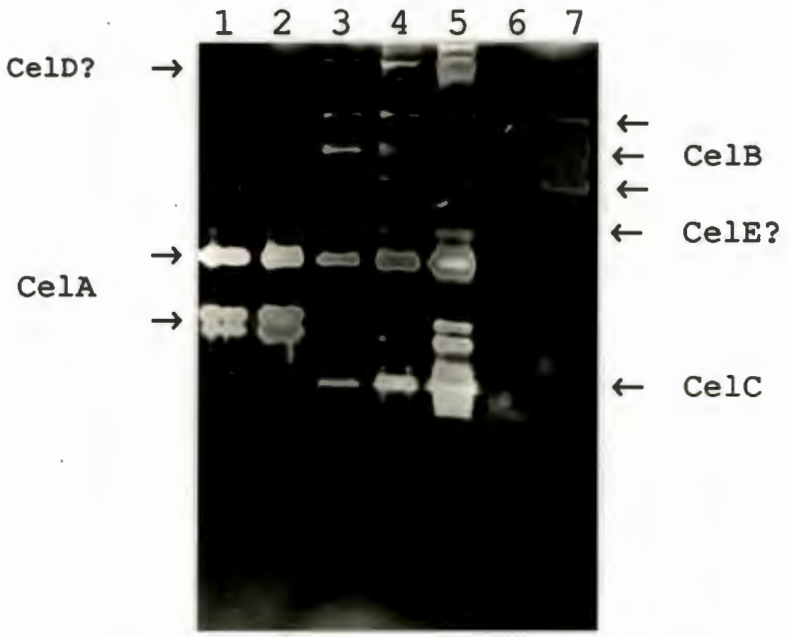


Fig. 4.3 ß-Glucan zymogram. The bands represent ß-glucanases present in the culture supernatant of a culture grown on ß-glucan as carbon source. Lanes 1 and 2 contain 0.05 and 0.11 µg purified CelA as positive control, respectively; 3, 4 and 5 contain 0.84, 2.1 and 6.3 µg extracellular proteins from *C. longisporum*; 6 and 7 contain 1 and 5 µl cell free extract of *E. coli* (pCL12) expressing CelB. The major endoglucanases CelA to CelE are indicated.

Fig. 4.4 Xylan-containing zymogram. Extracellular and cell-associated xylanase activities of a *C. longisporum* culture grown on ß-glucan. Lane 1 contains the positive control, 0.57 µg of purified CelA; lanes 2, 3 and 4 contain 4.2, 10.5 and 17 µg extracellular protein; lanes 5, 6 and 7 contain 7.2, 18 and 29 µg cell-associated protein.

4.5 DISCUSSION

4.5.1 Induction of *celA*

The generally accepted model of the regulation of cellulase genes involves the induction of transcription by enzymatic degradation products of the large extracellular substrate, e.g. cellodextrins or derivatives thereof. Some low constitutive expression of extracellularly located cellulases would be required to produce these for uptake into the cell (Béguin & Aubert, 1994). The transcriptional induction of the *celA* gene when *C. longisporum* was cultured on ß-glucan medium is not unexpected since CelA was found to be an endo-ß-glucanase with a high substrate specificity for barley ß-glucan (section 3.4.5). This induction was reflected by the increase in the activity of CelA as well as other cellulases on zymograms (Fig. 4.2 C) and by the visible increase in the amount of CelA detected on the western blot (Fig. 4.2 B). Significant amounts of these proteins were however present in the culture supernatant of uninduced cells, which indicates a low constitutive expression of these genes during growth on cellobiose. The quantitative increase in the total ß-glucanase activity in the *C. longisporum* supernatant, when grown under inducing conditions, has not been determined and these experiments still have to be done to conclude the induction study. In addition it is possible that more stringent catabolite repression of these cellulase genes would be observed during growth on carbon sources such as glucose or sucrose (see section 2.4.1.3). Transcriptional induction of *celA*, *celB* and *celC* from the ruminal fungus *N. patriciarum* by cellulose has been reported, but small quantities of mRNA from all three genes could still be detected during growth on glucose (Xue *et al.*, 1992b), unlike *C. longisporum celA* mRNA which could not be detected during growth on cellobiose. It is possible that the non-radioactive probe-detection protocol used in the northern blot was not sensitive enough to detect very low concentrations of *celA* mRNA. Transcription of *celD* from *N. patriciarum*, which encodes three different cellulase activities on a multicistronic transcript, was constitutive (Xue *et al.*, 1992a).

Very few studies on cellulase gene regulation have measured the transcriptional induction of individual genes and compared this with the induction of total cellulase activity. While individual cellulases may be regulated at the transcriptional level, the response of the total activity, being the sum of the various individual activities, would reflect a variety of different types of regulation. One example is the case of *R. flavefaciens* cellulase regulation,

where enzymatic data suggest that cellulase activity (specific CMCase and ^{14}C–cellulase activities) is higher during growth on cellobiose than on cellulose (Doerner *et al.*, 1992). Northern blots, however, clearly indicated that two cellulase genes (out of four investigated) were induced by cellulose. It is possible that some of the constitutively expressed activities were so high that the induction of other cellulases by cellulose could not be detected. In another rumen bacterium, *Fibrobacter succinogenes* subsp. *succinogenes* S85, total culture endoglucanase specific activity was only 2.1 to 2.2–fold greater in cellulose-grown cultures than in cellobiose- or glucose-grown cultures (McGavin *et al.*, 1990) and cellobiosidase activity (cellodextrinases and Cl-stimulated cellobiosidase) was found to be produced constitutively (Huang & Forsberg, 1990). In three different strains of *C. thermocellum* almost as much cellulase activity was produced during growth on cellobiose as on Avicel, and the authors found cellobiose to act as inducer of cellulase activity (Bhat *et al.*, 1993). In *B. fibrisolvens* NCFB 2249 enzymatic induction studies of xylanolytic enzymes showed up to 20–fold differences in the activity of specific enzymes when xylan or xylose-containing saccharides were the growth substrates (Williams & Withers, 1992). In the latter case transcriptional induction of the individual genes would probably correlate directly with the measurements of (total) enzyme activity, which is not the case with the examples mentioned previously.

The molecular mechanisms which cause the induction and repression of cellulase genes in cellulolytic microorganisms are unknown. Regulatory elements have been identified upstream of some genes (see section 2.5.4) and the phosphoenolpyruvate-dependent phosphotransferase system (PTS) has been implicated in some cases. The analysis of an operon encoding PTS genes from *C. longisporum*, obtained from the genomic library described in section 2.4.2, will hopefully contribute to our understanding of the molecular mechanisms of cellulase regulation. It will be interesting to see if we can find equivalents of central *pts* genes which have been found in other bacteria (refer to section 1.9). It is possible that the direct repeats upstream of *celA* are involved in binding the regulatory proteins. The sequencing and transcriptional regulation analysis of two more cellulase genes from *C. longisporum*, *celB* and *celC*, which were obtained from the genomic library, will hopefully provide more insights.

4.5.2 Hemicellulase activity of *C. longisporum*

C. longisporum was found to produce several xylanases (section 4.4.4) and CelA had significant xylanase activity (section 3.4.5). This is in agreement with reports that *C. longisporum* degraded the hemicellulose from alfalfa cell walls more extensively than various other rumen bacteria, even though it was not able to grow on larchwood xylan, xylose or arabinose (Varel, 1989). A comparison of the cellulolytic activities of *C. longisporum* and *R. albus* SY3 showed that the latter degraded alfalfa and barley straw more extensively (Varel *et al.*, 1989). In the present study it was found that *C. longisporum* was not able to grow on Avicel or ball-milled filter paper (section 2.4.1.3). It is therefore speculated that the sporadic occurrence of these clostridia and their relatively low numbers in the rumen is due to the lack, or the insufficient production, of essential cellulases. This feature could possibly be exploited in the genetic modification of its cellulolytic activities.

4.5.3 Work forthcoming from this project

Several additional features of *C. longisporum* warrant further investigation. Its lack of ability to grow on crystalline cellulose could act as a model system where the introduction of certain other cellulase genes, e.g. cellobiohydrolases, could be used to study synergistic enzyme activity. A genetic transfer system will be required to introduce such genes and various options are under investigation. An advantage of *C. longisporum* is that it grows as easily identifiable colonies on plates, in contrast to *R. flavefaciens* which only grows in liquid medium. The question whether *C. longisporum* synthesizes a cellulosome-like complex should also be addressed. Hopefully these studies will help to define the role of *C. longisporum* in ruminal cellulose degradation.

REFERENCES

Aiba, H., S. Adhya, and B. de Crombrugghe. 1981. Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. J. Biol. Chem. 256:11905-11910.

Aminov, R.I., K. Kaneichi, T. Miyagi, and K. Ohmiya. 1993. Conjugal transfer of plasmid pAMß1 to *Ruminococcus albus*, p. 179-187. In S. Hoshino, S. Karita, Y. Kobayashi, K. Ohmiya, K. Sakka, and K. Shimada (eds.), Genetics, biochemistry and ecology of lignocellulose degradation, Proceedings of MIE Bioforum 93. Uni Publishers Co., Ltd, Tokyo, Japan.

Anderson, M., S. Ratnayake, L. Kenne, L. Ericsson, and R.J. Stack. 1993. Structural studies of the extracellular polysaccharide from *Butyrivibrio fibrisolvens* strain X6C61. Carbohydrate Research 246:291-301.

Arase, A., T. Yomo, I. Urabe, Y. Hata, Y. Katsube, and H. Okada. 1993. Stabilization of xylanase by random mutagenesis. FEBS Lett. 316:123-127.

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.

Bayer, E.A., E. Morag, and R. Lamed. 1994. The cellulosome - a treasure-trove for biotechnology. Trends in Biotech. in press.

Barichievich, E.M., and R.E. Calza. 1990. Supernatant protein and cellulase activities of the anaerobic ruminal fungus *Neocallimastix frontalis* EB188. Appl. Environ. Microbiol. 56:43-48.

Barros, M.E.C., and J.A. Thomson. 1987. Cloning and expression in *Escherichia coli* of a cellulase gene from *Ruminococcus flavefaciens*. J. Bacteriol. 169:1760-1762.

Béchet, M., P. Pheulpin, H.J. Flint, J. Martin, and H.-C. Dubourguier. 1993. Transfer of hybrid plasmids based on the replicon pRR17 from *Escherichia coli* to *Bacteroides* and *Prevotella* strains. J. Appl. Bacteriol. 74:542-548.

Bedarkar, S., N.R. Gilkes, D.G. Kilburn, E. Kwan, D.E. Rose, R.C. Miller, Jr., R.A.J. Warren, and S.G. Withers. 1992. Crystallization and preliminary X-ray diffraction analysis of the catalytic domain of Cex, an exo-ß-1,4-glucanase and ß-1,4-xylanase from the bacterium *Cellulomonas fimi*. J. Mol. Biol. 228:693-695.

Béguin, P. 1983. Detection of cellulase activity in polyacrylamide gels using congo red-stained agar replicas. Analytical Biochemistry 131, 333-336.

Béguin, P., M. Rocancourt, M.-C. Chebrou, and J.-P. Aubert. 1986. Mapping of mRNA encoding endoglucanase A from *Clostridium thermocellum*. Mol. Gen. Genet. 202:251-254.

Béguin, P., J. Millet, O. Grapinet, A. Navarro, M. Juy, A. Amit, R. Poljak, and J.-P. Aubert. 1988. The *cel* (cellulose degradation) genes of *Clostridium thermocellum*, p. 267-282. In J.P. Aubert, P. Béguin and J. Millet (eds.), Biochemistry and genetics of cellulose degradation, FEMS symposium No. 43. Academic Press Inc., San Diego, CA 92101.

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

Béguin, P., J. Millet, S. Chauvaux, S. Salamitou, K. Tokatlidis, J. Navas, T. Fujino, M. Lemaire, O. Raynaud, M.-K. Daniel, and J.-P. Aubert. 1992. Bacterial cellulases. Biochemical Society Transactions 20, 42-46.

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

Berger, E., W.A. Jones, D.T. Jones, and D.R. Woods. 1989. Cloning and sequencing of an endoglucanase (*end1*) gene from *Butyrivibrio fibrisolvens* H17c. Mol. Gen. Genet. 219:193-198.

Berger, E., W.A. Jones, D.T. Jones, and D.R. Woods. 1990. Sequencing and expression of a cellodextrinase (*ced1*) gene from *Butyrivibrio fibrisolvens* H17c cloned in *Escherichia coli*. Mol. Gen. Genet. 223:310-318.

Bernalier, A., G. Fonty, F. Bonnemoy, and P. Gouet. 1993. Inhibition of the cellulolytic activity of *Neocallimastix frontalis* by *Ruminococcus flavefaciens*. J. Gen. Microbiol. 139:873-880.

Bhat, S., P.W. Goodenough, E. Owen, and M.K. Bhat. 1993. Cellobiose: A true inducer of cellulosome in different strains of *Clostridium thermocellum*. FEMS Microbiol. Lett. 111:73-78.

Brandt, W.F., H. Alk, M. Chauhan and C. Von Holt. 1984. A simple modification converts the spinning cup sequencer into a vapour phase sequencer. FEBS Letters 174, 228-232.

Brickman, E. and J. Beckwith. 1975. Analysis of the regulation of *E. coli* alkaline phosphatase synthesis using deletions and pho80 transducing phages. J. Mol. Biol. 96, 307-316.

Brown, G.D., T. Jorgensen, E.J. Morris, and J.A. Thomson. 1993. Analysis of a cellodextrinase cloned from *Ruminococcus flavefaciens* FD-1. FEMS Microbiol. Lett. 111:57-62.

Brown, G.D. 1994. Microbiology Department, University of Cape Town. Personal communications.

Caldwell, D.R., and M.P. Bryant. 1966. Medium without rumen fluid for nonselective enumeration and isolation of rumen bacteria. Appl. Microbiol. 14:794-801.

Chauvaux, S., P. Béguin, and J.-P. Aubert. 1992. Site-directed mutagenesis of essential carboxylic residues in *Clostridium thermocellum* endoglucnasase CelD. J. Biol. Chem. 267:4472-4478.

Chesson, A., and C.W. Forsberg. 1988. Polysaccharide degradation by rumen microorganisms, p. 251-284. In P.N. Hobson (ed.), The rumen microbial ecosystem. Elsevier Science Publishers, Crown House, Linton Road, Barking, Essex IG11 8JU, England.

Chow, J.M., and J.B. Russel. 1992. Effect of pH and monensin on glucose transport by *Fibrobacter succinogenes*, a cellulolytic ruminal bacterium. Appl. Environ. Microbiol. 58:1115-1120.

Chung, C.H., and A.L. Goldberg. 1981. The product of the *lon* (*capR*) gene in *Escherichia coli* is the ATP-dependent protease, protease La. Proc. Natl. Acad. Sci. USA 78:4931-4935.

Cocconcelli, P.S., E. Ferrari, F. Rossi, and V. Bottazzi. 1992. Plasmid transformation of *Ruminococcus albus* by means of high-voltage electroporation. FEMS Microbiol. Lett. 94:203-208.

Cotta, M.A. 1992. Interaction of ruminal bacteria in the production and utilization of maltooligosaccharides from starch. Appl. Environ. Microbiol. 58:48-54.

Coughlan, M.P., and L.G. Ljungdahl. 1988. Comparative biochemistry of fungal and bacterial cellulolytic enzyme systems, p. 11-30. In J.P. Aubert, P. Béguin and J. Millet (eds.), Biochemistry and genetics of cellulose degradation, FEMS symposium No. 43. Academic Press Inc., San Diego, CA 92101.

Coutinho, J.B., N.R. Gilkes, R.A.J. Warren, D.G. Kilburn, and R.C. Miller, Jr. 1992. The binding of *Cellulomonas fimi* endoglucanase C (CenC) to cellulose and Sephadex is mediated by the N-terminal repeats. Molecular Microbiol. 6:1243-1252.

Covert, S.F., A. Vanden Wymelenberg, and D. Cullen. 1992. Structure, organization, and transcription of a cellobiohydrolase gene cluster from *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. 58:2168-2175.

Demain, A.L., and L.R. Lynd. 1993. Turning garbage into motor fuel: fanciful dream or feasible scheme?, p. 573-583. In S. Hoshino, S. Karita, Y. Kobayashi, K. Ohmiya, K. Sakka, and K. Shimada (eds.), Genetics, biochemistry and ecology of lignocellulose degradation, Proceedings of MIE Bioforum 93. Uni Publishers Co., Ltd, Tokyo, Japan.

Davies, G., S. Tolley, K. Wilson, M. Schülein, H.F. Wöldike, and G. Dodson. 1992. Crystallization and preliminary X-ray analysis of a fungal endoglucanase I. J. Mol. Biol. 228:970-972.

Davies, G.J., G.G. Dodson, R.E. Hubbard, S.P. Tolley, Z. Dauter, K.S. Wilson, C. Hjort, J.M. Mikkelsen, G. Rasmussen, and M. Schülein. 1993. Structure and function of endoglucanase V. Nature 365:362-364.

Deguchi, H., Y. Watanabe, T. Sasaki, T. Matsuda, S. Shimizu, and K. Ohmiya. 1991. Purification and properties of the endo-1,4-ß-glucanase from *Ruminococcus albus* and its gene products in *Escherichia coli*. J. Ferment. Bioengineering 71:221-225.

Dehority, B.A., and C.G. Orpin. 1988. Development of, and natural fluctuations in, rumen microbial populations, p. 151-813. In P.N. Hobson (ed.), The rumen microbial ecosystem. Elsevier Science Publishers, Crown House, Linton Road, Barking, Essex IG11 8JU, England.

Deutscher, J., J. Reizer, C. Fischer, A. Galinier, M.H. Saier, Jr., and M. Steinmetz. 1994. Loss of protein kinase-catalyzed phosphorylation of HPr, a phosphocarrier protein of the phosphotransferase system, by mutation of the *ptsH* gene confers catabolite repression resistance to several catabolite genes in *Bacillus subtilis*. J. Bacteriol. 176:3336-3344.

Divne, C., J. Ståhlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J.C. Knowles, T.T. Teeri, and T.A. Jones. 1994. The three-dimensionsal crystal structure of the catlalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science 265:524-528.

Din, N., N.R. Gilkes, B. Tekant, R.C. Miller, Jr., R.A.J. Warren, and D.G. Kilburn. 1991. Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. Bio/Technology 9:1096-1099.

Din, N., I.J. Forsythe, L.D. Burtnick, N.R. Gilkes, R.C. Miller, Jr., R.A.J. Warren, and D.G. Kilburn. 1994. The cellulose-binding domain of endoglucanase A (CenA) from *Cellulomonas fimi*: evidence for the involvement of tryptophan residues in binding. Molecular Microbiol. 11:747-755.

Doerner, K.C., and B.A. White. 1990. Assessment of the endo-1,4-ß-glucanase components of *Ruminococcus flavefaciens* FD-1. Appl. Environ. Microbiol. 56:1844-1850.

Doerner, K.C., G.T. Howard, R.I. Mackie, and B.A. White. 1992. ß-Glucanase expression by *Ruminococcus flavefaciens* FD-1. FEMS Microbiol. Lett. 93:147-154.

Durrant, A.J., J. Hall, G.P. Hazlewood, and H.J. Gilbert. 1991. The non-catalytic C-terminal regions of endoglucanase E from *Clostridium thermocellum* contains a cellulose-binding domain. Biochemical J. 273, 289-293.

Erfle, J.D., and R.M. Teather. 1991. Isolation and properties of a (1,3)-ß-D-glucanase from *Ruminococcus flavefaciens*. Appl. Environ. Microbiol. 57:122-129.

Felix, C.R., and L.G. Ljungdahl. 1993. The cellulosome: The exocellular organelle of *Clostridium*. Annu. Rev. Microbiol. 47:791-819.

Fernández-Abalos, J.M., P. Sánchez, P.M. Coll, J.R. Villanueva, P. Pérez, and R.I. Santamarìa. 1992. Cloning and nucleotide sequence of *celA1*, an endo-ß-1,4-glucanase-encoding gene from *Streptomyces halstedii* JM8. J. Bacteriol. 174:6368-6376.

Ferreira, L.M.A., A.J. Durrant, J. Hall, G.P. Hazlewood, and H.J. Gilbert. 1990. Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase. Biochemical J. 269, 261-264.

Ferreira, L.M.A., G.P. Hazlewood, P.J. Barker, and H.J. Gilbert. 1991. The cellodextrinase from *Pseudomonas flourescens* subsp. *cellulosa* consists of multiple functional domains. Biochem. J. 279:793-799.

Fierobe, H.-P., C. Gaudin, A. Belaich, M. Loutfi, E. Faure, C. Bagnara, D. Baty, and J.-P. Belaich. 1991. Characterization of endoglucanase A from *Clostridium cellulolyticum*. J. Bacteriol. 173:7956-7962.

Fierobe, H.-P., C. Bagnara-Tardif, C. Gaudin, F. Guerlesquin, P. Sauve, A. Belaich, and J.-P. Belaich. 1993. Purification and characterization of endoglucanase C from *Clostridium cellulolyticum*. Eur. J. Biochem. 217:557-565.

Flint, H.J., A.M. Thomson, and J. Bisset. 1988. Plasmid-associated transfer of tetracycline resistance in *Bacteroides ruminicola*. Appl. Environ. Microbiol. 54:855-860.

Flint, H.J., C.A. McPherson, and J. Martin. 1991. Expression of two xylanase genes from the rumen cellulolytic bacterium *Ruminococcus flavefaciens* 17 cloned in pUC13. J. Gen. Microbiol. 137:123-129.

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

Flores, D.A. 1991. Biotechnology and the improvement of silage (tropical and temperate) rumen digestion: a mini-review. Appl. Microbiol. Biotechnol. 35:277-282.

Fourney, R.M., J. Miyakoshi, R.S. Day III, and M.C. Paterson. 1988. Northern blotting: efficient RNA staining and transfer. Focus 10:5-6.

Fujino, T., P. Béguin, and J.-P. Aubert. 1992. Cloning of a *Clostridium thermocellum* DNA fragment encoding polypeptides that bind the catalytic components of the cellulosome. FEMS Microbiol. Lett. 94:165-170.

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.

Gardner, R.M., K.C. Doerner, and B.A. White. 1987. Purification and characterization of an exo-ß-1,4-glucanase from *Ruminococcus flavefaciens* FD-1. J. Bacteriol. 169:4581-4588.

Genetics Computer Group. 1991. Program Manual for the GCG Package, Version 7, April 1991. 575 Science Drive, Madison, Wisconsin, USA 53711.

Gerngross, U.T., M.P.M. Romaniec, T. Kobayashi, N.S. Huskisson, and A.L. Demain. 1993. Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal S_L-protein reveals an unusual degree of internal homology. Molecular Microbiol. 8:325-334.

Ghose, T.K. 1987. Measurement of cellulase activities. Pure & Applied Chem. 59:257-268.

Gibbs, M.D., D.J. Saul, E. Lüthi, and P.L. Bergquist. 1992. The ß-mannanase from "*Caldocellum saccharolyticum*" is part of a multidomain enzyme. Appl. Environ. Microbiol. 58:3864-3867.

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. Molecular 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., R.A.J. Warren, R.C. Miller, Jr., and D.G. Kilburn. 1988. Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem. 263:10401-10407.

Gilkes, N.R., D.G. Kilburn, R.C. Miller, Jr., and R.A.J. Warren. 1989. Structural and functional analysis of a bacterial cellulase by proteolysis. J. Biol. Chem. 264:17802-17808.

Gilkes, N.R., B. Henrissat, D.G. Kilburn, R.C. Miller, Jr., and R.A.J. Warren. 1991a. Domains in microbial ß-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. 55:303-315.

Gilkes, N.R., M. Claeyssens, R. Aebersold, B. Henrissat, A. Meinke, H.D. Morrison, D.G. Kilburn, R.A.J. Warren, and R.C. Miller, Jr. 1991b. Structural and functional relationships in two families of ß-1,4-glycanases. Eur. J. Biochem. 202:367-377.

Goldstein, M.A., M. Takagi, S. Hashida, O. Shoseyov, R.H. Doi, and I.H. Segel. 1993. Characterization of the cellulose-binding domain of the *Clostridium cellulovorans* cellulose-binding protein A. J. Bacteriol. 175:5762-5768.

Gonsalbes, M.J., J.A. Pérez-González, R. González, and A. Navarro. 1991. Two ß-glycanase genes are clustered in *Bacillus polymyxa*: Molecular cloning, expression, and sequence analysis of genes encoding a xylanase and an endo-ß-(1,3)-(1,4)-glucanase. J. Bacteriol. 173:7705-7710.

Gräbnitz, F., M. Seiss, K.P. Rücknagel, and W.L. Staudenbauer. 1991. Structure of the ß-glucosidase gene *bglA* of *Clostridium thermocellum*. Eur. J. Biochem. 200:301-309.

Greenberg, N.M., R.A.J. Warren, D.G. Kilburn, and R.C. Miller, Jr. 1987a. Regulation, initiation, and termination of the *cenA* and *cex* transcripts of *Cellulomonas fimi*. J. Bacteriol. 169:646-653.

Greenberg, N.M., R.A.J. Warren, D.G. Kilburn, and R.C. Miller, Jr. 1987b. Regulation and initiation of *cenB* transcripts of *Cellulomonas fimi*. J. Bacteriol. 169:4674-4677.

Grodberg, J. and J.J. Dunn. 1988. *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. J. Bacteriol. 170:1245-1253.

Hall, J., G.P. Hazlewood, P.J. Barker, and H.J. Gilbert. 1988. Conserved reiterated domains in *Clostridium thermocellum* endoglucanases are not essential for catalytic activity. Gene 69:29-38.

Hall, J., G.P. Hazlewood, N.S. Huskisson, A.J. Durrant, and H.J. Gilbert. 1989. Conserved serine-rich sequences in xylanases and cellulases from *Pseudomonas fluorescens* subsp. *cellulosa*: internal signal sequence and unusual protein processing. Mol. Microbiol. 3:1211-1219.

Hanke, C., J. Hess, G. Schumacher, and W. Goebel. 1992. Processing by OmpT of fusion proteins carrying the HlyA transport signal during secretion by the *Escherichia coli* hemolysin transport system. Mol. Gen. Genet. 233:42-48.

Hansen, C.K., B. Diderichsen, and P.L Jørgensen. 1992. *celA* from *Bacillus lautus* PL236 encodes a novel cellulose-binding endo-ß-1,4-glucanase. J. Bacteriol. 174:3522-3531.

Hazlewood, G.P., and R.M. Teather. 1988. The genetics of rumen bacteria, p. 323-341. In P.N. Hobson (ed.), The rumen microbial ecosystem. Elsevier Science Publishers, Crown House, Linton Road, Barking, Essex IG11 8JU, England.

Hazlewood, G.P., K. Davidson, J.H. Clarke, A.J. Durrant, J. Hall, and H.J. Gilbert. 1990. Endoglucanase E, produced at high level in *E. coli* as a lacZ' fusion protein, is part of the *Clostridium thermocellum* cellulosome. Enzyme Microbial Technology 12:656-662.

Hazlewood, G.P., K. Davidson, J.I. Laurie, M.P.M. Romaniec, and H.J. Gilbert. 1990b. Cloning and sequencing of the *celA* gene encoding endoglucanase A of *Butyrivibrio fibrisolvens* strain A46. J. Gen. Microbiol. 136:2098-2097.

Hazlewood, G.P., J.I. Laurie, L.M.A. Ferreira, and H.J. Gilbert. 1992. *Pseudomonas fluorescens* subsp. *cellulosa*: an alternative model for bacterial cellulases. J. Appl. Bacteriol. 72:244-251.

Hebraud, M., and M. Fevre. 1990. Purification and characterization of an extracellular ß-xylosidase from the rumen anaerobic fungus *Neocallimastix frontalis*. FEMS Microbiol. Lett. 72:11-16.

Hefford, M.A., R.M. Teather, and R.J. Forster. 1993. The complete nucleotide sequence of a small cryptic plasmid from a rumen bacterium of the genus *Butyrivibrio*. Plasmid 29:63-69.

Helaszek, C.T., and B.A. White. 1991. Cellobiose uptake and metabolism by *Ruminococcus flavefaciens*. Appl. Environ. Microbiol. 57:64-68.

Henderson, T.A., P.M. Dombrosky, and K.D. Young. 1994. Artifactual processing of penicillin-binding proteins 7 and 1b by the OmpT protease of *Escherichia coli*. J. Bacteriol. 176:256-259.

Henrissat, B., M. Claeyssens, P. Tomme, L. Lemesle, and, J.-P. Mornon. 1989. Cellulase families revealed by hydrophobic cluster analysis. Gene 81:83-95.

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

Henrissat, B., and A. Bairoch. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 293:781-788.

Hespell, R.B., and T.R. Whitehead. 1991a. Conjugal transfer of Tn916, Tn916ΔE, and pAMß1 from *Enterococcus faecalis* to *Butyrivibrio fibrisolvens* strains. Appl. Environ. Microbiol. 57:2703-2709.

Hespell, R.B., and T.R. Whitehead. 1991b. Introduction of Tn916 and pAMß1 into *Streptococcus bovis* JB1 by conjugation. Appl. Environ. Microbiol. 57:2710-2713.

Hespell, R.B., and P.J. O'Bryan. 1992. Purification and characterization of an α-L-arabinofuranosidase from *Butyrivibrio fibrisolvens* GS113. Appl. Environ. Microbiol. 58:1082-1088.

Hewick, R.M., M.W. Hunkapiller, L.E. Hood, and W.J. Dreyer. 1981. A gas-liquid solid phase peptide and protein sequencer. J. Biol. Chem. 256:7990-7997.

Honda, H., T. Saito, S. Iijima, and T. Kobayashi. 1988. Molecular cloning and expression of a ß-glucosidase from *Ruminococcus albus* in *Escherichia coli*. Enzyme Microb. Technol. 10:559-562.

Howard, G.T., and B.A. White. 1988. Molecular cloning and expression of cellulase genes from *Ruminococcus albus* 8 in *Escherichia coli* bacteriphage λ. Appl. Environ. Microbiol. 54:1752-1755.

Hsing, W., and E. Canale-Parola. 1992. Cellobiose chemotaxis by the cellulolytic bacterium *Cellulomonas gelida*. J. Bacteriol. 174:7996-8002.

Hu, M.C.-T., and N. Davidson. 1986. Mapping transcription start points on cloned genomic DNA with T4 DNA polymerase: a precise and convenient technique. Gene 42:21-29.

Huang, L., and C.W. Forsberg. 1987. Isolation of a cellodextrinase from *Bacteroides succinogenes*. Appl. Environ. Microbiol. 53:1034-1041.

Huang, L., and C.W. Forsberg. 1990. Cellulose digestion and cellulase regulation and distribution in *Fibrobacter succinogenes* subsp. *succinogenes* S85. Appl. Environ. Microbiol. 56:1221-1228.

Hungate, R.E. 1957. Microorganisms in the rumen of cattle fed a constant ration. Can. J. Microbiol. 3:289-311.

Hungate, R.E. 1988. Introduction: The ruminant and the rumen, p. 1-19. In P.N. Hobson (ed.), The rumen microbial ecosystem. Elsevier Science Publishers, Crown House, Linton Road, Barking, Essex IG11 8JU, England.

Jauris, S., K.P. Rücknagel, W.H. Schwarz, P. Kratzsch, K. Bronnenmeier, and W.L. Staudenbauer. 1990. Sequence analysis of the *Clostridium stercocarium celZ* gene encoding a thermoactive cellulase (Avicelase I): identification of catalytic and cellulose-binding domains. Mol. Gen. Genet. 223:258-267.

Johnson, E.A., M. Sakajoh, G. Halliwell, A. Madia, and A.L. Demain. 1982. Saccharification of complex cellulosic substrates by the cellulase system from *Clostridium thermocellum*. Appl. Environ. Microbiol. 43:1125-1132.

Jung, E.D., G. Lao, D. Irwin, B.K. Barr, A. Benjamin, and D.B. Wilson. 1993. DNA sequences and expression in *Streptomyces lividans* of an exoglucanase gene and an endoglucanase gene from *Thermomonospora fusca*. Appl. Environ. Microbiol. 59:3032-3043.

Juy, M., A.G. Amit, P.M. Alzari, R.J. Poljak, M. Claeyssens, P. Béguin, and J.-P. Aubert. 1992. Three-dimensional structure of a thermostable bacterial cellulase. Nature 357:89-91.

Katayeva, I.A., N.P. Golovchenko, N.A. Chuvilskaya, and V.K. Akimenko. 1992. *Clostridium thermocellum* ß-glucosidases A and B: Purification, properties, localization, and regulation of biosynthesis. Enzyme Microb. Technol. 14:407-412.

Kawazu, T., T. Ohta, M. Tsuji, K. Ito, M. Shibata, K. Sakka, and K. Ohmiya. 1993. Expression of a *Ruminococcus albus* cellulase gene in plant cells, p. 566-572. *In* S. Hoshino, S. Karita, Y. Kobayashi, K. Ohmiya, K. Sakka, and K. Shimada (eds.), Genetics, biochemistry and ecology of lignocellulose degradation, Proceedings of MIE Bioforum 93. Uni Publishers Co., Ltd, Tokyo, Japan.

Kim, A.Y., G.T. Attwood, S.M. Holt, B.A. White, and H.P. Blaschek. 1994. Heterologous expression of endo-ß-1,4-D-glucanase from *Clostridium cellulovorans* in *Clostridium acetobutylicum* ATCC 824 following transformation of the *engB* gene. Appl. Environ. Microbiol. 60:337-340.

Kraulis, P.J., G.M. Clore, M.J. Nilges, T.A. Jones, G. Pettersson, J. Knowles and A.M. Gronenborn. 1989. Determination of the three-dimensional structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. Biochemistry 28:7241-7257.

Krüger, S., J. Stülke, and M. Hecker. 1993. Catabolite repression of ß-glucanase synthesis in *Bacillus subtilis*. J. Gen. Microbiol. 139:2047-2054.

Kubicek, C.P., R. Messner, F. Gruber, R.L. Mach, and E.M. Kubicek-Pranz. 1993. The *Trichoderma* cellulase regulatory puzzle: From the interior life of a secretory fungus. Enzyme Microb. Technol. 15:90-99.

Kurasawa, T., M. Yachi, M. Suto, Y. Kamagata, S. Takao, and F. Tomita. 1992. Induction of cellulase by gentiobiose and its sulfur-containing analog in *Penicillium purpurogenum*. Appl. Environ. Microbiol. 58:106-110.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685.

Langsford, M.L., N.R. Gilkes, B. Singh, B. Moser, R.C. Miller, Jr., R.A.J. Warren, and D.G. Kilburn. 1987. Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. FEBS Lett. 225:163-167.

Lao, G., G.S. Ghangas, E.D. Jung, and D.B. Wilson. 1991. DNA sequences of three ß-1,4-glucanase genes from *Thermomonospora fusca*. J. Bacteriol. 173:3397-3407.

Lauková, A., and M. Mareková. 1993. Antimicrobial spectrum of bacteriocin-like substances produced by rumen staphylococci. Folia Microbiol. 38:74-76.

Lin, E., and D.B. Wilson. 1988. Identification of a *celE*-binding protein and its potential role in induction of the *celE* gene in *Thermomonospora fusca*. J. Bacteriol. 170:3843-3846.

Lin, L.-L., E. Rumbak, H. Zappe, J.A. Thomson, and D.R. Woods. 1989. Cloning, sequencing and analysis of expression of a *Butyrivibrio fibrisolvens* gene encoding a ß-glucosidase. J. Gen. Microbiol. 136:1567-1576.

Lin, L.-L., and J.A. Thomson. 1991. Cloning, sequencing and expression of a gene encoding a 73 kDa xylanase enzyme from the rumen anaerobe *Butyrivibrio fibrisolvens* H17c. Mol. Gen. Genet. 228:55-61.

Lin, L.-L., and J.A. Thomson. 1991. An analysis of the extracellular xylanases and cellulases of *Butyrivibrio fibrisolvens* H17c. FEMS Microbiol. Lett. 84:197-204.

Lindner, C., J. Stülke, and M. Hecker. 1994. Regulation of xylanolytic enzymes in *Bacillus subtilis*. Microbiology 140:753-757.

Lottspeich, F. 1985. Microscale isocratic separation of the phenylthiohydantoin amino acid derivatives. J. Chromatography 326:321-327.

Maglione, G., O. Matsushita, J.B. Russel, and D.B. Wilson. 1992. Properties of a genetically reconstructed *Prevotella ruminicola* endoglucanase. Appl. Environ. Microbiol. 58:3593-3597.

Malburg, L.M., D.C. Smith, H.E. Schellhorn, and C.W. Forsberg. 1993. *Fibrobacter succinogenes* S85 has multiple xylanase genes. J. Appl. Bacteriol. 75:564-573.

Mandels, M., F.W. Parrish, and E.T. Reese. 1962. Sophorose as an inducer of cellulase in *Trichoderma viride*. J. Bacteriol. 83:400-408.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Martin, S.A., and J.B. Russel. 1986. Phosphoenolpyruvate-dependent phosphorylation of hexoses by ruminal bacteria: evidence for the phosphotransferase transport system. Appl. Environ. Microbiol. 52:1348-1352.

Martin, S.A., and J.B. Russel. 1987. Transport and phosphorylation of disaccharides by the ruminal bacterium *Streptococcus bovis*. Appl. Environ. Microbiol. 53:2388-2393.

Martin, S.A., and J.B. Russel. 1988. Mechanisms of sugar transport in the rumen bacterium *Selenomonas ruminantium*. J. Gen. Microbiol. 134:819-827.

Matte, A., and C.W. Forsberg. 1992. Purification, characterization, and mode of action of endoxylanases 1 and 2 from *Fibrobacter succinogenes* S85. Appl. Environ. Microbiol. 58:157-168.

McGavin, M., J. Lam, and C.W. Forsberg. 1990. Regulation and distribution of *Fibrobacter succinogenes* subsp. *succinogenes* S85 endoglucanases. Appl. Environ. Microbiol. 56:1235-1244.

Meinke, A., N.R. Gilkes, D.G. Kilburn, R.C. Miller, Jr., and R.A.J. Warren. 1993. Cellulose-binding polypeptides from *Cellulomonas fimi*: endoglucanase D (CenD), a family A ß-glucanase. J. Bacteriol. 175:1910-1918.

Meinke, A., N.R. Gilkes, E. Kwan, D.G. Kilburn, R.A.J. Warren, and R.C. Miller, Jr. 1994. Cellobiohydrolase A (CbhA) from the cellulolytic bacterium *Cellulomonas fimi* is a ß-1,4-exocellobiohydrolase analogous to *Trichoderma reesei* CBH II. Molecular Microbiol. 12:413-422.

Miller, J.H. 1972. Experiments in molecular genetics, p. 352-355. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Mishra, S., P. Béguin, and J.-P. Aubert. 1991. Transcription of *Clostridium thermocellum* endoglucanase genes *celF* and *celD*. J. Bacteriol. 173:80-85.

Mittendorf, V., and J.A. Thomson. 1993. Cloning of an endo-(1-4)-ß-glucanase gene, *celA*, from the rumen bacterium *Clostridium* sp. (*'C. longisporum'*) and characterization of its product, CelA, in *Escherichia coli*. J. Gen. Microbiol. 139:3233-3242.

Miyazaki, K., T. Hino, and H. Itabashi. 1992. Effects of extracellular pH on the intracellular pH and membrane potential of cellulolytic ruminal bacteria, *Ruminococcus flavefaciens*, and *Fibrobacter succinogenes*. J. Gen. Appl. Microbiol. 38:567-573.

Moormann, M., A. Schlochtermeier, and H. Schrempf. 1993. Biochemical characterization of a protease involved in the processing of a *Streptomyces reticuli* cellulase (Avicelase). Appl. Environ. Microbiol. 59:1573-1578.

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.

Morris, J.E., and O.J. Cole. 1987. Relationship between cellulolytic activity and adhesion to cellulose in *Ruminococcus albus*. J. Gen. Microbiol. 133:1023-1032.

Morrison, M., R.I. Mackie, and A. Kistner. 1990. 3-Phenylpropanoic acid improves the affinity of *Ruminococcus albus* for cellulose in continuous culture. Appl. Environ. Microbiol. 56:3220-3222.

Morrison, M., R.I. Mackie, and A. Kistner. 1990b. Evidence that cellulolysis by an anaerobic ruminal fungus is catabolite regulated by glucose, cellobiose, and soluble starch. Appl. Environ. Microbiol. 56:3227-3229.

Morrison, M., R.I. Mackie, and B.A. White. 1992a. Partial characterization of a DNA restriction endonuclease from *Ruminococcus flavefaciens* FD-1 and its inhibition by site-specific adenine methylation. Appl. Environ. Microbiol. 58:66-69.

Morrison, M., R.I. Mackie, and B.A. White. 1992b. Partial purification and characterization of *Ral*8I, a class-IIS restriction endonuclease from *Ruminococcus albus* 8 which recognizes 5'-GGATC. Gene 111:105-108.

Moser, B., N.R. Gilkes, D.G. Kilburn, R.A.J. Warren, and R.C. Miller. 1989. Purification and characterization of endoglucanase C of *Cellulomonas fimi*, cloning of the gene, and analysis of *in vivo* transcripts of the gene. Appl. Environ. Microbiol. 55:2480-2487.

Nakamura, S., K. Wakabayashi, R. Nakai, R. Aono, and K. Horikoshi. 1993. Purification and some properties of an alkaline xylanase from alkaliphilic *Bacillus* sp. strain 41M-1. Appl. Environ. Microbiol. 59:2311-2316.

Nossal, N.G., and L.A. Heppel. 1966. The release of enzymes by osmotic shock from *E. coli* in exponential phase. J. Biol. Chem. 13:3055-3062.

Ohmiya, K., T. Kajino, A. Kato, and S. Shimizu. 1989. Structure of a *Ruminococcus albus* endo-1,4-ß-glucanase gene. J. Bacteriol. 171:6771-6775.

Ohmiya, K., H. Deguchi, and S. Shimizu. 1991. Modification of the properties of a *Ruminococcus albus* endo-1,4-ß-glucanase by gene truncation. J. Bacteriol. 173:636-641.

Ong, E., N.R. Gilkes, R.C. Miller, Jr., R.A.J. Warren, and D.G. Kilburn. 1993. The cellulose-binding domain (CBD_{Cex}) of an exoglucanase from *Cellulomonas fimi*: production in *Escherichia coli* and characterization of the polypeptide. Biotechnology and Bioengineering 42:401-409.

Ong, E., D.G. Kilburn, R.C. Miller, Jr., and R.A.J. Warren. 1994. *Streptomyces lividans* glycosylates the linker region of a ß-1,4-glycanase from *Cellulomonas fimi*. J. Bacteriol. 176:999-1008.

Orpin, C.G. 1988. Genetic approaches to the improvement of lignocellulose degradation in the rumen, p. 171-179. In J.P. Aubert, P. Béguin and J. Millet (eds.), Biochemistry and genetics of cellulose degradation, FEMS symposium No. 43. Academic Press Inc., San Diego, CA 92101.

Orpin, C.G., and K.N. Joblin. 1988. The rumen anaerobic fungi, p. 129-150. In P.N. Hobson (ed.), The rumen microbial ecosystem. Elsevier Science Publishers, Crown House, Linton Road, Barking, Essex IG11 8JU, England.

Osborne, J.M., and B.A. Dehority. 1989. Synergism in degradation and utilization of intact forage cellulose, hemicellulose, and pectin by three pure cultures of ruminal bacteria. Appl. Environ. Microbiol. 55:2247-2250.

Pérez-González, J.A., R. González, A. Querol, J. Sendra, and D. Ramón. 1993. Construction of a recombinant wine yeast strain expressing ß-(1,4)-endoglucanase and its use in microvinification processes. Appl. Environ. Microbiol. 59:2801-2806.

Pohlschröder, M., S.B. Leschine, and E. Canale-Parola. 1994. Multicomplex cellulase-xylanase system of *Clostridium papyrosolvens* C7. J. Bacteriol. 176:70-76.

Poole, D.M., G.P. Hazlewood, J.I. Laurie, P.J. Barker, and H.J. Gilbert. 1990. Nucleotide sequence of the *Ruminococcus albus* SY3 endoglucanase genes *celA* and *celB*. Mol. Gen. Genet. 223:217-223.

Poole, D.M., A.J. Durrant, G.P. Hazlewood, and H.J. Gilbert. 1991. Characterization of hybrid proteins consisting of the catalytic domains of *Clostridium* and *Ruminococcus* endoglucanases, fused to *Pseudomonas* non-catalytic cellulose-binding domains. Biochem. J. 279:787-792.

Poole, D.M., G.P. Hazlewood, N.S. Huskisson, R. Virden, and H.J. Gilbert. 1993. The role of conserved tryptophan residues in the interaction of a bacterial cellulose binding domain with its ligand. FEMS Microbiol. Letters 106:77-84.

Postma, P.W., J.W. Lengeler, and G.R. Jacobson. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiological Rev. 57:543-594.

Ramalingam, R., J.E. Blume, and H.L. Ennis. 1992. The *Dictyostelium discoidium* spore germination-specific cellulase is organized into functional domains. J. Bacteriol. 174:7834-7837.

Rasmussen, M.A. 1993. Isolation and characterization of *Selenomonas ruminantium* strains capable of 2-deoxyribose utilization. Appl. Environ. Microbiol. 59:2077-2081.

Roig, V., H.-P. Fierobe, V. Ducros, M. Czjzek, A. Belaich, C. Gaudin, J.-P. Belaich, and R. Haser. 1993. Crystallization and preliminary X-ray analysis of the catalytic domain of endoglucanase from *Clostridium cellulolyticum*. J. Mol. Biol. 233:325-327.

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 249:380-386.

Russel, J.B., and R.L. Baldwin. 1978. Substrate preferences in rumen bacteria: evidence of catabolite regulatory mechanisms. Appl. Environ. Microbiol 36:319-329.

Russel, J.B., and D.B. Dombrowski. 1980. Effect of pH on the efficiency of growth by pure cultures of rumen bacteria in continuous culture. Appl. Environ. Microbiol. 39:604-610.

Russel, J.B. 1992. Glucose toxicity and inablility of *Bacteroides ruminicola* to regulate glucose transport and utilization. Appl. Environ. Microbiol. 58:2040-2045.

Saier, M.H. Jr., and J. Reizer. 1994. MicroReview. The bacterial phosphotransferase system: new frontiers 30 years later. Mol. Microbiol. 13:(in press).

Salamitou, S., K. Tokatlidis, P. Béguin, and J.-P. Aubert. 1992. Involvement of separate domains of the cellulosomal protein S1 of *Clostridium thermocellum* in binding to cellulose and in anchoring of catalytic subunits to the cellulosome. FEBS Letters 1:89-92.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sato, T., K. Daimon, and H. Sakaguchi. 1993. Application of cellulases for textile industry, p. 602-603. *In* S. Hoshino, S. Karita, Y. Kobayashi, K. Ohmiya, K. Sakka, and K. Shimada (eds.), Genetics, biochemistry and ecology of lignocellulose degradation, Proceedings of MIE Bioforum 93. Uni Publishers Co., Ltd, Tokyo, Japan.

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. 56:3117-3124.

Sharp, R., G.P. Hazlewood, H.J. Gilbert, and A.G. O'Donnell. 1994. Unmodified and recombinant strains of *Lactobacillus plantarum* are rapidly lost from the rumen by protozoal predation. J. Appl. Bacteriol. 76:110-117.

Sharma, S., and D.K. Sandhu. 1986. An improved method for detection of isozymes of endocellulases. Indian J. Experim. Biol. 24:732-733.

Shen, H., M. Schmuck, I. Pilz, N.R. Gilkes, D.G. Kilburn, R.C. Miller, Jr., and R.A.J. Warren. 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. J. Biol. Chem. 266:11335-11340.

Shibata, D., T. Hibino, and T. Higuchi. 1993. Potential application of lignin biosynthesis genes in tree breeding, p628-631. *In* S. Hoshino, S. Karita, Y. Kobayashi, K. Ohmiya, K. Sakka, and K. Shimada (eds.), Genetics, biochemistry and ecology of lignocellulose degradation, Proceedings of MIE Bioforum 93. Uni Publishers Co., Ltd, Tokyo, Japan.

Shoemaker, N.B., K.L. Anderson, S.L. Smithson, G.-R. Wang, and A.A. Salyers. 1991. Conjugal transfer of a shuttle vector from the human colonic anaerobe *Bacteroides uniformis* to the ruminal anaerobe *Prevotella* (*Bacteroides*) *ruminicola* B_14. Appl. Environ. Microbiol. 57:2114-2120.

Shoemaker, N.B., G.-R. Wang, and A.A. Salyers. 1992. Evidence for natural transfer of a tetracycline resistance gene between bacteria from the human colon and bacteria from the bovine rumen. Appl. Environ. Microbiol. 58:1313-1320.

Shoseyov, O., M. Takagi, M.A. Goldstein, and R.H. Doi. 1992. Primary sequence analysis of *Clostridium cellulovorans* cellulose binding protein A. Proc. Nat. Acad. Sci. USA 89:3483-3487.

Sinha, R.N., and B. Ranganathan. 1983. Cellulolytic bacteria in buffalo rumen. J. Appl. Bacteriol. 54:1-6.

Simpson, H.D., U.R. Haufler, and R.M. Daniel. 1991. An extremely thermostable xylanase from the thermophilic eubacterium *Thermotoga*. Biochem. J. 277:413-417.

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

Smith, D.C., and C.W. Forsberg. 1991. α-Glucuronidase and other hemicellulase activities of *Fibrobacter succinogenes* S85 grown on crystalline cellulose or ball-milled barley straw. Appl. Environ. Microbiol. 57:3552-3557.

Spezio, M., D.B. Wilson, and P.A. Karplus. 1993. Crystal structure of the catalytic domain of a thermophilic endocellulase. Biochemistry 32:9906-9916.

Stangl, H., F. Gruber, and C.P. Kubicek. 1993. Characterization of the *Trichoderma reesei cbh2* promoter. Curr. Genet. 23:115-122.

Stewart, B.J., and J.M. Leatherwood. 1976. Derepressed synthesis of cellulase by *Cellulomonas*. J. Bacteriol. 128:609-615.

Stewart, C.S., and M.P. Bryant. 1988. The rumen bacteria, p. 21-75. In P.N. Hobson (ed.), The rumen microbial ecosystem. Elsevier Science Publishers, Crown House, Linton Road, Barking, Essex IG11 8JU, England.

Stewart, C.S., S.H. Duncan, and H.J. Flint. 1990. The properties of forms of *Ruminococcus flavefaciens* which differ in their ability to degrade cotton cellulose. FEMS Microbiol. Lett. 72:47-50.

Stewart, C.S., S.H. Duncan, A.J. Richardson, C. Backwell, and R. Begbie. 1992. The inhibition of fungal cellulolysis by cell-free preparations from ruminococci. FEMS Microbiol. Lett. 97:83-88.

Strobel, H.J. 1993. Evidence for catabolite inhibition in regulation of pentose utilization and transport in the ruminal bacterium *Selenomonas ruminantium*. Appl. Environ. Microbiol. 59:40-46.

Strobel, H.J. 1993b. Pentose utilization and transport by the ruminal bacterium *Prevotella ruminicola*. Arch. Micobiol. 159:465-471.

Studier, F.W., and B.A. Moffat. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189:113-130.

Szakmary, K., A. Wotawa, and K.P. Kubicek. 1991. Origin of oxidized cellulose degradation products and mechanism of their promotion of cellobiohydrolase I biosynthesis in *Trichoderma reesei*. J. Gen. Microbiol. 137:2873-2878.

Teather, R.M., and P.J. Wood. 1982. Use of congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl. Environ. Microbiol. 43:777-780.

Thomson, A.M., and H.J. Flint. 1989. Electroporation induced transformation of *Bacteroides ruminicola* and *Bacteroides uniformis* by plasmid DNA. FEMS Microbiol. Lett. 61:101-104.

Thomson, J.A. 1993. Molecular biology of xylan degradation. FEMS Microbiol. Rev. 104:65-82.

Thurston, B., K.A. Dawson, and H.J. Strobel. 1993. Cellobiose versus glucose utilization by the ruminal bacterium *Ruminococcus albus*. Appl. Environ. Microbiol. 59:2631-2637.

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 Letters 291:185-188.

Tomme, P., H. Van Tilbeurgh, G. Pettersson, J. Van Damme, J. Vandekerckhove, J. Knowles, T. Teeri, and M. Claeyssens. 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Eur. J. Biochem. 170:575-581.

Törrönen, A., A. Harkki, and J. Rouvinen. 1994. Three-dimensional structure of endo-1,4-ß-xylanase II from *Trichoderma reesei*: two conformational states in the active site. EMBO J. 13:2493-2501.

Van Gylswyk, N.O., K. Wejdemar, and K. Kulander. 1992. Comparative growth rates of various rumen bacteria in clarified rumen fluid from cows and sheep fed different diets. Appl. Environ. Microbiol. 58:99-105.

Van Nevel, C.J., and D.I. Demeyer. 1988. Manipulation of rumen fermentation, p. 387-443. In P.N. Hobson (ed.), The rumen microbial ecosystem. Elsevier Science Publishers, Crown House, Linton Road, Barking, Essex IG11 8JU, England.

Varel, V.H. 1989. Reisolation and characterization of *Clostridium longisporum*, a ruminal sporeforming cellulolytic anaerobe. Arch. Microbiol. 152:209-214.

Varel, V.H., A.J. Richardson, and C.S. Stewart. 1989. Degradation of barley straw, ryegrass, and alfalfa cell walls by *Clostridium longisporum* and *Ruminococcus albus*. Appl. Environ. Microbiol. 55:3080-3094.

Varel, V.H., and W.G. Pond. 1992. Characteristics of a new cellulolytic *Clostridium* sp. isolated from pig intestinal tract. Appl. Environ. Microbiol. 58:1645-1649.

Vercoe, P.E., and K. Gregg. 1992. DNA sequence and transcription of an endoglucanase gene from *Prevotella (Bacteroides) ruminicola* AR20. Mol. Gen. Genet. 233:284-292.

Wachenheim, D.E., and J.A. Patterson. 1992. Anaerobic production of extracellular polysaccharide by *Butyrivibrio fibrisolvens* nyx. Appl. Environ. Microbiol. 58:385-391.

Wallace, R.J. 1992. Rumen microbiology, biotechnology and ruminant nutrition: The application of research findings to a complex microbial ecosystem. FEMS Microbiol. Lett. 100:529-534.

Walsh, G.A., R.F. Power, and D.R. Headon. 1993. Enzymes in the animal-feed industry. TIBTECH 11:424-430.

Wang, W., S.J. Reid, and J.A. Thomson. 1993. Transcriptional regulation of an endoglucanase and a cellodextrinase gene in *Ruminococcus flavefaciens* FD-1. J. Gen. Microbiol. 139:1219-1226.

Wang, W.K., K. Kruus, and J.H.D. Wu. 1993a. 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, Q., D. Tull, A. Meinke, N.R. Gilkes, R.A.J. Warren, R. Aebersold, and S.G. Withers. 1993b. Glu^{280} is the nucleophile in the active site of *Clostridium thermocellum* CelC, a family A endo-ß-1,4-glucanase. J. Biol. Chem. 268:14096-14102.

Ware, C.E., T. Bauchop, and K. Gregg. 1989. The isolation and comparison of cellulase genes from two strains of *Ruminococcus albus*. J. Gen. Microbiol. 135:921-930.

Watson, J.D., N.H. Hopkins, J.W. Roberts, J.A. Steitz and A.M. Weiner. 1987. Molecular biology of the gene, Fourth edition, p. 377-378. The Benjamin/Cummings Publishing Company, Inc. Menlo Park, California 94025.

Whitehead, T.R., M.A. Cotta, and R.B. Hespell. 1991. Introduction of the *Bacteroides ruminicola* xylanase gene into the *Bacteroides thetaiotaomicron* chromosome for production of xylanase activity. Appl. Environ. Microbiol. 57:277-282.

Whitehead, T.R. 1992. Genetic transformation of the ruminal bacteria *Butyrivibrio fibrisolvens* and *Streptococcus bovis* by electroporation. Lett. Appl. Microbiol. 15:186-189.

Williams, A.G., and G.S. Coleman. 1988. The rumen protozoa, p. 77-128. In P.N. Hobson (ed.), The rumen microbial ecosystem. Elsevier Science Publishers, Crown House, Linton Road, Barking, Essex IG11 8JU, England.

Williams, A.G., and S.E. Withers. 1992. The regulation of xylanolytic enzyme formation by *Butyrivibrio fibrisolvens* NCFB 2249. Lett. Appl. Microbiol. 14:194-198.

Wolin, M.J., and T.L. Miller. 1988. Microbe-microbe interactions, p. 343-359. In P.N. Hobson (ed.), The rumen microbial ecosystem. Elsevier Science Publishers, Crown House, Linton Road, Barking, Essex IG11 8JU, England.

Wood, W.B. 1966. Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. J. Mol. Biol. 16:118-133.

Wood, T.M., S.I. McCrae, C.A. Wilson, K.M. Bhat, and L.A. Gow. 1988. Aerobic and anaerobic fungal cellulases, with special reference to their mode of attack on crystalline cellulose, p. 31-52. In J.P. Aubert, P. Béguin and J. Millet (eds.), Biochemistry and genetics of cellulose degradation, FEMS symposium No. 43. Academic Press Inc., San Diego, CA 92101.

Wood, T.M. 1992. Fungal cellulases. Biochemical Society Transactions 20:46-53.

Workshop on nomenclature of cellulases and other related enzymes. June 1994. Stanford University Sierra Camp, Lake Tahoe Basin, USA. S. Shoemaker, B. Henrissat, M. Claeyssens, M. Himmel, and T. Teeri, organizing committee.

Xue, G.-P., K.S. Gobius, and C.G. Orpin. 1992a. A novel polysaccharide hydrolase cDNA (*celD*) from *Neocallimastix patriciarum* encoding three multi-functional catalytic domains with high endoglucanase, cellobiohydrolase and xylanase activities. J. Gen. Microbiol. 138:2397-2403.

Xue, G.-P., C.G. Orpin, K.S. Gobius, J.H. Aylward, and G.D. Simpson. 1992b. Cloning and expression of multiple cellulase cDNAs from the anaerobic rumen fungus *Neocallimastix patriciarum* in *Escherichia coli*. J. Gen. Microbiol. 138:1413-1420.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.

Yoshikawa, M., and K. Sugimoto. 1993. A specific binding site on soybean membranes for a phytoalexin elicitor released from fungal cell walls by ß-1,3-endoglucanase. Plant Cell Physiol. 34:1229-1237.

Zabeau, M., and K.K. Stanley. 1982. Enhanced expression of cro-ß-galactosidase fusion proteins under the control of the P_R promoter of bacteriophage λ. EMBO J. 1:1217-1224.

Zappe, H., D.T. Jones, and D.R. Woods. 1986. Cloning and expression of a *Clostridium acetobutylicum* endoglucanase, cellobiase and amino acid biosynthesis genes in *E. coli*. J. Gen. Microbiol. 132:1367-1372.

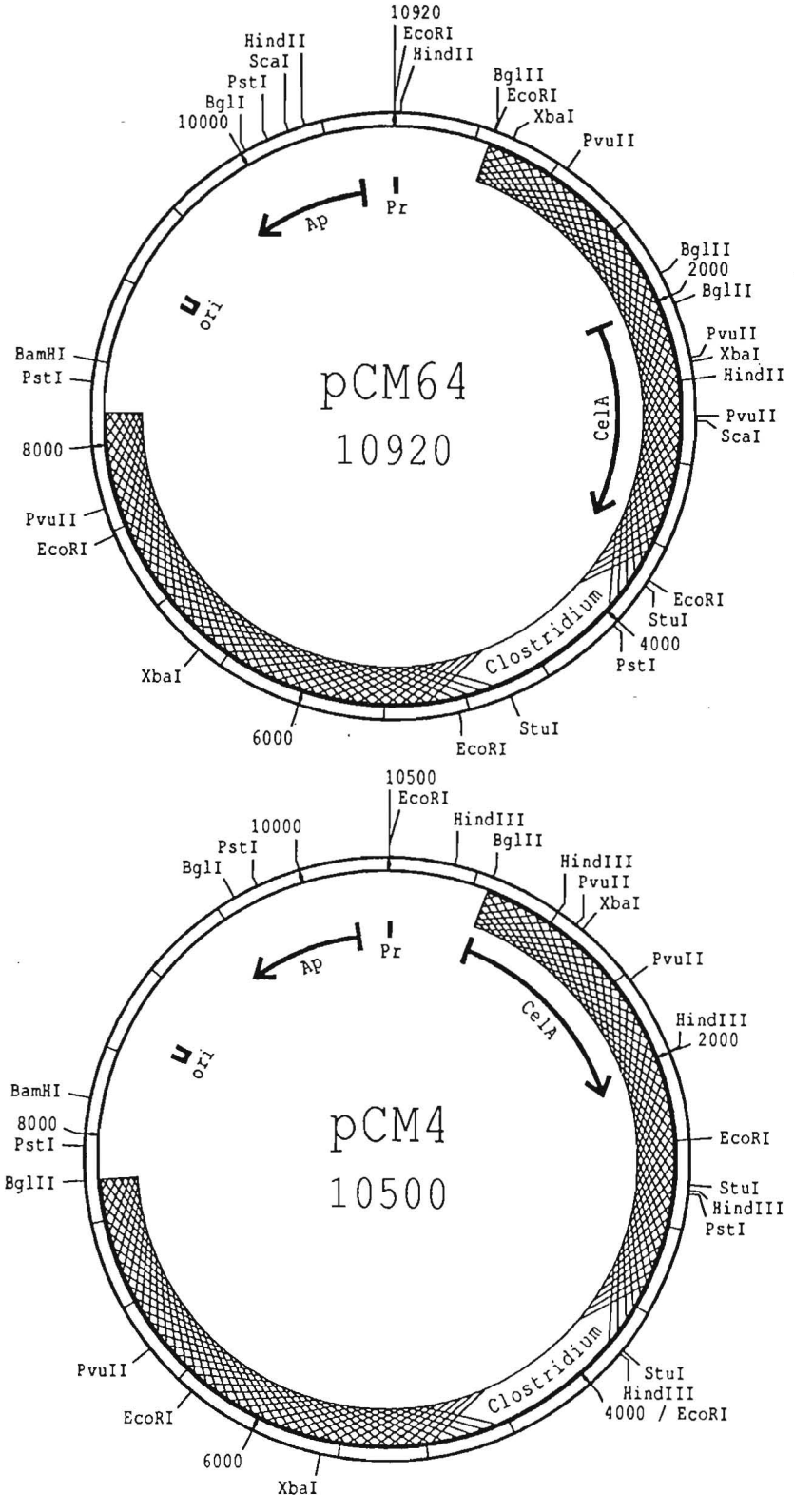
Zhang, J.-X., and H.J. Flint. 1992. A bifunctional xylanase encoded by the *xynA* gene of the rumen cellulolytic bacterium Ruminococcus flavefaciens 17 comprises two dissimilar domains linked by an asparagine/glutamine-rich sequence. Mol. Microbiol. 6:1013-1023.

Zhang, N., and J.D. Brooker. 1993. Characterization, sequence, and replication of a small cryptic plasmid from *Selenomonas ruminantium* subspecies *lactilytica*. Plasmid 29:125-134.

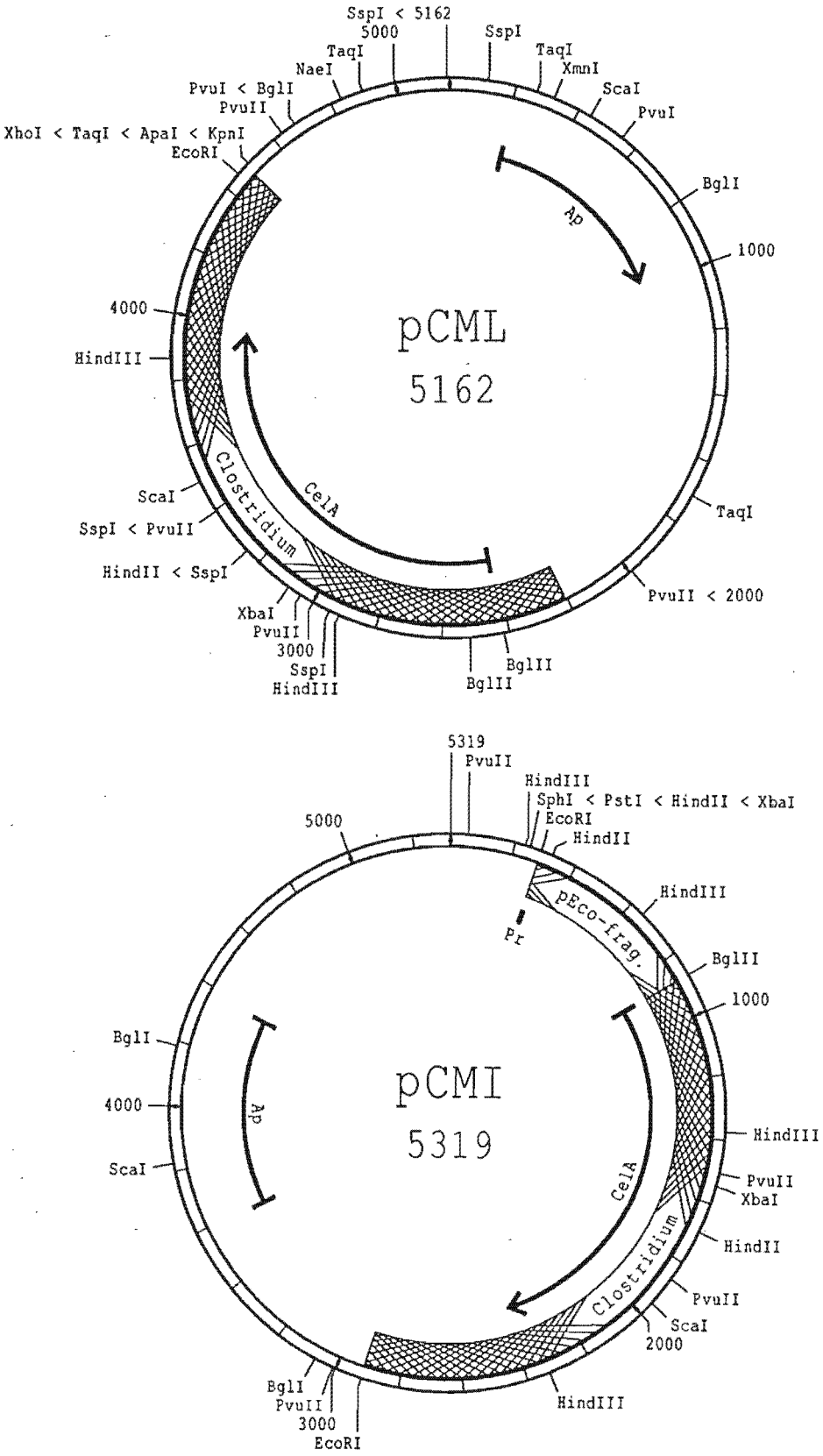
Zhou, L., G. Xue, C.G. Orpin, G.W. Black, H.J. Gilbert, and G.P. Hazlewood. 1994. Intronless *celB* from the fungus *Neocallimastix patriciarum* encodes a modular family A endoglucanase. Biochem. J. 297:359-364.

Zuker, M., and P. Stiegler. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxilary information. Nucleic Acids Res. 9:133-148.

APPENDIX



Plasmid maps of pCM64 and pCM4. The insert DNA is indicated with "Clostridium", and the positions of the *celA* ORF, the λPr promoter, the origin of replication (ori) and the ß-lactamase gene (Ap, Ampicillin resistance) are shown.



Plasmid maps of pCML and pCMI. The insert DNA is indicated with "Clostridium", and the positions of the *celA* ORF, the λPr promoter, the origin of replication (ori) and the ß-lactamase gene (Ap, Ampicillin resistance) are shown.

ABBREVIATIONS

Ap	Ampicillin
ATG	methionine / start codon
bp	base pairs
BSA	bovine serum albumine
CELAΔCBD	CelA enzyme without cellulose-binding domain
CBD	cellulose-binding domain
CBH	cellobiohydrolases
CD	catalytic domain
cDNA	complementary DNA
CFE	cell free extract
cfu	colony forming units
CMC	carboxymethyl-cellulose
CMCase	CMC degrading enzyme
Da	Dalton
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanine triphosphate
dTTP	deoxythymine triphosphate
ddA	dideoxyadenosine triphosphate
ddT	dideoxythymine triphosphate
DIG	digoxygenin
DNA	deoxyribonucleic acid
DNAse	DNA degrading enzyme
dNTPs	mixture of dATP, dCTP, dGTP and dTTP
EP	extracellular polysaccharides
EtOH	ethanol
g	gram
g	gravitational acceleration
GCG	Genetics Computer Group
G1	glucose
G2	cellobiose
G3	cellotriose
G4	cellotetraose
G5	cellopentaose
HPLC	high pressure liquid chromatography
kb	kilobase
LB	Luria-Bertani medium
mcs	multiple cloning site
Mr	molecular weight in Dalton
mRNA	messenger RNA
NBT	nitro-blue tetrazolium chloride

NRF	non-rumen fluid medium
OD_{540}	optical density measured at 540 nm wavelength
ON	overnight
ORF	open reading frame
P_R	phage lambda rightward promoter
PCR	polymerase chain reaction
RNA	ribonucleic acid
RNAse	RNA degrading enzyme
RNA:ssDNA	mRNA hybridized to single-stranded DNA
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
ssDNA	single stranded DNA
TAE	Tris-acetate-EDTA electrophoresis buffer
TE	Tris-EDTA buffer (10 mM Tris.Cl, 1 mM EDTA)
Tris	tris(hydroxymethyl)aminomethane
Tris.Cl	Tris solution, pH adjusted with HCl
U	units
w/v	weight per volume (in g per 100 ml)
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactoside
3-D	three-dimensional
α	alpha
ß	beta
Δ	delta
γ	gamma
λ	lambda (referring to phage)
μ	micro