MOLECULAR CLONING AND CHARACTERIZATION OF CELLULASE GENES OF *RUMINOCOCCUS FLAVEFACIENS* STRAIN 186

by

CHUNG-MING HUANG

A thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Biotechnology at Massey University, Palmerston North, New Zealand

1990

TABLE OF CONTENTS

ABSTRACT

A genomic library of *Ruminococcus flavefaciens* 186 was constructed using the lambda vector λNM1149. It constituted 2.1 x 10^4 recombinant clones which was large enough to represent the entire genome of this bacterium. From this library, 26 CMC^+ clones were identified after screening about 2500 recombinant clones. These CMC^+ clones were divided into four groups according to their insertion fragment size. Partial restriction maps of these clones have been achieved. Eight representative clones from these four groups showed different mapping patterns. One of the three 9 kb insert clones (λCM903) was selected for further study based on its ability to express all three types of cellulase activities.

The locations of endoglucanase and exoglucanase genes in λCM903 were determined. Two separate fragments from λCM903 were subcloned and named λCMEH1 (CMC^+) and λCMEH2 (MUC^+). Cross hybridization experiment using *cel*A gene of *C. thermocellum* and the 26 recombinant clones showed no significant homology. However, different degrees of homology were found among the 26 ruminococcal clones.

The difficulty of subcloning the ruminococcal DNA fragments into plasmid vectors was one of the major obstacles in the study of cellulase gene in foreign hosts. Vectors with different functions were tried but all the recombinant plasmid clones showed instability.

The cellular location of cellulase enzymes in *E. coli* cell was determined. Most of the endo- and exo-glucanases were found in the periplasmic space. Partial purification of these cellulase enzymes from *E. coli* cells using chromatography was then performed and the characterization of these enzymes was achieved.

Using *Exo*III deletion, the locations of endoglucanase and exoglucanase genes were determined in λCM903. The internal *Hin*dIII-*Hin*dIII fragment of 7.3 kb from λCM903 was sequenced. Five ORFs were detected using computer software (UWGCG) analyses.

The first ORF which coded for an endoglucanase gene (*ren*A) was 2157 bp long with putative 680 amino acid residue. The SD sequence and promoter sequence were present. The best fitting cellulase gene tested was that of the *cen*A gene of *Ce. fimi.* An unusual structure of Pro-Thr-Ser rich region, which had 38 out of 42 a.a. residue of proline, threonine or serine, was found in the N-terminal of the putative peptide. The second ORF which was 1821 bp long coded for an exoglucanase gene (*rex*). The putative amino acid sequence had 572 a.a. residue, also a SD sequence and a promoter sequence were found. A Pro-Thr-Ser rich region, which was highly conserved with PTS of *ren*A, was found at the C-terminal of the putative peptide. Again, the cellulase gene, *cex* gene of *Ce. fimi* showed the best similarity. It is suggested that the gene structure of cellulase in *R. flavefaciens* strain 186 was similar to cellulase genes in family A as described by Ong et al, 1989. The third ORF was found overlapping with ORF1 using the transcriptional second reading frame. This ORF had a putative SD sequence but lacked a promoter sequence. The coding region of this ORF has the characteristics of a β-glucosidase gene which was 1300 bp long (a putative sequence of 443 amino acid residues). The fourth ORF which used the second reading frame was 1300 bp long with a SD sequence 5' upstream of the ATG codon, overlapped with ORF1 gene. This ORF coded for a protease gene. The fifth ORF located closely to the 3' end of the 7.3 kb fragment used the second reading frame. This ORF had both putative SD sequence and promoter sequence. It was 1080 bp long with putative 341 amino acid residues and showed the structure of a xylanase gene. A short PTS region was also found in this ORF.

ACKNOWLEDGEMENTS

I wish to sincerely thank my chief supervisor Dr. Pak-Lam Yu for his constant interest over these years and for his patience in advising, encouragement, guidance as well as supporting the grants for this Ph.D study.

I would also like to thank my co-supervisor Dr. R.V. Asmundson and my ex-supervisor Dr. W.J. Kelly for their helpful discussions, encouragement and guidance.

Special thanks also goes to the following people for their help and friendship which are greatly appreciated.

- Dr. J.J. Patel, Dr. K.N. Joblin, Dr. D.R. Greenwood, Dr. D.R. Biggs, Dr. M.S. Grant, Dr. D.E. Ottor, Dr. S. Phua, Mr. G.E. Naylor, Ms. M. Curry, Mr. R. Townsend, Ms. A. Ede and Ms. M. Carpenter
- Assoc. Prof. I. Maddox, Assoc. Prof. B. Chong, Mr. CX Xu, and Mr. FF Xu
- K.L. & W. Teo, C. & F. Chua, Y.H. & C. Chi, K. & K. Lee, F. & T. Yen, W. Yu and C.D. & A.I. Lai

To my family: my mother, my wife, brother, sisters and my son, I would like to show my greatest gratitude, for their unending love and supports.

Also, thanks goes to DSIR biotechnology Division which provided me the facilities to carry out this study.

LIST OF PUBLICATIONS

Publications or abstracts arising from this thesis

1. Molecular cloning and expression of multiple cellulase genes of *Ruminococcus flavefaciens* strain 186 in *Escherichia coli*. Huang CM, Kelly WJ, Asmundson RV and Yu PL, Appl. Microbiol. Biotechnol. (1989) 31: 265-271.
2. Purification and cloning of cellulase from *Ruminococcus flavefaciens*. Huang CM, Kelly WJ, Yu PL and Asmundson RV, Collogue FEMS symposium, Paris. 1987.
3. Molecular cloning of cellulase genes of *Ruminococcus flavefaciens* strain 186 in *E. coli*. Huang CM, Kelly WJ, Asmundson RV and Yu PL, New Zealand, Australia Microbiological combined meeting, Auckland. 1987.
4. Characterization of the cellulases and cellulase genes of *Ruminococcus flavefaciens* involved in cellulolytic fermentation. Huang CM, Asmundson RV and Yu PL, Fermentation technologies: Industrial application. (P.L. Yu ed.) Elsevier Applied Science, UK. 1990.
5. Molecular analyses of cellulase genes and cellulase complexes from *Ruminococcus flavefaciens*. Huang CM, Asmundson RV and Yu PL, Microbiological Society annual meeting, New Zealand. Lincoln, 1990.
6. Structure of a cellulase gene from *Ruminococcus flavefaciens*. Huang CM, Yu PL and Asmundson RV, 6th International Symposium on the Genetics of Industrial Microorganisms. Paris. 1990.
7. Structure of the genes encoding endo-, exo-glucanase and β-glucosidase of *Ruminococcus flavefaciens* strain 186. (in preparation)
8. Characterization of endo- and exo-glucanase and β-glucosidase of *Ruminococcus flavefaciens* strain 186 expressed in *E. coli*. (in preparation)
9. Structure of the gene encoding xylanase of *Ruminococcus flavefaciens* strain 186. (in preparation)

LIST OF FIGURES

Chapter I

Chapter III

Chapter IV

Chapter V

Chapter VI

LIST OF TABLES

CHAPTER I

INTRODUCTION

Table of contents

1-1 Introduction

For a long-term solution to our resource problems of energy, chemical and food, cellulose should be considered because it is a renewable carbon source and the most abundant organic molecule in nature. Cellulose is the major structural component of higher plants which is synthesized efficiently as a result of photosynthesis. Decomposition of cellulolytic materials in industry is processed normally by harsh physical or chemical treatments which could result in the pollution of the environment and low yield of substrate. Bioenergy technology, which is often called "soft energy paths" as opposed to the types of energies that are commonly used, is an alternative way to overcome these problems. It is the technology which uses microbial enzymes to degrade biomass into products that could be directly or indirectly used as energy, chemical, feedstock or food.

Cellulose decomposition in nature is a complex process, involving mixed populations of many microorganisms. It is carried out by fungi, unicellular bacteria and actinomycetes (Waksman and Skinner, 1926; Alexander, 1961; Imsenecki, 1968). In general, fungi are the initial and principal decomposing microorganisms and these are followed by the bacteria and actinomycetes (Waksman and Skinner, 1926; Tribe, 1960). The cellulolytic fungi include several ascomycetes, (orders Eurotiales and Sphaerials), or their imperfect counterparts such as *Penicillium*, *Aspergillus*, *Trichoderma*, *Fusarium*, and *Chaetomium* (Mandels, 1981). Basidiomycetes are also involved in cellulose decomposition, especially as wood-rotters (Gilbertson, 1980), and the white rot fungi (eg. *Phanaerochaete*) have been implicated in lignin decay. Cellulolytic bacteria include the aerobes *Cytophaga*, *Sporocytohaga*, *Cellvibrio*, *Cellulomonas*, *Pseudomonas* (Imsenecki, 1968), *Erwinia* (Kotoujansky et al, 1985) and *Bacillus* (Murphy et al, 1984), the anaerobic mesophiles *Clostridium* (Hungate, 1944), *Rumino-*

coccus (Hungate, 1966), *Bacteroides* (Groleau and Forsberg, 1981) and *Acetivibrio* (Patel et al, 1980), and the thermophiles *Clostridium thermocellum* (Viljoen et al, 1926; McBee, 1948), *Caldocellum saccharolyticum* (Love et al, 1987) and *Thermomonospora* sp. (Stutzenberger, 1987; Hu and Wilson, 1988). Each can easily be isolated from most decomposing samples of cellulose.

The rumen is a very specific microenvironment for cellulose decomposition. Most polysaccharides entering the rumen can be considered as belonging to one of two general types: plant storage polysaccharides, or structural polysaccharides. The structural polysaccharides constitutes the greater part of all plant cell walls and are loosely considered to form the fibrous component of animal feedstuffs. Storage polysaccharides can function as food reserves and as such must be readily mobilised when required by the plant. As a consequence they are easily degraded by plant hydrolytic activities and are susceptible to attack by enzymes secreted by rumen microorganisms. The structural polysaccharides have a skeletal function in the living plant and are far more resistant to microbial attack. However, it is the ability to utilise such materials as an energy source that provides ruminants with their particular ecological niche, and the ability of rumen microorganisms to degrade plant polysaccharides efficiently is of paramount importance.

Recent progress in recombinant DNA methodology has made it possible to "genetically engineer" bacteria so that they are capable of performing functions very different from those of the wild-type strains. There is great potential for applying this technique to enzyme technology to improve the production of enzymes or to study the regulation of the expression of particular genes in industrial microorganisms.

1-2 Cellulolytic microorganisms

A summary of cellulolytic microorganisms which have been studied is shown in Table 1-1.

Table 1-1. Cellulase-producing microorganisms

Bacteria	Fungi
Acetivibrio cellulolyticus	*Agaricus bisporus*
Bacillus spp.	*Aspergilus* spp.
Cellulomonas fimi	*Botryodiplodia mxp.*
Cellvibrio fulvus	*Chaetomium thermop*
C. gilvus	*Eupenicillium javenicum*
C. vulgaris	*Fusarium solani*
Clostridium thermocellum	*Humicola insolens*
C. chartabidum	*Macrophomina pseudolina*
C. thermocellulaseum	*Myrothecium verrucaria*
C. thermomonospora	*Myceliophthora thermophila*
Erwinia chrysanthema	*Pellicularia filamentosa*
Fibrobacter cellulosolvents	*Penicillium* spp.
F. succinogenes	*Pestalotiopsis westerdijkii*
Pseudomonas fluorescens	*Polyporus* spp.
Ruminococcus albus	*Poria* spp.
R. flavefaciens	*Sporotrichum* spp.
	Talaromyces emersonii
Actinomycetes	*Thermoascus aurantiacus*
	Thielavia terrestris
Streptomyces griseus	*Tranetes sanguinea*
Thermoactinomycete spp.	*Trichoderma* spp.
Thermomonospora curvata	*Trichosporon* spp.
T. fusca	

based on Coughlan, 1985

1-2-1 Fungi

Although several fungi (eg. *Aspergillus* sp., *Penicillium* sp.) which degraded cellulose were found very early, the study of cellulase production from these fungi other than the genus *Trichoderma reesei* was limited.

Trichoderma reesei (formerly *T. viride*)

The ascomycetous, brown-rot fungus *T. reesei* is well studied and probably the most efficient cellulase producer. Its enzyme system consists of three different hydrolases: endoglucanase, cellobiohydrolase and β-glucosidase which act synergistically in the degradation of crystalline cellulose. However, knowledge about the pathway of cellulase formation and secretion in this organism is limited.

Although the mechanism of secretion of cellulase in *T. reesei* is not yet clear, what is known is summarized as follows: the primary peptide with its core carbohydrate (O-linked) attached to asparagine residues is synthesized in the endoplasmic reticulum. In the Golgi, carbohydrates are added to serine and threonine residues and the glycoprotein is then packaged into secretory vesicles and transported to the plasma membrane. The vesicles fuse with the membrane and the enzymes are released into the medium. (Merivuori et al, 1985, Salovuori et al, 1987; Messner and Kubicek, 1988).

Several mutation and cloning studies to improve the productivity of this organism have been done and considerable success has been achieved. The genealogy of these mutants has been summarized by Montenecourt (1983).

1-2-2 Bacteria

Three major sources of cellulolytic bacteria occur: soil, hot springs and the rumen. Some of the representative genera or species from these environments are described below:

Clostridium thermocellum

C. thermocellum was first isolated from manure. It is Gram positive, rod shaped (5.0 x 0.4 μm), produces oval shaped

spores (0.9 x 0.6 μm), and has irregular colonies when grown on agar (Viljoen et al, 1926). All *C. thermocellum* strains are able to ferment cellulose and its hydrolysis product, cellobiose. Also it degrades xylan, but the resulting xylose and xylobiose are not fermented and accumulate in the broth (Gordo, 1981).

The cellulase of *C. thermocellum* is active and stable in the absence of substrate from 37 °C to 70 °C, but loses all its activity in a short time at 80 °C. It shows resistance to cellobiose inhibition when growing on artificial celluloses such as dyed Avicel (Ng and Zeikus, 1981), but is strongly inhibited by cellobiose when growing on Avicel (Johnson et al, 1982). Cellobiose is converted into glucose 1-phosphate and glucose by the enzyme cellobiose phosphorylase produced by this organism (Alexander, 1972; Swisher et al, 1964).

Cellulomonas fimi

Ce. fimi is a cellulolytic, Gram positive coryneform bacterium of 72 mol% G+C (Stackerandt and Kandler, 1979). Only part of the extracellular activity responsible for the hydrolysis of cellulose is present in the supernatant of grown cultures, and the remainder is bound to cellulose (Beguin et al, 1977). Three extracelluar cellulases have been purified (Beguin and Eisen, 1978). One was found in the cell-free supernatant and two were found to be bound to the cellulose. Endoglucanase (CenA) and exoglucanase (Cex) have been identified by binding to cellulose (Langsford et al, 1984; Gilkes et al, 1984).

Bacillus sp.

Cellulolytic *Bacillus* sp. were very easy to isolate from soil or cellulosic substrates (Priest, 1977). Several species have been reported including *B. subtilis* (Lo et al, 1988;

Robson and Chambliss, 1984), alkalophilic *Bacillus* (Fukumori et al, 1987; Horikoshi et al, 1984), and anaerobic *Bacillus cellulosae dissolvens* (Khouvine, 1923). The general characterization is rod shaped (2.0 x 0.4 µm), oval shaped spores (2.0 x 2.5 µm), Gram positive and containing about 42 mol% G+C.

The first case of bacterial cellulase used in commercial and industrial applications was that from alkalophilic *Bacillus* sp. (Kawai et al, 1988).

Fibrobacter succinogenes (formerly *Bacteroides succinogenes*)

F. succinogenes, first described by Hungate (1950), is now considered to be one of the most widespread cellulolytic bacteria of the rumen. The characteristic of this organism was Gram negative, rod shaped to lemon shaped or oval, 0.8 to 1.6 µm in diameter, occurring singly or in short chains. The major substrates fermented are listed in Table 1-2. Valerate, isobutyrate, CO_2, NH_4^+, Na^+, K^+, Ca^{+2}, Mg^{+2} and PO_4^{-3} are essential for growth; biotin is usually essential.

The cellulases of *F. succinogenes* and its mode of attack on plant cell walls have been reviewed by Chesson and Forsberg (1988). Their action is in the hydrolysis of highly ordered celluloses such as cotton fibres (Halliwell and Bryant, 1963) and plant materials (Dehority and Scott, 1967; Stewart et al, 1979; Morris and Van Gylswyk, 1980; Morris, 1984; Kolankaya et al, 1985; Graham et al, 1985; Chesson et al, 1986). When the cell attaches to cell walls of plant materials, *F. succinogenes* exhibits a tight adhesion, frequently conforming to the surface of the material being digested (Forsberg et al, 1981; Cheng et al, 1983,1984).

Ruminococcus

The name *Ruminococcus flavefaciens* was conferred by

Table 1-2 Fermentation characteristics of *Ruminococci* and *Fibrobacterium*.

	R. albus	*R. flavefaciens*	*F. succinogenes*
Acid from:			
starch	-	-	d
cellulose	+	+	+
xylan	+	+	-
pectin			d
maltose	-	-	d
cellobiose	+	+	+
sucrose	d	d	-
D-xylose	d	d	-
L-arabinose	d	d	-
glucose	d	d	+
fructose	d		-
galactose	-	-	-
mannose	d	d	-
lactose	d	d	d
mannitol	d	-	-
glycerol	-	-	-
lactate	-	-	-
Aesculin hydrolysis	d	d	-
H_2S production	-	-	-
Nitrate reduction	-	-	-
Fermentation product:			
Major	AE	AS	AS
Minor/some strains	FL	FL	FPiV
Gas produced	H_2, CO_2	H_2	

Abbreviations: A, Acetate; E, ethanol; F, Formate; L, Lactate; P, Propionate; S, Succinate; iV, isovalerate; d, reaction varies between strains; blank, no test.
From: Stewart and Bryant, 1988.

Sijpesteijn (1951) on Gram-positive, non-motile, cellulolytic streptococci with cells 0.8-0.9µm in diameter, which occurred singly, in pairs and chains. A yellow pigment was produced, particularly during growth on cellulose. The species *R. albus* and *R. flavefaciens* are among the most active bacteria involved in plant cell wall digestion in the rumen (Dehority and Scott, 1967; Stewart et al., 1979; Morris and Van Gylswyk, 1980; Morris, 1984; Dehority, 1986; Bryant, 1986; Stewart, 1986; Chesson et al., 1986).

Ruminococcus flavefaciens

The morphological characteristics of *R. flavefaciens* are as described above. Almost all strains are cellulolytic, and

additional fermentation characteristics are summarised in Table 1-2. Trace amounts of branched-chain fatty acids are required for growth and NH_3 is essential (Bryant, 1986).

The role of *R. flavefaciens* in plant cell wall breakdown, established in a series of studies on ruminococci and other rumen bacteria, has been further elucidated by scanning electron microscopy. When incubated with leaves of perennial ryegrass, *R. flavefaciens* mainly colonised the cutting edges of the epidermis, sclerenchyma and phloem cells (Latham et al, 1978). When incubated with orchard grass and bermuda grass leaves and the distribution and activity of the bacteria studied, the digestion of epidermis and parenchyma bundle-sheath cells was accomplished by attached bacteria. However, bacteria did not become attached either to the readily degraded mesophyll cells or to the indigestible xylem vessels. Transmission electron micrographs (TEM) of *R. flavefaciens* strain 186 attached to Avicel (unpublished data, Asmundson) indicate a similar mode of cellulose attach (Fig. 1-1).

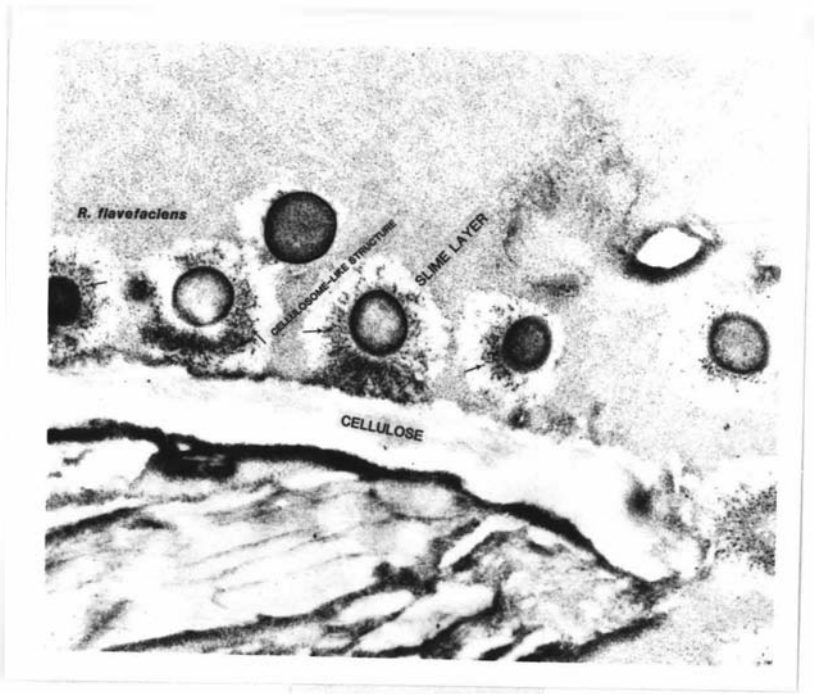


Figure 1-1 TEM of *R. flavefaciens* attach to Avicel.

Ruminococcus albus

A simple way to distinguish *R. albus* and *R. flavefaciens* is the colony colour when growing on cellulose: white and yellow, respectively. *R. albus* typically ferments cellulose, cellobiose and glucose, and may ferment a range of other carbohydrates (Table 1-2) together with rhamnose, and salicin (Bryant, 1986).

R. albus has attracted particular interest as a result of its production of high-molecular weight, cell-bound cellulases (Wood et al, 1982; Wood and Wilson, 1984), cellobiosidase (Ohmiya et al, 1982) and β-glucosidase (Ohmiya et al 1985). Cultivation of *R. albus* on different substrates influenced the production of polysaccharide depolymerases (Williams and Withers, 1982). It has been shown that the 3-phenylpropanoic acid (PPA) required for growth of some strains (Hungate and stack, 1982) enhances the production of vesicular structures by the bacterium and increases the association of the cellulase enzyme with the cells in a high molecular-weight form (Stack and Hungate, 1984). This requirement for PPA is shared by type-strain 7 and strain 8 of *R. albus*, but not by *R. flavefaciens* (Stack and Cotta, 1986). In the presence of PPA, phenylacetic acid further enhances cellulose digestion by *R. albus* strain 8 (Stack et al, 1983).

1-3 Biochemical studies of cellulose degradation

1-3-1 Cellulose structure

Cellulose is a linear polymer of up to 14000 anhydro-glucose residues in a chair configuration held together by β-(1,4)- linkages. Each residue is rotated 180 degrees about the main axis with respect to its neighbouring residues (Fig. 1-2). The basic recurring unit is cellobiose. Cellulose chains are held together by hydrogen bonding and orientated in parallel and staggered with respect to their partners to form insoluble

fibrils. The hydrogen-bonding network consists of inter- and intramolecular bonds between successive and adjacent dextrose residues (Fig.1-2) (Gardner and Blackwell, 1974; Winterburn, 1974; Rees et al., 1982). The intramolecular bonds help to maintain the rigidity of the cellulose chain. Bundles of fibrils aggregate to form the inert, insoluble fibres of great strength characteristic of the primary and secondary cell walls of higher plants (Ree et al., 1982; McNeil et al., 1984)(Fig.1-3). Within cellulose fibres there are areas of complete order (crystalline regions) and also less well-ordered or amorphous regions.

1-3-2 Cellulolytic systems

The numerous published papers dealing with cellulase have been reviewed by many authors, notably by Bailey, Enari and Linko (1975), Sternberg (1976), Gong and Tsao (1979), Ryu and Mandels (1980), Natick programme (Anonymous, 1981), Bisaria and Ghose (1981), Linko et al (1983), Wiseman et al (1985), Russell et al (1985), Kennedy et al (1987), Chynoweth et al (1987), Lamed and Bayer (1988), Hobson et al (1988) and Ong et al, (1989).

Several cellulase enzymes with different specificities and modes of action are produced by cellulolytic organisms. Three types of cellulase enzymes are suggested to be involved in the degradation of cellulose based on substrate specificity: endoglucanase (C_1, EC 3.2.1.4) referred to as carboxymethyl-cellulase (CMCase), exoglucanase (C_x, EC 3.2.1.91) also referred to as methylumbelliferylcellulase (MUCase), Avicelase, or cellobiohydrolase, and β-glucosidase or cellobiase (EC 3.2.1.21). In principle, the endoglucanase hydrolyzes cellulose by randomly breaking down internal β-(1,4) glycosidic bonds. The exoglucanase removes cellobiose units from the reducing end of the cellulose chain. Finally, β-glucosidase breaks the final β-(1,4) linkage of cellobiose. The conversion of cellulose to

Figure 1-2 Cellulose chains showing the β-1,4-linked residues rotated through 180 degrees with respect to their neighbours in the chain. Intermolecular hydrogen bonds tightly crosslink adjacent chains within a microfibril. Intramolecular hydrogen bonds stabilize each chain. (From: Alberts et al, 1983)

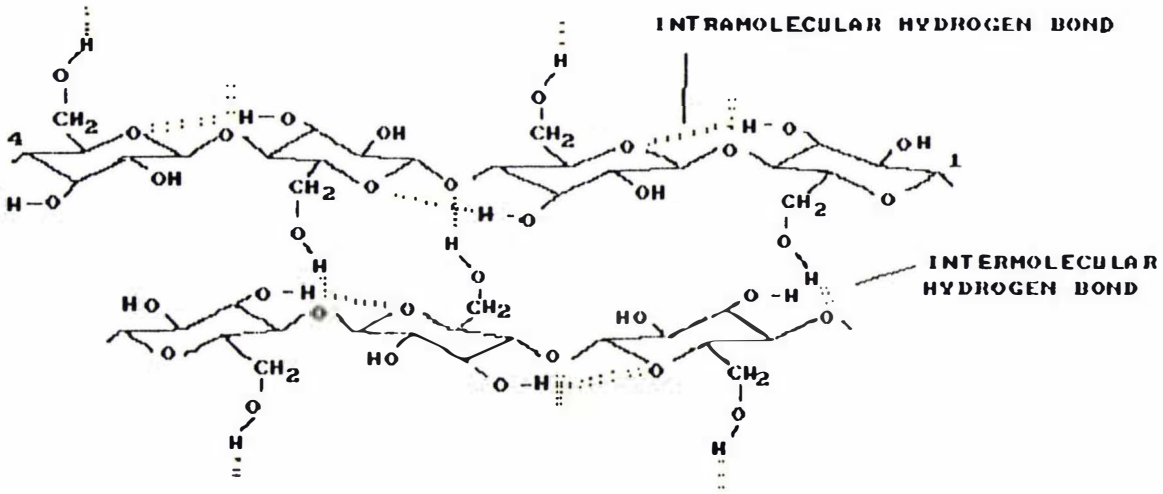
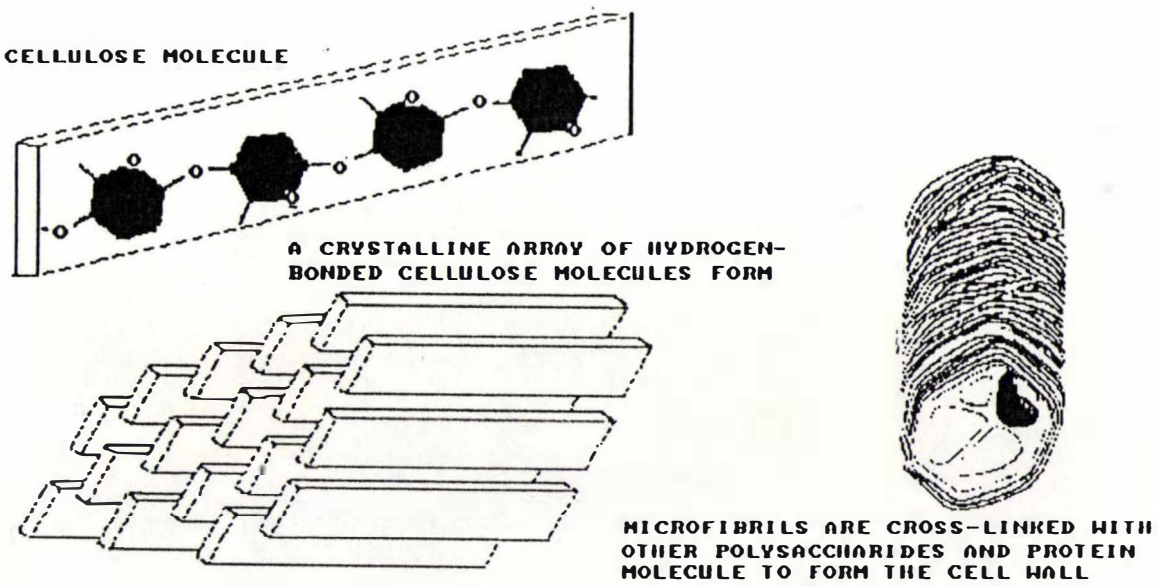


Figure 1-3 Organization of cellulose molecules



glucose is the combined effect of all three activities (Fig. 1-4).

Overall, the enzymatic degradation of cellulose can be divided into two steps: adsorption and hydrolysis.

1-3-2-1 Adsorption

Factors affecting the adsorption of cellulases to cellulose may include the nature of the substrate, its purity, pretreatment, the extent of crystallinity or amorphous cellulose content, the enzyme/substrate ratio, the affinity of the multicomponent enzyme system used, the topography of the substrate which changes as digestion proceeds, factors causing inactivation of the bound or free enzyme, and the effects of reaction products or other substances that might promote or inhibit adsorption. However, the adsorption of cellulase usually parallels the rate of hydrolysis of cellulose.

Endoglucanase and exoglucanase adsorb at distinctly different sites on cellulose, these sites correspond to the sites of hydrolysis. The addition of exoglucanase to substrate to which endoglucanase is bound causes the latter to speed up its action and bring about its desorption. Conversely, endoglucanase on binding to cellulose speeds up the rate of scission by exoglucanase and brings about its desorption (Ryu et al., 1984; Fig. 1-5). Each enzyme is generally found to exist in several forms and these forms may differ in substrate specificity, ability to adsorb to the substrate and in their capacity to interact synergistically with other enzymes in the system.

Synergism is most marked when highly crystalline substrates are used, is low with amorphous cellulose and absent with soluble derivatives (Wood and McCrae, 1979). A separated endoglucanase in a single filtrate may or may not act synergis

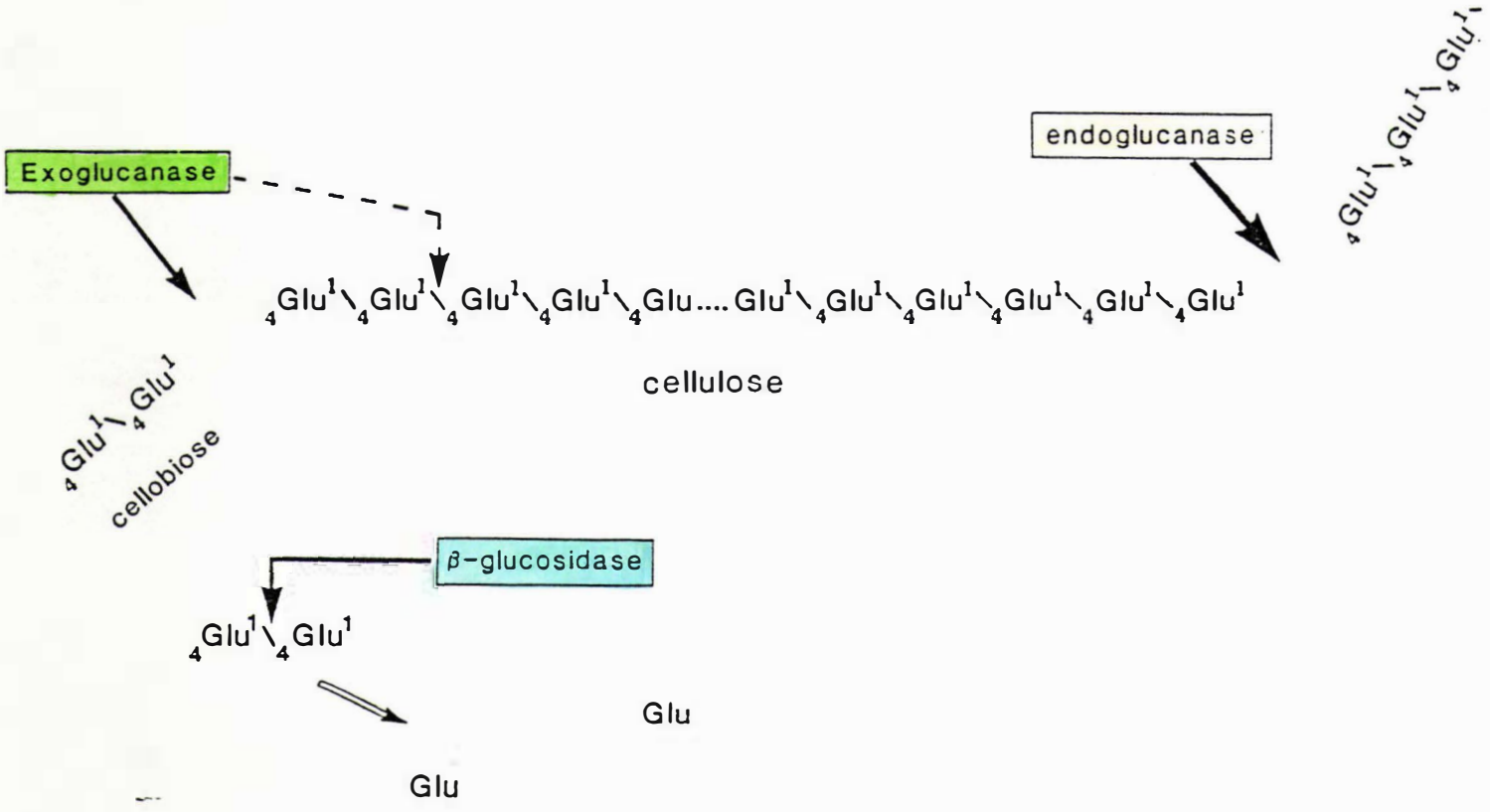


Figure 1-4 Diagram of cellulases acting in cellulose chain

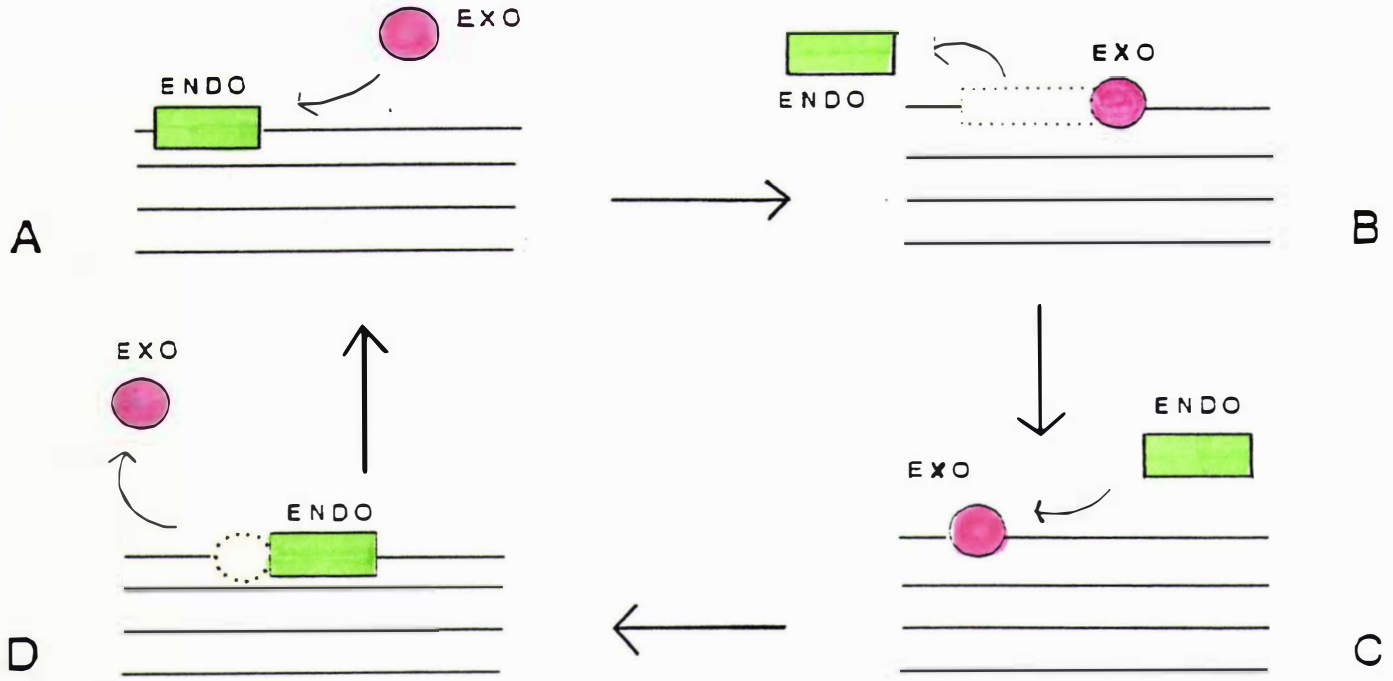


Figure 1-5 Competition between endo- and exo-glucanase for adsorption (based on Ryu et al, 1984)

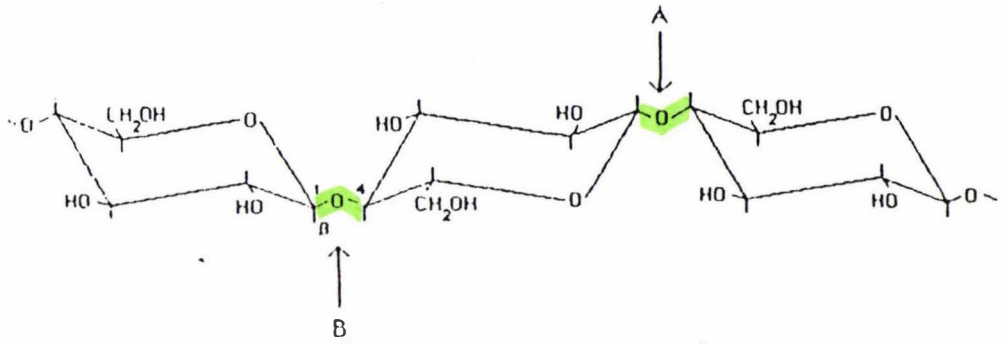
tically or equally synergistically with the exoglucanase fraction from the same filtrate (Eriksson, 1975; Wood, 1975). The two cellobiohydrolases I and II from *Trichoderma reesei* exhibit different substrate stereospecificities, each attacking one of the two different non-reducing end groups that may be found in the substrate (Fig. 1-6). Removal of cellobiose units from one type of non-reducing chain end could expose, on an adjacent chain, a non-reducing end of the other type from which the second stereospecific cellobiohydrolase would remove cellobiose. The successive operation of each enzyme on the adjacent chains could then explain the observed synergism (Wood, 1975, 1980).

1-3-2-2 Hydrolysis mechanism

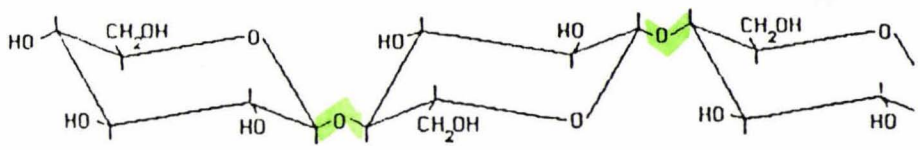
The N-terminal sequence of an endoglucanase (EGI) of *Schizophyllum commune* (Yaguchi et al., 1983), from Glu-33 to Tyr-51 was homologous with the active-site sequence of hen egg-white lysozyme including the lysozyme catalytic residues (Glu-35, Asp-52) and the substrate-binding residue, Asn-44. Also, the sequence of active site region, residues Glu-65 to Asp-74 in an exocellobiohydrolase produced by *Trichoderma reesei* is homologous with the catalytic sequence Glu-11 to Asp-20 in the lysozyme produced by phage T4 (Paice et al., 1984). These similarities support the hypothesis that cellulases and lysozymes utilize a common catalytic mechanism (Fig. 1-7)

The hydrolysis of cellulose by bacterial systems is illustrated in Fig.1-8. The cell-associated high-molecular-weight entity, the cellulosome which is a multienzyme complex, is thought to be responsible for adherence of the bacterium to the substrate, and the endoglucanase and exoglucanase in the cellulosome are to permit extensive hydrolysis (Hofsten and Berg, 1972; Wood, et al., 1982; Lamed et al, 1983,1988). Cellobiose and short chain oligosaccharides produced by the cellulosome are metabolized by β-glucosidases or cellobiose

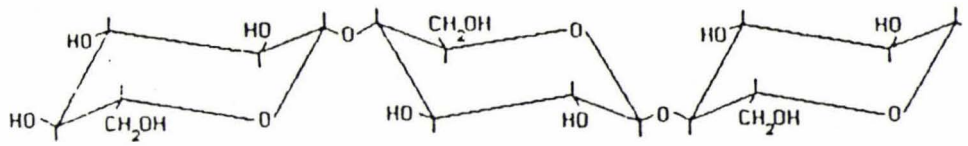
Figure 1-6 Substrate stereospecificity of two exoglucanase (based on Wood, 1985)

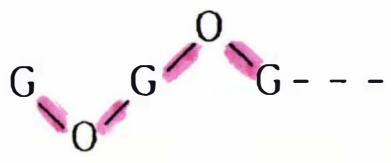


Cleavage at A gives end group type I

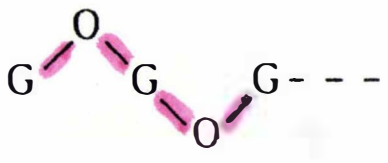


Cleavage at B gives end group type II





Type I



Type II

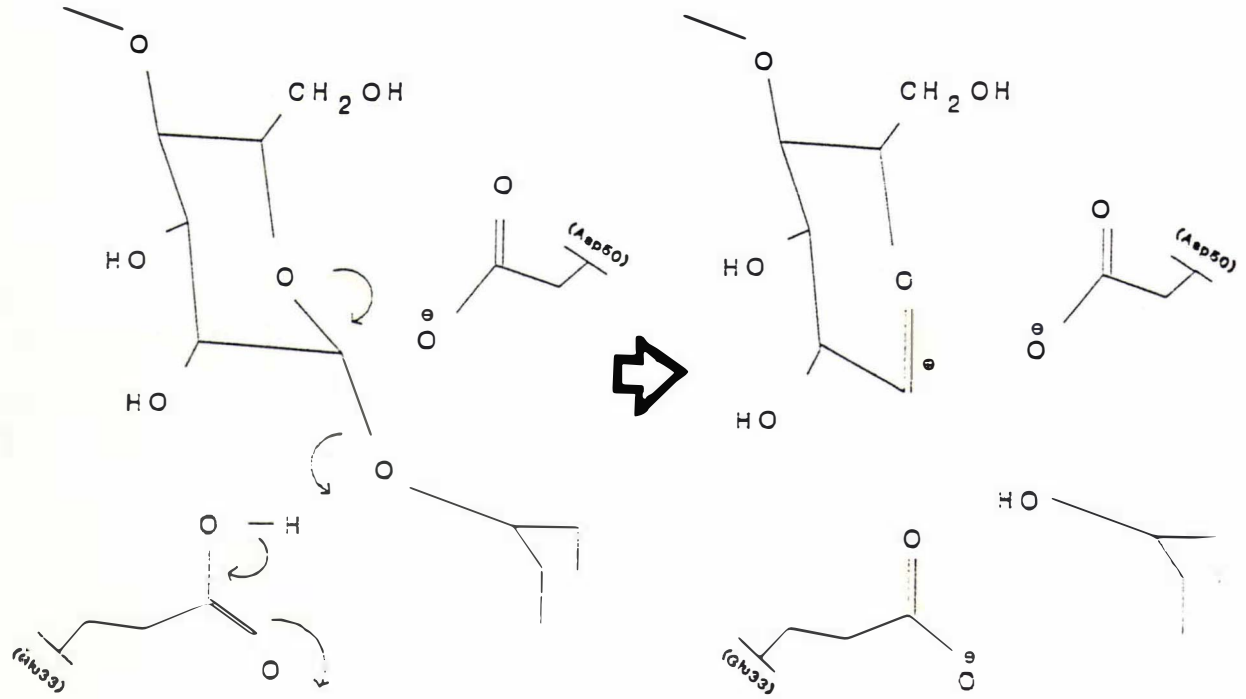


Figure 1-7 Proposed mechanism for the endoglucanase-catalysed cleavage of β-1,4-linkages in cellulose (from: Yaguchi et al, 1983)

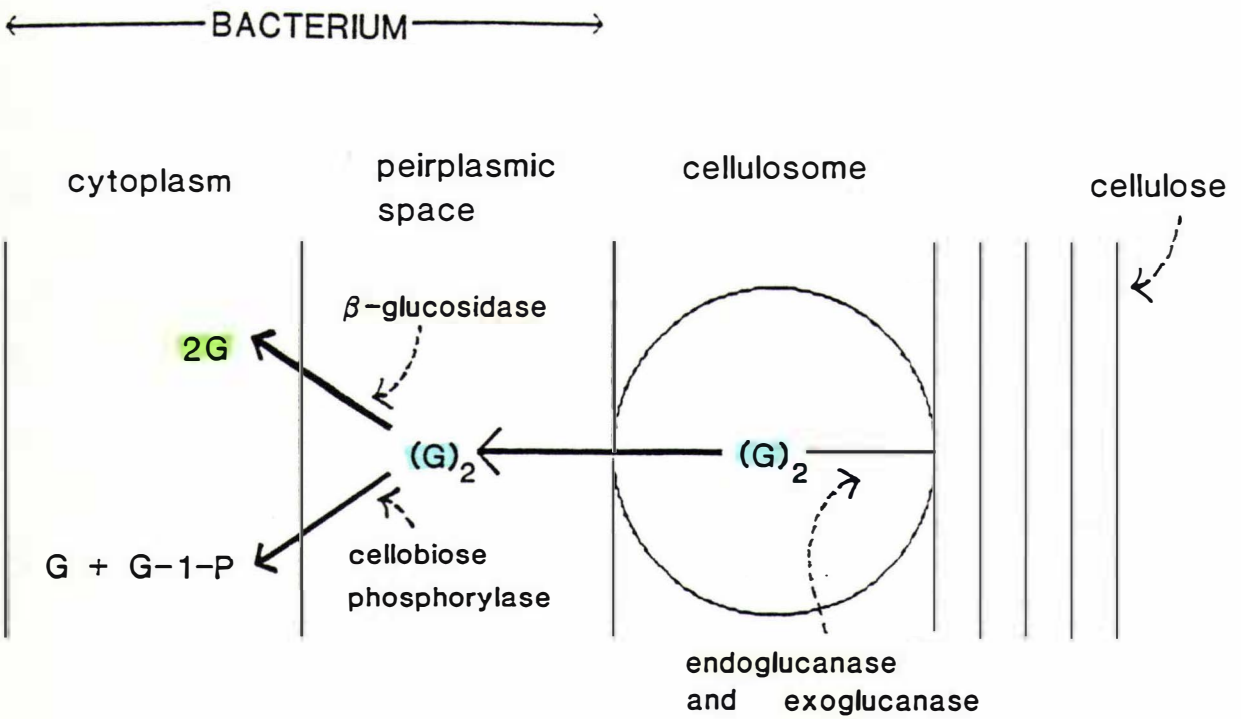


Figure 1-8 Hydrolysis of cellulose by bacterial cellulolytic enzyme system (based on Lamed et al, 1983)

phosphorylase in the periplasmic space and glucose or glucose-1-phosphate enters intracellular metabolism via an ATP-dependent permease or phosphorylase and hexokinase (Hernandez, 1982; Ng and Zeikus, 1982; Lamed and Zeikus, 1980).

1-4 Genetic studies of cellulase genes

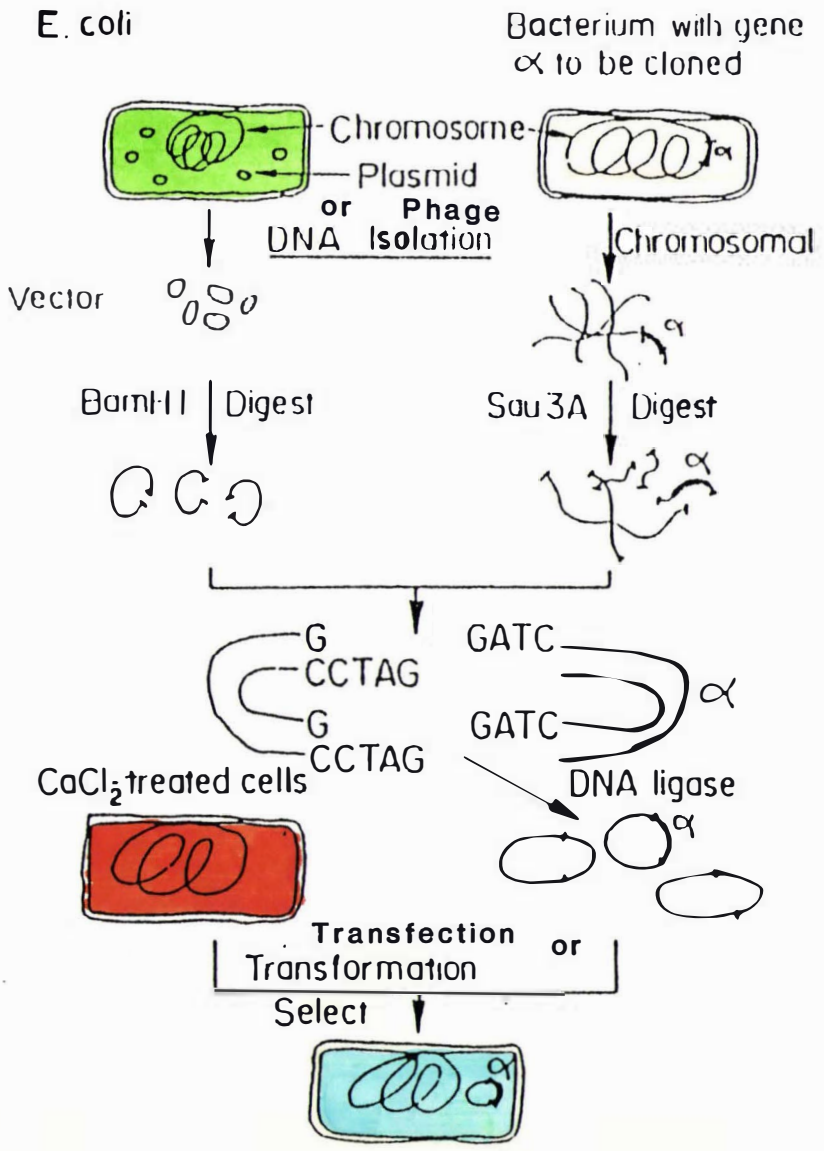
1-4-1 General concept of gene manipulation

Gene manipulation has been defined as the formation of new combinations of heritable material by the insertion of nucleic acid molecules, produced outside the cell, into any virus, bacterial plasmid or other vector system and allowing their incorporation into a host organism in which they do not naturally occur, but in which they are capable of continued propagation (Old and Primrose, 1985). This technology enables genes to be transferred between species where there is no natural exchange.

Cloning of genes involves a number of basic procedures which are exemplified schematically by Figure 1-9. Initially, total cellular DNA is extracted from the host organism. The DNA is then treated with a restriction endonuclease that cleaves it at specific sites to give fragments of different sizes. These cleaved DNA fragments are inserted into a plasmid or bacteriophage vector that will replicate in the host bacterium into which the desired gene is to be cloned. This collection of DNA fragments inserted in a vector forms a "library", ideally containing in total a complete genome. The "Library" of cells can be screened for the gene of interest by plating cells on appropriate selective media.

Escherichia coli is the most commonly used organism for the bacterial expression of cloned genes, and the general principles for gene expression in this organism will apply when dealing with genes cloned into its original host. Synthesis of a functional protein depends upon transcription of the

Figure 1-9 A general diagram of gene cloning procedure



appropriate gene, transcript stability, efficient translationof the mRNA and post-translational processing and correct compartmentalization or extracelluar secretion of the nascent polypeptide. Stability of the foreign protein in the presence of the host bacterium's proteases may also become an important factor. Because the genes often reside on a plasmid, plasmid stability is also a critical factor. Assuming the gene has already been cloned, the desired level of activity of the gene product can be achieved by having the gene under the control of an appropriate promoter for RNA transcription. In many cases this may be the original promoter present on the gene. A higher synthesis rate could put an excessive demand on the cell and lead to a slower rate of growth.

Instability of foreign proteins coded for by cloned genes has been observed in *E. coli*, and the cellular location of the proteins was shown to affect stability. Talmadge and Gilbert (1982) measured the half-life of rat preproinsulin molecules synthesized in *E. coli* and found it to be 2 min for molecules in the cytoplasm, but over 20 min for those molecules in the periplasm. Various strategies have been developed to reduce the instability of foreign proteins caused by proteolytic degradation in *E. coli* (Carrier et al., 1983), including in-frame fusion of the foreign gene with a host gene. Typically, the host gene has been modified using the entire *E. coli lac*Z gene (Itakura et al, 1977) or a small fragment of the *lac*Z gene with convenient restriction enzyme sites for cloning or translational fusion in the correct reading frame (Guo et al, 1984; Stark, 1987). This approach would require selection of a suitable promoter to avoid overproduction of the fusion protein. Whether or not the fused protein produced active biological activity depended on the copy number, correct reading frames and other physiological factors.

The use of *E. coli* mutants lacking the *lon* protease (the major protease for degradation of abnormal proteins) is another

avenue to reduce post-translational degradation (Carrier et al., 1983). Simon et al., (1983) have developed a strategy based on the observation that during a T4 bacteriophage infection of *E. coli*, turnover of proteins continued, but there was a marked decrease in the degradation of abnormal proteins. They cloned the T4 protease inhibition gene (*pin*) responsible for reduced proteolysis and inserted it into an expression vector. They found that labile eukaryotic proteins encoded by genes cloned in *E. coli* were stabilized in cells in which the T4 *pin* gene was being simultaneously expressed.

Plasmids used as vectors for foreign genes usually replicate in the cytoplasm independently of the chromosomal replication cycle. Each plasmid replicates on average once every cell cycle. The number of plasmid copies within cells may differ from six to eight per chromosome for a low-copy-number plasmid to 50 per chromosome for the high-copy-number plasmid vector, which is commonly used for cloning experiments (Old and Primrose, 1985), to 100 or more per chromosome for small plasmids (Vieria and Messing, 1982).

Recombinant plasmids containing inserts of foreign DNA often tend to be eliminated or lost during the growth and division of microorganisms. The stable maintenance of a plasmid is dependent upon the host phenotype, the plasmid phenotype and copy number, and on the environmental conditions (Carrier et al., 1983; Old and Primrose, 1985). Low-copy-number plasmids have a lower probability of segregating equally between daughter cells (Chruchward et al., 1983). It has also been reported that the presence of a gene expressing a heterologous protein is liable to increase the incidence of plasmid-free cells (Imanaka et al, 1980). A conventional approach to maintenance of a recombinant plasmid is to use a selectable mechanism e.g. resistance to an antibiotic (Old and Primrose, 1985).

A technique that also may be used to stabilize plasmid born genes to be expressed in bacteria is to have on the plasmid of interest, a gene that codes for a function missing on the bacterial host chromosome. For example, inserting a recombinant plasmid with the *pyr*E gene into *E. coli* which enables the synthesis of uracil, thereby allowing the cell to grow in a uracil-free medium. Another possibility would be to introduce plasmids coding for bacteriocin production and immunity into these strains. This would have two advantages; it would enable the introduced bacteria to compete with the indigenous bacteria in the environment, and cells that lose the plasmid and therefore immunity to the bacteriocin would be killed by the bacteriocin-producing cells.

Another type of vector system which was often used is lambda vectors. The ever increasing use of bacteriophage lambda as a cloning vector both for genomic and cDNA has led to the development of techniques for the manipulation of lambdoid phage. The development of the phage λchromosome as a receptor for fragments of DNA was made by staggering cuts within specific DNA sequences (targets) to produce discrete fragments with short cohesive ends (Hedgpeth et al, 1972; Mertz and Davis, 1972; Bigger et al, 1973) such that the fragments of DNA could be spliced into the severed arms of a λ vector molecule. Two types of vectors, insertion and replacement, have been constructed. A λ genome with some nonessential DNA deleted, but retaining a single target for a restriction enzyme, is known as an insertion vector; replacement vectors are phages that retain two targets flanking a replaceable segment of DNA (Murray, 1983).

The efficiency of cloning being sufficiently high to allow the recovery of a complete library and rapid high density screening techniques make lambda vectors a very convenient system to use. Also the lytic growth of lambda recombinants overcomes the problem of secretion, thus making screening of

the library easier. A limitation of lambda vector systems is the maximum amount of foreign DNA that can be accommodated (approximately 23 kb) and the difficulty of direct sequencing from these vectors due to the difficulty of preparing large quantities of DNA and handling of large number of samples (Murray, 1983). In many cases, the capacity of insertion was not really critical. The large scale preparation of DNA from a large number of samples was solved by a method described by Manfioletti and Schneider (1988). This improvement allows the use of lambda vector for general cloning and directly through to sequencing without the necessity for laborious subcloning steps.

1-4-2 Genetic manipulation in cellulase genes

The cloning of cellulase genes, the determination of their nucleotide sequences, and the analysis of the amino acid sequences predicted from the nucleotide sequences have led to a greatly increased understanding of some of these important enzymes (Aubert, et al 1988). Several cellulase genes from different microorganism have been cloned and studied.

Over 20 apparently distinct DNA fragments coding for cellulase have been cloned from *C. thermocellum* into *E. coli*. (Millet et al, 1985; Schwarz et al, 1985; Romaniec et al, 1987). The clones containing the genes *cel*A, *cel*B, *cel*C, *cel*D and *cel*E have been extensively characterized and the DNA sequences of the coding and control regions determined (Beguin et al, 1985; Grepinet and Beguin, 1986; Joliff et al, 1986; Hall et al, 1988). Also, the xylanase gene *xyn*Z (Grepinet et al, 1988) and β-glucosidase gene (Kadam and Demain, 1988) have been cloned, characterized and the DNA sequence of *xyn*Z was determined.

Supernatant from cultures of *Ce. fimi* contained up to 10 components with cellulase activity (Langsford et al, 1984). The

clones containing the genes *cen*A, *cen*B, *cen*C and *cex* have been identified and sequenced (O'Neil et al, 1986; Wong et al, 1986; Owolabi et al, 1988; Moser et al, 1989).

Extracelluar β-glucanases and xylanase are produced by many members of the genus *Bacillus* (Priest, 1977). Cellulase cloning in this genus has had good progress. The endoglucanase and xylanase of *B. subtilis* have been cloned in *E. coli* or other species of *Bacillus* (Nakamure et al, 1987; MacKay et al, 1986; Robson and Chambliss, 1987; Koide et al, 1986; Hinchliffe, 1984; Kim and Pack, 1988; Bernier Jr. et al, 1983). Cloning from alkalophilic *Bacillus* sp. has also been done (Fukumori et al, 1989; Kim et al, 1987; Sharma et al, 1987; Fukumori et al, 1986; Sashihara et al, 1984). Cloning work from other cellulolytic *Bacillus* sp. was achieved (Park and Pack, 1986;Yang et al, 1988; Sandhu and Kennedy, 1986; Panbangred et al, 1983; Yang et al, 1989).

The cellulase gene from *F. succinogenes* has been cloned (Taylor et al, 1987; Gong et al, 1989; Sipat et al, 1987; Irvin and Teather, 1988; McGavin et al, 1989). A xylanase gene from *F. ruminicola* also has been cloned (Whitehead and Hespell, 1989).

Cloning studies of cellulase genes from ruminococci also have given exciting results. The endoglucanase gene, exoglucanase gene, xylanase gene and β-glucosidase gene of *R. albus* and *R. flavefaciens* have been cloned (Romaniec et al, 1989; Honda et al, 1988; Flint et al, 1989; Barros and Thomson, 1987; Ware et al, 1989; Howard and White, 1988; Kawai et al, 1987; Ohmiya et al, 1988; Huang et al, 1989). From the data of our laboratory (Unpublished data, Asmundson), more than 15 protein components have been found which purify with cellulase activity. Also at least 8 non-homologous clones have been identified (this study, Huang et al, 1989).

Other cellulase genes have been cloned from *Pseudomonas fluorescens* (Lejeune et al, 1986, 1988; Gilbert et al, 1987, 1988); yeast (Nebreda et al, 1986; Leclerc et al, 1986); *Erwinia chrysanthemi* (Brestic-Goachet et al, 1989; Boyer et al, 1987; Gijsegem et al, 1985; Kotoujansky et al, 1985; Aymeric et al, 1988); *Cryptococcus albidus* (Morosoli and Durand, 1988); *Thermomonospora fusca* (Hu and Wilson, 1988; Lin and Wilson, 1988; Ghangas and Wilson, 1988; Collmer and Wilson, 1988); *Streptomyces* sp. (Coppolecchia et al, 1987; Jaurin and Granstrom, 1989; Nakai et al, 1988); *Caldocellum saccharolyticum* (Love et al, 1987); *Agrobacterium* sp. (Wakarchuk et al, 1988) and ripe avocado fruit (Tucker et al, 1987).

Several successful cloning studies of fungi, particularly *T. reesei*, have been reported (Chen et al, 1987; Sim et al 1988; Teeri et al, 1983, 1987; Shoemaker et al, 1983; Arsdell et al, 1987; Saloheimo et al, 1988).

1-4-3 Cellulase gene structure

A given microorganism will produce a number of cellulases which differ in overall amino acid sequence but which may share short conserved sequences. Cellulases can be grouped into families on the basis of these conserved sequences (Table 1-3; Beguin, et al 1989; Henrissat et al, 1989). The location of conserved sequences within the enzymes of a family suggests that the family of genes arose by region or domain shuffling. In bacteria, the conserved sequences of a particular enzyme family occur in more than one genus and the conserved sequences occur in both Gram-positive and Gram-negative genera. Such sequence conservation implies an important function for the different enzymes. The sequences form discrete functional domains within the enzymes.

The enzyme family of *Cellulomonas fimi*, which includes both an endoglucanase (CenA) and an exoglucanase (Cex), contain two

Table 1-3 Families of cellulases

Family	Organism	Enzymes
A	*Cellulomonas fimi*	exoglucanase and endoglucanase A
	Microbispora bispora	endoglucanase A
	Clostridium thermocellum	xylanase Z
B	*Trichoderma reesei*	cellobiohydrolase I and II, and endoglucanase I and III
	Phanaerochaete chrysosporium	endoglucanase
C	*C. thermocellum*	endoglucanase E
	C. cellulolyticum	endoglucanase A
D	*C. thermocellum*	endoglucanase D
	Pseudomonas fluorescens var. *cellulosa*	endoglucanase
E	*Erwinia chrysanthemi*	endoglucanase Y
	Ce. uda	endoglucanase
F	*E. chrysanthemi*	endoglucanase Z
	C. acetobutylicum	endoglucanase
	Bacillus subtilis	endoglucanase
	alkalophilic *Bacillus* sp. strain N-4	endoglucanase A and B
	strain 1139	endoglucanase F

From: Ong et al, 1989.

conserved segments: (1) a sequence of about 100 amino acids (at the N-terminus ends of CenA but at the C-terminus ends of Cex) is 50% conserved, and (2) a conserved sequence of 20 amino acids containing only proline and threonine, the Pro-Thr box, which separates the former conserved region from the rest of the protein (Fig. 1-10)

CenA and Cex from *Ce. fimi* bind tightly to cellulose. Two major fragments were released from both CenA and Cex. One fragment has catalytic activity, but does not bind to cellulose. The second, smaller fragment binds to cellulose, but is catalytically inactive. The site of cleavage in both enzymes is at the C-terminus of the Pro-Thr box. Therefore, the conserved sequences in CenA and Cex constitute the cellulose-binding domains of these enzymes (Gilkes et al,1988).

Another enzyme family includes two endoglucanases, EGI and EGIII, and two cellobiohydrolases, CBHI and CBHII, from the fungus *Trichoderma reesei* (Knowles, et al, 1987). A 70% conserved sequence of about 35 amino acids occurs in all four

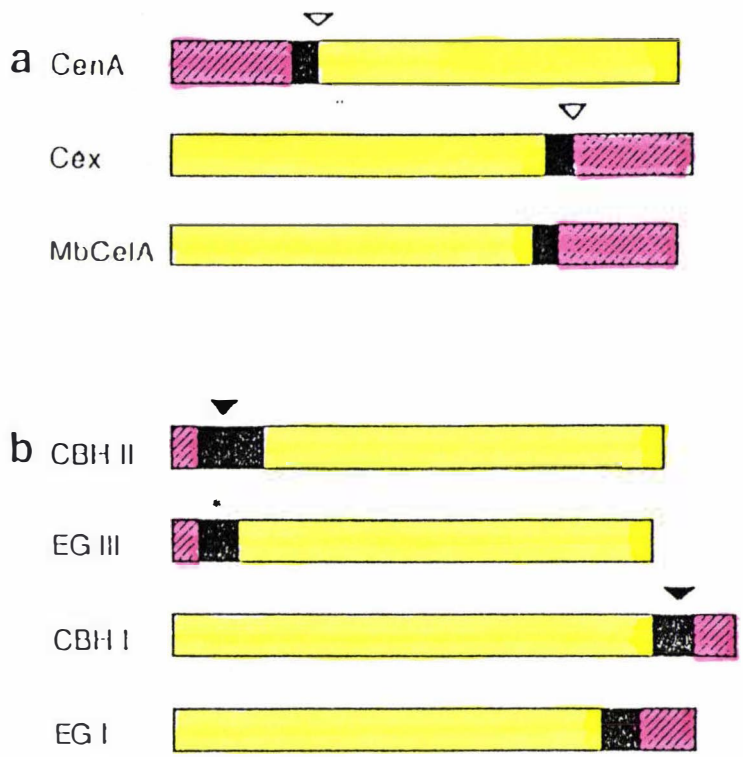
enzymes, at the N-terminal ends of EGIII and CBHII, but at the C-termini of EGI and CBHI. In all four, this conserved sequence is linked to the remainder of the polypeptide by a stretch of amino acids rich in proline and hydroxyamino acids (Fig. 1-10; Teeri, et al, 1987).

CBHI and CBHII of *T. reesei* also bind to cellulose. Papain cleaves about 80 amino acids from the N-terminus of CBHII and about 70 amino acids from the C-terminus of CBHI. The truncated proteins retain catalytic activity but no longer bind to cellulose. Thus the conserved sequences in CBHI and CBHII constitute the cellulose-binding domains of these enzymes (Tomme et al, 1988). The C-terminal fragment released from CBHI still binds to cellulose, as does a synthetic peptide of 36 amino acids which corresponds to the 70% conserved sequence at the C-terminus of the enzyme (Johnson, et al, 1989).

Cleavage of cellulases appears to occur naturally. The culture filtrate obtained after growth of *T. reesei* on cellulose contains a catalytically active form of EGIII which lacks the first 61 amino acids at the N-terminus of the native enzyme. Unlike the native enzyme, the truncated form does not bind to cellulose (Stahlberg, et al 1988). The protease-susceptible site in the *T. reesei* enzymes are in the proline and hydroxyamino acid rich sequences connecting the conserved sequences to the catalytic domains.

From both cases, *Cellulomonus* and *Trichoderma,* a short sequence of amino acids which is rich in proline and hydroxy-amino acids is significantly more protease-susceptible than the bulk of the protein. Small-angle X-ray scattering shows that the *T. reesei* CBHI and CBHII and *Ce. fimi* CenA enzymes are tadpole-shaped, whereas the catalytically active cores produced by proteolysis are ellipsoidal (Schmuck et al, 1986; Teeri et al, 1987; Abuja et al, 1988, ; Ong et al, 1989). The distal end of the 'tail' is the CBD (cellulose binding domain), the

Figure 1-10 The organization of functional domains in two cellulase families (from: Ong et al, 1989)



The organization of functional domains in two cellulase families (a) the C. fimi *(CenA and Cex) and* M. bispora *(MbCelA) cellulases; (b) the* T. reesei *cellulases. [ ], catalytic domain; [ ], proline- and hydroxyamino acid-rich linker region; [ ], cellulose binding domain. ▽ and ▼ are the primary sites of hydrolysis by* C. fimi *protease and papain, respectively; *indicates the amino terminus of the truncated form of EGIII found in culture filtrate. In both families, a cellulose-binding domain is joined to a catalytic domain by a sequence of amino acids which is proline- and hydroxyamino acid-rich, and protease-susceptible.*

proximal end is the proline-hydroxyamino acid rich linker. A sequence such as a Pro-Thr box probably assumes a very extended configuration within a protein (Radford, et al, 1989). The presence of such terminal domains in cellulases from *C. thermocellum* also has been reported, suggesting that these functionally related enzymes might have a similar architecture (Henrissat et al, 1989).

1-5 Aim of this study:

Enzyme-production costs represent 40-60% of overall costs of processes designed to saccharify cellulosic substrates and ferment the products (Ryu and Mandels, 1980). The poor enzyme productivity of fungi, a result of the relatively prolonged fermentation times required, might be solved by isolation of new organisms, mutants of existing species or cloning the cellulase genes in rapidly dividing bacteria which synthesize more efficient cellulases or exhibit enhanced enzyme productivity.

Other prospects of genetic manipulation of micro-organisms for use in programmes aimed at improving biomass conversion, fermentation resource or improvement in animal physiology has also led to interest in cloning cellulase genes.

The rumen is among the most abundant source of cellulolytic microorganisms. *Ruminococcus flavefaciens*, *R. albus*, *Butyrivibrio fibrisolvens* and *Bacteroides succinogenes* are the four major cellulolytic bacteria found in the rumen (Bryant 1973). *Ruminococcus flavefaciens* is one of the most important species of rumen bacteria and is capable of degrading plant cell walls (Pettipher and Latham 1979). Also, *R. flavefaciens* degrades crystalline cellulose more efficiently than *R. albus* (Bryant 1973).

Ruminococcus flavefaciens strain 186, which was isolated from a bovine rumen in New Zealand, expressed all three types of cellulase activity and showed effective cellulose degradation.

The aim of this study was to clone, sequence and characterize the cellulase genes and gene products from *Ruminococcus flavefaciens* strain 186, and in particular a 9 kb clone which expressed all three types of cellulase activities and showed effective cellulose degradation.

CHAPTER II

MATERIALS AND METHODS

Table of contents

2-1 Bacterial strains and vectors

Bacteria and vectors used are summarized in Table 2-1.

Table 2-1. Bacterial strains and vectors

Strains	Genotype or Phenotype	Source or Reference
Ruminococcus flavefaciens strain 186	Wild type	Asmundson and Kelly 1987
Escherichia coli		
BHB2600	*Sup*E, *Sup*F, *hsd*R, *hsd*M, r_k^-, m_k^+	Hohn and Murray 1977
POP-13	*Sup*E, *hsd*R, r_k^-, m_k^+, P2	Hohn and Murray 1977
JM109	*rec*1A, *hsd*R, *Sup*E, *end*A1, *gyr*A, *thi*, *rel*A1, λ^-, Δ(*lac-pro*A,B), F'*tra*D, *Pro*A,B, *lac*I^q, ΔM15	Messing et al. 1981
LE392	F^-, *hsd*R(r_k,m_k), *sup*E, *sup*F, *lac*Y, *gal*K, *gal*T, metB, *trp*R, λ^-	Murray 1977
MB406	*sup*E, *rec*B21, *rec*C22, *sbc*B, *hfl*A, *hfl*B, *hsd*R^-	Murray 1977
Q358	*hsd*R_k^-, *hsd*M_k^+, *sup*F, $\phi 80^r$	Karn et al. 1980
Q359	*hrd*R_k^-, *hsd*M_k^+, *sup*F, $\phi 80^r$, P2	Karn et al. 1980
Bacillus subtilis		
4712	*Aro*B2, *Trp*2Z, *Tyr*A1, *His*H2	NZDRI
Vectors		
λNM1149	λb358, *sr*Iλ3 , *imm*434, *sr*Iλ4 , *shn*dIIIλ6 , -*sr*Iλ5	Murray 1983
λGem11		Promega, USA
pUC19	Amp^r	Celeste et al. 1985
pGem3-blue	Amp^r	Promega, USA
pOU71	Amp^r	Larsen et al. 1984
pCK17	$Cm^r km^r$	Gasson et al. 1986
pBD64	$Cm^r km^r$	Gryczan et al. 1980
Recombinant clones		
λCM200- series	2 kb cellulase gene inserts	this study
λCM300- series	3 kb cellulase gene inserts	this study
λCM400- series	4 kb cellulase gene inserts	this study
λCM900- series	9 kb cellulase gene inserts	this study
λCMEE- series	*Eco*RI-*Eco*RI fragment of λCM903	this study
λCMEH- series	*Eco*RI-*Hin*dIII fragment of λCM903	this study
λCMHH- series	*Hin*dIII-*Hin*dIII fragment of λCM903	this study
λdGCM- series	deletions of subclones	this study

NZDRI, New Zealand Dairy Research Institute culture collection

2-2 Media and growth conditions

The *Ruminococcus* isolate was grown using the medium shown in Table 2-2 with rubber sealed test tubes or bottles under 100 % CO_2 at 37 °C (Kelly et al. 1987). *E. coli* was grown in Luria Broth (LB)(0.5 % w/v NaCl, 0.5 % w/v yeast extract and 1.0 % w/v tryptone) with 0.2 % (w/v) maltose and 20 mM $MgSO_4$ at 37 °C with shaking. When needed, LB medium was solidified with 1.5 % (w/v) agar (Davis, Christchurch, New Zealand). Overlay agar contained 0.7 % agar in LB medium. *Bacillus subtilis* was grown in LB broth with shaking at 37 °C. M9 medium (Table 2-3) was prepared for substrate specificity assay (Maniatis et al. 1982).

Table 2-2. Rumen fluid cellobiose medium for culture of *Ruminococcus* sp.

Na_2CO_3	0.4 g	Salt A*	7.5 ml
Trypticase	0.1 g	Salt B*	7.5 ml
Yeast extract	0.1 g	Resazurin	0.1 ml
Cellobiose	0.2 g	Rumen fluid	20 ml
(or cellulose sol'n20 ml)		Distill water	make up to 100ml

*, Salt A: K_2HPO_4 0.6 % w/v

Salt B: KH_2PO_4 0.6 % w/v
$(NH_4)_2SO_4$ 1.2 % w/v
NaCl 1.2 % w/v
$MgSO_4.7H_2O$ 0.25 % w/v
$CaCl_2.2H_2O$ 0.16 % w/v

Table 2-3. M9 minimal medium

per liter:

A:	Na_2HPO_4	6 g	B:	1 M $MgSO_4$	2 ml
	KH_2PO_4	3 g		1 M $CaCl_2$	0.1 ml
	NaCl	0.5 g			
	NH_4Cl	1 g			

Adjust pH of solution A to 7.4, autoclave, cool and then add solution B.

2-3 Enzymes and chemicals

Restriction enzymes, proteinase K, RNase A, T4 DNA ligase and Lambda packaging extracts were purchased from either Bethesda Research Laboratories (Gaithersburg, Md, USA) or

Promega Biotech (Madison, Wis, USA) and used under the conditions described by the suppliers. Agarose, mutanolysin, carboxymethyl-β-cellulose (CMC), 4-methylumbelliferyl-β-D-cellobioside (MUC), 4-methylumbelliferyl-β-D-glucoside (MUG), para-nitrophenyl-β-D-glucoside (pNPG), para-nitrophenyl-β-D-cellobioside (pNPC), Avicel (PH101), lichenan and other chemicals were purchased from Sigma Chemical Co. (St. Louis, Mo, USA), BDH Chemicals Ltd., (Poole, UK) and FMC Corporation (Philadelphia, Pa, USA).

2-4 Molecular biological methods

2-4-1 Gel electrophoresis

2-4-1-1 Agarose gel for DNA

DNA was mixed with gel-loading buffer (0.25 % bromophenol blue, 40 % sucrose in H_2O) and was analyzed by electrophoresis on 0.8 % to 1.5 % horizontal agarose slab gels using TBE buffer (89 mM TRIS-borate, 89 mM boric acid, 2 mM ethylenediaminetetraacetate (EDTA), pH 8.0). The gels were stained with 0.5 μg/ml ethidium bromide and DNA was detected with a short wavelength UV transilluminator. DNA fragments of Lambda phage digested with *Hin*d III were used as a molecular size standard.

2-4-1-2 Polyacrylamide gel for protein

Ten percent gels with a discontinuous SDS buffer system gel in a cooled vertical slab gel apparatus was used (Hames, D. 1982). Resolving gel was prepared by adding 10 ml of acrylamide-bisacrylamide stock solution (30%:0.8%, w/v), 3.75 ml of 3.0 M Tris.HCl pH 8.8, 0.3 ml of 10 % SDS, 1.5 ml of 1.5 % ammonium persulphate, 14.45 ml of water and 0.015 ml of TEMED and degassing before pouring into mounted glass plates (0.5 mm gap). After the resolving gel was set, stacking gel containing 2.5 ml of acrylamide-bisacrylamide stock solution, 5 ml of 0.5 M Tris.Hcl Ph 6.8, 0.2 ml of 10 % SDS, 1.0 ml of 1.5 % ammonium persulphate, 11.3 ml of water and 0.015 ml of TEMED was mixed,

degassed and added to the top of the gel with the comb inserted. Protein sample was mixed with gel-loading buffer (0.0625 M Tris.HCl pH 6.8, 2 % SDS, 5% 2-mercaptoethanol, 10 % glycerol and 0.002 % bromophenol blue) and boiled for 2 min or left unboiled before loading into gel. The running condition was set as constant voltage at 100 V until the dye front reached the end of the gel.

Three of the staining and blotting procedures used were as follows:

(1) Coomassie blue R250 staining: Dissolve Coomassie blue R250 (0.1%) in water:methanol:glacial acetic acid (5:5:2 by volume) and filter through a Whatman No.1 filter paper to remove any insoluble material before use. Each gel was placed in a plastic tray filled with stain solution and left for at least 4 hr at room temperature. After staining was complete, excess stain was removed, the gel was transferred to the destain solution (12.5 % isopropanol, 10 % acetic acid) and the destain solution was simply renewed as stain leached out of the gel over a period of about 24 hr at room temperature with gentle rocking.

(2) Silver staining: Bio-Rad Silver Stain kit was used according to instructions supplied by the company. The solution was added in the following order (per gel): Fixation solution (40% methanol, 10 % acetic acid (v/v)), 200 ml, 30 min; (10% ethanol, 5% acetic acid (v/v)), 200 ml, 15 min; (10% ethanol, 5% acetic acid (v/v)), 200 ml, 15 min; Oxidizer (potassium dichromat and nitric acid. store at 4 °C), 100 ml, 5 min; Deionized water, 200 ml, 5 min; Deionized water, 200 ml, 5 min; Silver reagent (silver nitrate, store at 4 °C), 100 ml, 20 min; Deionized water, 200 ml, 1 min; Developer (sodium carbonate and para-formaldehyde, store at 4 °C), 100 ml, until brown precipitate appeared; Stop solution (5% acetic acid (v/v)), 200 ml, 5 min.

(3) Zymogram staining: The probing gel was prepared by dissolving 1.5% (w/v) of agar in PC buffer (50 mM K_2HPO_4, 12 mM citric acid, pH 6.0) containing 0.5% CMC, MUC or MUG and pouring into the same size glass plate assembly (1.5 mm thick). After the electrophoresis gel was partially dried (37 °C for 15 min), the probing gel was put on the top and wrapped with plastic membrane. The sandwich was gently pressed with a roller to prevent bubbles remaining between the two gels. Then the sandwich was incubated at 30 °C for 1 hr (CMCase) or 2 min (MUCase and MUGase). The probing gel was removed and stained with the Congo red method (CMCase) or immediately detected using fluorescence under UV light (MUCase or MUGase).

2-4-1-3 Polyacrylamide gel for DNA sequencing

Six percent polyacrylamide gels were used for sequencing. The gel was prepared by mixing 50 g of urea, 15.8 ml 40 % acrylamide-bisacrylamide (38 g acrylamide and 2 g bisacrylamide to 100 ml H_2O), 10 ml of 10 x sequencing buffer (1.35 M Tris base, 25 mM EDTA and 450 mM boric acid) and making to 100 ml with H_2O. After vacuum degassing, 1.2 ml 5 % ammonium persulphate and 15 µl TEMED were added, mixed well gently and poured into assembled glass plates (BRL model S2) with the flat edge of the sharktooth comb inserted in position. After the gel was set, the comb was removed and reinserted with the sharktooth just inserting the gel by 1-2 mm.

2-4-1-4 Preparation and assembly of glass plates

To prevent formation of bubbles when pouring the gel, the plates were cleaned thoroughly with a scourer, washed well, and wiped with ethanol. The inside of the top plate was siliconised with Sigmacote, and the plates were wiped and polished dry with paper towels. After checking that the plates were spotlessly clean, the spacers were placed on the sides of the back plate and the top plate was placed on top of the back plate. The

plates were wrapped with "Sleek" tape ensuring no gaps. The sides were then clamped and a line was drawn from the top of the top plate as a guide for the comb.

2-4-2 Preparation of DNA

2-4-2-1 Chromosomal DNA

Ruminococcus cells were cultured and lysed by Mutanolysin treatment (Asmundson and Kelly, 1987) as described below. *R. flavefaciens* strain 186 was cultured in 1200 ml of rumen fluid cellobiose broth described above. Cells were harvested by centrifugation at 10,000 rpm for 20 min. After washing twice with TP7 buffer (10 mM K_2HPO_4, 0.1 % Triton X-100, pH 7.0), the cells were resuspended in 10 ml of 10 mM K_2HPO_4 and 1.2 ml Mutanolysin (5000 U/ml), 600 µl proteinase K (5 mg/ml) and 300 µl RNase A (5 mg/ml) added. The mixture was incubated at 55 °C for 5 min, then diluted with 50 ml of 10 mM K_2HPO_4, pH 7.0 containing 25 mM EDTA to complete lysis. The volume of mixture was measured and 1.028 x V grams of CsCl and 0.11 x V of 5 mg/ml ethidium bromide solution added. After the CsCl had dissolved, the cell debris were removed by centrifugation at 15,000 rpm for 15 min. The preparation was spun in an ultra-centrifuge (VTi50 rotor, Beckman, USA) for 18 hr at 25 °C. The DNA band was harvested and the ethidium bromide was removed by extraction with CsCl saturated isopropanol. CsCl was removed by dialyzing four times against 1000 ml TE buffer (10 mM Tris.HCl, 1 mM EDTA, pH 8.0). The DNA was extracted with phenol, phenol-chloroform and chloroform and was precipitated with 2.5 x V of cold ethanol. Finally, The precipitated DNA was washed with 70 % ethanol, dried under vacuum and dissolved in 1 ml of distilled water (Asmundson and Kelly 1987).

2-4-2-2 Phage DNA

The λNM1149 vector DNA was isolated by a method modified from Maniatis et al.(1982). The phage was concentrated by

polyethylene glycol 6000 precipitation and dialyzed against SM buffer (0.1 M NaCl, 10 mM $MgSO_4$, 10 mM TRIS-Cl, 0.1 % gelatine, pH 7.4). The concentrated phage solution was lysed by 1 % sodium dodecyl sulphate in TE buffer and incubated at 80 °C for 10 min. After phenol-chloroform extraction, the vector DNA was precipitated with cold ethanol, dried and resuspended in TE buffer containing 50 µg/ml of RNase A.

λGem- serial phage DNA was isolated by the method modified from Manfioletti and Schneider (1988).

(1) DE52 exchanger preparation: The weighed ion exchanger was stirred into 15 volume of 0.5 N HCl and left for at least 30 min. After filtering off the supernatant liquor in a funnel and washing with H_2O until effluent reached pH 4 or above, another 15 volumes of 0.5 N NaOH were added and stirred for a further 30 min. The slurry was washed with distilled water in a funnel until the filtered effluent was near neutral, 0.8 volume of TE buffer (pH 8.0) containing 0.01 % sodium azide was added and the slurry stored at 4 °C. Before using in phage clearing steps, the treated ion exchanger was washed in LB broth twice and resuspended in 0.4 volume of LB.

(2) Phage lysate preparation: Semiconfluent lysis plates were prepared by adding phage into bacteria soft agar plates with a multiple of infection (M.O.I) at 0.5. After about 5 hr incubation at 37 °C, 5 ml per plate of SM buffer were added and gently shaken in a platform shaker overnight in a cold room (4 °C). The supernatant was collected after centrifugation at 5000 x g for 10 min. To 10 ml of phage lysate, gelatine was added to a final concentration of 50 µg/ml. An equal volume of DE52 was added and the mixture was shaken by inverting the tube for 10 min at room temperature. The DE52 resin was spun down and the supernatant was filtered through glass fibre. The phage lysate was then ready for DNA extraction.

(3) DNA extraction: One ml of cleared phage lysate was transferred into 1.5 ml Eppendorf tube and a final concentration of 20 mM EDTA and 50 μg/ml proteinase K was added. After incubation at 45 ℃ for 15 min, 20 μl of cetyl-trimethyl ammonium bromide (CTAB) from stock solution (5 % CTAB in 0.5 M NaCl) was added and further incubated at 68 ℃ for 3 min. The pellet was collected with centrifugation at 8000 x g for 10 min, and resuspended in 0.2 ml of 1.2 M NaCl. The DNA was then precipitated with 2.5 x V of ethanol and dissolved in H_2O.

2-4-2-3 Plasmid DNA

Vector DNA from pUC19, pOU71, pGem3-blue and pCK17 were prepared by the method modified from Monstein and Geijer (1986). Bacterial strains carrying vectors were cultured overnight in 100 ml LB broth with antibiotics at 37 ℃. The cells were centrifuged at 5000 x g for 10 min and resuspended in 10 ml TELT buffer (50 mM Tris.HCl pH 7.5, 62.5 mM EDTA, 0.4 % Triton X-100 and 2.5 M LiCl). After adding 1 ml of freshly prepared lysozyme (10 mg/ml in H_2O), the cells were heated in boiling water for 2 min and immediately centrifuged at 10,000 x g for 30 min at room temperature. The supernatant was mixed with an equal volume of isopropanol. The pellet, collected by centrifugation at 10,000 x g at 4 ℃ for 10 min, was dissolved in 400 μl TE buffer. The solution was loaded in a Sephadex G-50 column, prepared by equilibration with TE buffer in a 1 ml syringe, and spun down at 2000 x g for 10 min. The eluent was transferred to a new sterile Eppendorf tube. Ten μl of RNase A (20 μg/μl) was added and the tube was incubated at 37 ℃ for 30 min. SDS and proteinase K were added to a final concentration of 0.5 % and 1 mg/ml, respectively, and the mixture was incubated at 37 ℃ for a further 1 hr. After phenol-chloroform extraction, the DNA was precipitated with 2.5 x V ethanol at -20 ℃ for 3 hr, dried under vacuum and resuspended in H_2O.

Vector pBD64 DNA was prepared by the method of Hardy

(1985). The culture was grown in 1.5 ml LB broth to $O.D._{600}$ of about 1.0. (Ideally the culture should be in the logarithmic phase of growth.) The cells were harvested by centrifuging for 1 min in a microcentrifuge tube. After resuspending the pellet in 100 µl of solution I (50 mM Glucose, 10 mM EDTA, 25 mM Tris.HCl pH 8.0) containing 2 mg/ml lysozyme, the mixture was left on ice for 15 min. Then 200 µl of solution II (0.2 M NaOH, 1% w/v SDS) was added and the solution left on ice for 5 min. Another 150 µl of solution III (3M sodium acetate pH 4.8) was added to the mixture which was vortexed and left for 1 hr on ice. After centrifuging for 5 min at 4 °C at full speed in a microcentrifuge, the supernatant was removed into a fresh tube, 1 ml of ethanol was added and it was left at -70 °C for 10 min. The mixture was centrifuge again, the supernatant was carefully removed and the pellet was resuspended in 100 µl of solution IV (0.1 M sodium acetate, 50 mM Tris.HCl pH 8.0). 250 µl of ethanol was added to precipitate the DNA and the mixture was centrifuged for a further 10 min at 4 °C in a microcentrifuge. The pellet was washed with 400 µl of ethanol and centrifuged for 2 min at 4 °C in a microcentrifuge. The pellet was dried in a desiccator and then resuspended in 30 µl of TE buffer.

2-4-3 Extraction of DNA from agarose gel

2-4-3-1 Electroelution method

The dialysis tubing was prepared according to the method of Maniatis et al 1982. The dialysis tubing was cut into 15 cm lengths and boiled in a large volume of 2% (w/v) $NaHCO_3$ and 1 mM Na_2EDTA for 10 min. The tubing was rinsed thoroughly in distilled water, submerged in distilled water and autoclaved for 10 min at 15 lb (72 kPa) pressure. After cooling, the tubing was stored at 4 °C.

DNA fragments were electroeluted by cutting out the band from agarose gel, transferring the gel into a dialysis bag with as little TE buffer as possible and the DNA was run out of the

gel with an electric field of 100 V (5 V/cm) for 1 hr followed by field reversal for 2 min. The DNA solution was carefully collected, and purified by extracting twice with phenol-chloroform and then precipitated with ethanol.

2-4-3-2 DEAE membrane binding method

The gel was sliced in front of the required band and a piece of NA45 membrane (Schleicher and Schuell), which was just large enough to fit, was placed into the slit. After running the gel at 100 mA until the band has completely stacked up on the membrane, the membrane was carefully removed, rinsed in TE buffer (pH8.0) and placed into an Eppendorf tube with 0.4 ml of 1 M NaCl and 0.05 M arginine (free base). The tube was incubated at 70 ℃ for as long as necessary to elute the DNA. One hour was usually sufficient for fragments around one to two kb. After elution, the membrane was removed and ethanol was used to precipitate DNA as usual. The pellet was resuspended in water and was then ready for use.

2-4-4 Restriction digestion

Enzyme digestion was carried out with type II restriction endonucleases. One volume of 10 x restriction enzyme buffer was mixed with eight volume of DNA (about 1 μg) in H_2O. Another one volume of restriction enzyme (1 U) was added into the mixture and incubated at 37 ℃ or the optimal temperature for restriction enzyme reaction for 1 hr. One tenth of the volume was loaded into agarose gel to check for reaction completion. When satisfied of completion, the reaction mixture was heated at 65 ℃ for 10 min. *Eco*RI, *Bam*HI, *Hin*dIII, *Xho*I, *Sma*I, *Cla*I, *Pst*I, *Kpn*I, *Sal*I and *Bgl*II were the restriction enzymes used. The conditions used were those specified by the supplier.

2-4-4-1 Dephosphorylation of DNA

The method used for dephosphorylation was described by Maniatis et al. (1982). After completed digestion with restriction enzyme, the DNA was extracted once with phenol-chloroform and precipitated with ethanol. The pellet was dissolved in a minimum volume of 10 mM Tris.HCl pH 8.0 (usually 5 μl). One volume of 10 x CIP buffer (0.5 M Tris.HCl pH 9.0, 10 mM $MgCl_2$, 1 mM $ZnCl_2$, and 10 mM spermidine) was added to nine volume of H_2O. Two additions of calf intestinal alkaline phosphatase (CIP, 0.01 U per 1 pmole of 5′ ends of DNA) were made, one initially and again at 30 min. The reaction was stopped by adding 40 μl of H_2O, 10 μl 10 x STE buffer (100 Mm Tris.Hcl Ph 8.0, 1 M NaCl and 10 mM EDTA) and 5 μl of 10 % SDS and heating at 68 °C for 15 min. The reaction mixture was extracted twice with phenol-chloroform and passed through a spun column of Sephadex G-50 equilibrated in TE. The DNA was precipitated with ethanol and dissolved in H_2O.

2-4-5 Construction of recombinant phage library

The chromosomal DNA of *R. flavefaciens* 186 was partially digested with *Eco*RI and separated on a 0.8 % preparative gel. DNA fragments of 2-10 kb were electroeluted and the DNA solution was carefully collected. The DNA was purified with two phenol-chloroform extractions and precipitated with ethanol. The isolated DNA fragments were ligated to *Eco*RI digested vector λNM1149 DNA. One volume of 10 x ligation buffer (0.66 M Tris.HCl pH 7.5, 50 mM $MgCl_2$, 50 mM dithiothreitol and 10 mM ATP) was added into nine volume of DNA solution. One unit of T4 DNA ligase per μg DNA was then added. The reaction mixture was allowed to react at 4 °C overnight. The ligated DNA was packaged into lambda packaging extract according to the conditions of the supplier. Up to 5 μg of DNA in 10 μl was added to thawed packaging extract. After incubation at 22 °C for 2 hour, 0.5 ml of SM buffer and 25 μl of chloroform were added

and mixed well. The packaged lambda DNA was transfected into *E. coli* host POP-13 and amplified by harvesting the lysate and reinfection into the same host (Maniatis et al, 1982).

2-4-6 Screening of recombinant for cellulase activities

The Congo red dye method was used to screen for CMCase (endoglucanase) activity (Teather and Wood 1982). Plaques of recombinant phages on a LB plate were overlaid with soft agar containing 0.5 % CMC in PC buffer (50 mM K_2PO_4, 12 mM citric acid, pH 6.5). After incubation at 37°C until the plaque appeared (usually 5-8 hr), the plate was flooded with 0.5 % Congo red solution (0.5 % w/v Congo red in water) and left to stain for 15 min. The stained plate was then washed with 1 M NaCl.

The method of methylumbelliferone method (fluorescent method) was used for screening of MUCase and MUGase activities (exoglucanase and β-glucosidase, respectively). The genomic library in lambda was diluted, preinfected on POP-13 and plated onto LB plate containing 0.1% of MUC or MUG. The plate was incubated at 37°C until plaques were formed and the fluorescence immediately detected under long-wave UV light in the dark.

2-4-7 Hybridization

The DNA of λCM903 (a 9 kb clone described later) right-hand side (λCMEH1, encoding CMCase) and left-hand side (λCMEH2, encoding MUCase) were prepared as probes. Hybridizations were carried out using Chemiprobe (Organic Ltd, Yavne, Israel), a kit system for non-radioactive tagging of DNA probes and visualization of the tagged probes after hybridizations. The kit was used according to the instructions of the supplier as described below.

2-4-7-1 Denatured salmon sperm DNA

DNA was dissolved in H_2O at a concentration of 10 mg/ml and sheared by passing it several times through an 18 gauge hypodermic needle. The DNA was boiled for 10 min and chilled on ice quickly. This DNA could be stored at -20 ℃ for 1 year. Before using the DNA for hybridization, the DNA was heated for 5 min in a boiling water bath and then chilled immediately on ice.

2-4-7-2 Preparation of probe

DNA was diluted in distilled H_2O to a final concentration of 5 μg/ml. The DNA was denatured by boiling for 5 min and then chilled quickly on ice. 0.5 volume of Modified solution A (basically sodium bisulfite) and 0.125 volume of Modified solution B (basically methylhydroxylamine) were added to DNA solution and the mixture was incubated overnight at 25 ℃. The modified DNA probe could be stored at -20 ℃ for up to 1 year.

2-4-7-3 Dot hybridization

DNA was boiled for 5 min and immediately chilled on ice. After diluting the DNA in 10 x SSC (1.5 M NaCl, 0.15 M Sodium citrate) and spotting onto a nitrocellulose membrane, the membrane was baked at 80 C for 2 hr under vacuum. The membrane was then prehybridized with 100 μg/ml of denatured salmon sperm DNA in 6 x SSC, 0.1 % SDS and 5 x Denhardt's solution (1 % Ficoll 400, 1 % polyvinylpyrolidone (MW:360,000), and 1 % BSA) for 2 hr. The same solution as used for prehybridization was used for hybridization with denatured modified probe and hybridization was carried out overnight at 68 ℃. The membrane was then washed using a washing procedure as follows: (1) 2 x SSC, 0.1 % SDS at room temperature for 5 min. (2) 2 x SSC, 0.1 % SDS at 68 C for 30 min. (3) 0.1 x SSC, 0.1 % SDS at room temperature for 5 min.

2-4-7-4 Immunological visualization

Preparation of blocking solution by adding 3 g of powdered milk to 10 ml of diluent (25 mM NaCl, 50 mM Tris.HCl pH 7.5, 1 mM EDTA, and 0.3% Tween 20) and mixed well until completely dissolved. The hybridized membrane was transferred to a plastic bag and 50 μl of blocking solution per each cm^2 of membrane was added. The bag was sealed and incubated for 1 hr at 25 °C with gently agitation. With a final dilution of 1:250, the mouse monoclonal antibodies (antimodified DNA) was added and the membrane was incubated for another 1 hr at 25 °C with gentle shaking. The membrane was washed three times for 5 min with washing solution (0.5 M NaCl and 0.5% Brij 35T), and transferred to another bag. The alkaline phosphatase antimouse immunoglobulin conjugate solution diluted 1:250 with blocking solution was added. After incubation at 25 °C for 1 hr with shaking, the membrane was washed three times for 10 min each in 0.5 M NaCl and 0.3 % Brij 35T. A freshly prepared substrate solution (1.5 mg nitro blue tetrazolium, 13.5 mg glucose, 0.2 mg 5-bromo-4-chloro-3-indolylphosphate in 100 mM Tris.HCl, pH 9.5, 100 mM NaCl and 5 mM $MgCl_2$) was then added and the membrane incubated at 37 °C for 30 min to 3 hr under dim lighting until a blue color, indicating hybridization, appeared. The membrane was then placed between two sheets of 3MM paper to dry out.

2-4-8 Subcloning of insert fragment

2-4-8-1 Transformation

(1) Competent cell preparation: A single colony of JM109 was inoculated into 5 ml Psi broth (5 g yeast extract, 20 g tryptone, 4 g $MgSO_4$, 0.75 g KCl and KOH adjust to pH 7.6) and incubated at 37 °C with shaking until A_{600}= 0.3. To 100 ml of psi broth, 1 ml of JM109 culture was added and shaking continued at 37 °C until A_{600}= 0.5. The cells were chilled in an ice-water bath for 15 min and centrifuged at 5000 x g for 5 min at 4 °C.

The pellet was resuspended in 30 ml chilled Tƒβ1 buffer (30 mM KOAc, 50 mM $MnCl_2$, 100 mM RbCl, 10 mM $CaCl_2$ and 15 % glycerol (stored at 0 °C)) and left in the ice-water bath for a further 15 min. The cells were centrifuged again at 2500 x g for 5 min at 4 °C. The competent cells were gently resuspended into 4 ml of chilled Tƒβ2 buffer (10 mM NaMOPS (sodium salt of 3-N-morpholinopanesulfonic acid) pH 7.0, 75 mM $CaCl_2$, 10 mM RbCl, and 15% glycerol, (stored at 0 °C)). The competent cells were then transferred into a thin-walled glass tube (0.2 ml per tube) and stored at -70 °C for up to 3 month without decrease in efficiency (Maniatis, 1982).

(2) Transformation: Up to 50 ng of DNA in 10 µl was added to 0.2 ml thawed competent cells and the mixture was left on an ice-water bath for 20 min. After heating at 42 °C for 2 min, the competent cells were immediately chilled on ice-water for 2 min. Four volumes of psi broth were added and the cells incubated at 37 °C with gentle shaking for 1 hr. The transformant was poured onto a LB plate with LB soft agar containing antibiotics.

2-4-8-2 *Bacillus* transformation (protoplast transformation)

A few colonies from a fresh plate of the strain was inoculated to 50 ml of LB broth and grown with vigorous aeration to an $O.D._{600}$ of 0.4. The cells were centrifuged for 10 min at 5000 r.p.m. in the Sorvall GSA rotor. The cells were resuspended in 5 ml of SMMLBP buffer (0.5 M sucrose, 20 mM sodium maleate, 20 mM $MgCl_2$, pH 6.5, 10 g tryptone, 5 g yeast extract, 5 g NaCl, 40 g polyvinylpyrolidone) and 260 µl of 20 mg/ml lysozyme (freshly prepared in SMMLB and filter-sterilised) were added. During incubation at 37 °C, samples were taken at intervals and examined by phase-contrast microscopy to check how many protoplasts had been formed. When 99% or more of the cells had been converted to protoplasts, 1-5 µg of DNA dissolved in 10 µl of TE and 1.5 ml of 40% (w/v) polyethylene glycol were

added. The mixture was mixed gently and left for 2 min at room temperature. Five ml of SMMLBP were added, mixed gently and centrifuged at 3000 r.p.m. for 10 min in the Sorval SS-34 rotor. The protoplasts were resuspended in 1 ml SMMLBP and incubated at 37 °C for 90 min with gentle shaking. An aliquot (0.1 ml) was spread onto protoplast regeneration medium, (DMP) composed of 0.3 M disodium succinate, 8 % (w/v) polyvinyl-pyrolidone (MW=40,000) pH 7.3; 1 % (w/v)agar, 0.5 % (w/v) casamino acids, 0.5 % (w/v) yeast extract; 11 mM KH_2PO_4, 20 mM K_2HPO_4; 0.5 % (w/v) glucose; 20 mM $MgCl_2$ (autoclaved separately and then mixed together) and containing the appropriate antibiotics (1 mg/ml Kanamycin, 5 μg/ml chloramphenicol (Hardy, 1985)).

2-4-8-3 Transfection

A colony from a fresh culture (LE392, MB406, Q358 or Q359) was inoculated into 2 ml LB broth and incubated with shaking at 37 °C for about 2-3 hr (A_{600}=0.6). The culture was chilled in an ice-water bath for 15 min, and the cells were then centrifuged at 5000 x g for 5 min at 4 °C. After resuspending in 1 ml of CMT solution (50 mM $CaCl_2$, 10 mM $MgSO_4$, and 50 μg/ml of thymidine, freshly prepared and kept at -20 °C for 5 min before use), the cells were left in the ice-water bath for a further 5 min and centrifuged again. The competent cells were gently resuspended in 150 μl of CMTT buffer (CMT solution + 0.1 M Tris.HCl pH 8.0) containing 100 ng/ml of phage DNA or ligated recombinant DNA and transferred into a thin-walled glass tube in an ice-water bath. The mixture was kept for 45 min in the ice-water bath, then heated at 42 °C for 2 min. The entire heat-shock treated mixture was immediately plated onto a LB plate with LB soft agar.

2-4-9 DNA deletion

DNA deletion for preparation of sequencing and gene

location was carried out by a commercial package, Erase-a-Base (Promega, USA). DNA (5 μg) was digested with *Eco*RI or *Hin*dIII and then extracted once with phenol-chloroform, precipitated with ethanol and resuspended in 1 x Klenow buffer (20 mM Tris-HCl pH 8.0, 100 mM $MgCl_2$). A mixture of 40 μM α-phosphorothioate dNTP, 1 mM DTT (dithiothreitol) and Klenow DNA polymerase to 50 U/ml was added and incubated at 37 °C for 10 min. The reaction was heated at 70 °C for 10 min to inactivate the Klenow DNA polymerase. Then the DNA was extracted again with phenol-chloroform, precipitated with ethanol and suspended in H_2O. The second restriction enzyme digestion (*Hin*dIII or *Eco*RI) was then performed. After digestion, phenol-chloroform extraction and precipitation, the DNA was dissolved in 60 μl of 1 x *Exo*III buffer (66 mM Tris.HCl pH 8.0, 0.66 mM $MgCl_2$). Meanwhile 25 small tubes with 7.5 μl of S1 mix solution (172 μl H_2O, 27 μl 7.4 x S1 buffer (0.3 M potassium acetate pH 4.6, 2.5 M NaCl, 10 mM $ZnSO_4$ and 50% glycerol) and 60 U S1 nuclease) each were prepared and left on ice. The DNA tube was warmed to 37 °C in a water bath. There was a 30 sec lag for the reaction to begin. After adding 150 U of *Exo*III and mixing as rapidly as possible, 2.5 μl samples were removed at 30 sec intervals into the S1 tubes on ice, pipetting up and down briefly to mix. When all samples had been taken, tubes were placed at room temperature for 30 min. One μl of S1 stop buffer (0.3 M Tris base, 0.05M EDTA) was added and tubes were heated at 70 °C for 10 min to inactivate the S1 nuclease. To determine the extent of digestion, 2 μl (about 40 ng DNA) samples were removed from each time point for analysis on an agarose gel . When satisfied with the DNA lengths achieved, the time point tubes were transfer to 37 °C, and 1 μl Klenow mix (30 μl 1 x Klenow buffer, 3 U Klenow DNA polymerase) was added to each and incubated at 37 °C for 3 min. Then 1 μl of the dNTP mix was added and further incubated for 5 min at 37 °C. The reaction mixture was moved to room temperature and 40 μl ligase mix 890 μl H_2O, 100 μl 10 x ligase buffer (500 mM Tris.HCl pH 7.6, 100 mM $MgCl_2$, 10 mM ATP), 10 μl 100 mM DTT and 5 U T4 DNA ligase) was added to each

sample, mixed well and left to ligate at 4 ℃ overnight. The ligated DNA was then ready for transfection.

2-4-10 DNA sequencing analysis

Two μg of the lambda DNA was used for each sequencing reactions. The DNA was mixed with 100 ng of oligonucleotide primer (SP6 or T7, Promega) in a final volume of 10 μl of water. The sample was boiled for 4 min and then snap frozen in dry ice/ethanol. The sample was then left to defrost on ice. In a separate tube, always on ice, the indicated volumes of reagents from the Sequenase Kit (Tabor and Richardson, 1987)(USB, Cleaveland, Ohio, USA) were added in the following order: 2 μl DTT (0.1 M), 0.5 μl ^{35}S-α-dATP (sp.act. >1000 Ci/mmole (DuPont)), 0.7 μl of the diluted Sequenase labelling mix (1.5 μM dNTP-dATP), 2 μl of sequencing buffer (200 mM Tris.HCl pH 7.5, 100 mM $MgCl_2$ and 250 mM NaCl) and finally 2.5 μl of diluted Sequenase (1:8 dilution of the stock enzyme in TE buffer). This mixture was then transferred into the tube containing the lambda DNA and primer, treated as above. The reaction was incubated at room temperature for 5 min. After this time 4 μl were removed and added to 2 μl of each termination mix (80 μM dNTP, 8 μM of the relevant ddNTP, and 50 mM NaCl) which was prewarmed at 37 ℃ for one min. The reaction was stopped after 2-3 min by adding 4 μl of Stop solution (50 % formamide, 20 mM EDTA, 0.05 % bromophenol-blue and 0.05 % xylene-cyanol). After heating at 90 ℃ for 3 min, 2 μl of each reaction were loaded on a sequencing gel.

2-4-10-1 Assembly and analysis of sequence data

Compilation of the sequencing data, analysis of the compiled sequence to determine ORF's and comparison of sequence data to the known cellulase genes was performed using the programs available through the University of Winsconsin Genetics Computer Group (UWGCG) sequence analysis software

package installed on a Vax 750 computer (Devereux et al. 1984).

2-5 Biochemical methods

2-5-1 Partial purification of cellulase enzymes

2-5-1-1 Preparation of crude enzyme extract

To partially purify the cellulase, the recombinant phage lysate was mixed with an equal volume of pebble-milled paper (PMP) solution (0.5 % pebble-milled Whatman No.1 filter paper in water) on ice water for 10 min. The PMP was harvested by centrifugation at 4 °C and resuspended in 1/4 original volume of H_2O. The solution was incubated at 42 °C for 30 min with shaking and then was centrifuged again for 2 min at 40 °C. The supernatant containing cellulase protein was withdrawn and concentrated by freeze-drying.

2-5-1-2 Gel filtration chromatography

Gel filtration chromatography was carried out on a series of Bio-Gel P polyacrylamide gels: P-2, P-6, P-60, P-100 and P-300. The column sizes chosen were 3.8 cm I.D. x 25 cm, 3.8 cm I.D. x 60 cm and 2.5 cm I.D. x 75 cm. The beads were prepared by rehydrating in water with stirring overnight, using at least 22 ml of liquid for each gram of dry gel. The buffer chosen was water. The water used was purified by a Millipore Milli-Q system which included an Organex-Q cartridge and a 0.22 µm filtration step (Millipore Corp.). The chromatography system was composed of a LKB peristaltic pump (12000 varioperpex), an ISCO model UA-5 monitor and recorder and a Gilson microfractionator. Crude enzyme preparations (10 ml) were eluted through P-series columns with water. Fractions (1 ml/tube) were collected and assayed for protein and enzyme activity. Fractions with cellulase activities were freeze-dried to concentrate the proteins.

2-5-2 Biochemical assay of cellulase activities

2-5-2-1 Assay of endoglucanase activity

2-5-2-1-1 Preparation of CMC substrate

To prevent contamination by small sugar (monomers and oligomers), carboxylmethylcellulose was purified as follows : ten percent CMC (w/v) in water was mixed well with two volumes of ethanol, and left standing for 30 min. After centrifuging at 5000 rpm for 10 min, the pellet was dried under vacuum. A 2% (w/v) purified CMC in water was tubed, autoclaved and stored at 4 °C.

Endoglucanase activity was analyzed by measuring the release of reducing sugar using the p-hydroxybenzoic acid hydrazide (HBAH) method (Koziol 1981). The sample was added to 1 % CMC solution in PC buffer pH 6.5 and incubated at 30 °C for 3 hour. One volume of mixture was added to two volumes of HBAH reagent (1 % w/v of HBAH dissolved in 100 mM HCl, 50 mM Na_3citrate, 100 mM Na_2SO_3, and 1.5 M NaOH). The reaction mixture was boiled for 5 min or cooked with full-powered microwave for 20 second then immediately cooled in an ice-water bath. A_{410} was measured by spectrophotometer within one hour and reducing sugar determined from a standard curve.

2-5-2-2 Assay of exoglucanase and β-glucosidase activities

Exoglucanase activity and β-glucosidase activity were detected qualitatively by scoring for fluorescence under UV light or quantitatively either using an Aminco spectrofluorometer (SPF-500 Aminco, MA, USA) (Millet et al. 1985) or by measuring the release of p-nitrophenol from pNPC or pNPG, respectively (Ohmiya et al. 1985). One fifth volume of sample was added into four fifth volume of 0.1 % MUC or MUG in PC buffer pH 6.5. After incubation for 15 min at 30 °C, the fluorescence was read with emission set at 450 nm and excitation set at 370 nm. Another method was performed by

adding 10 mM pNPC or pNPG to sample and incubating for 3 hour at 30 °C. Absorbance was read immediately at A_{410}. One unit(U) of activity was defined as that releasing 1 μmol of glucose equivalent of reducing sugar, 1 μmol of p-nitrophenol or 1 μmol methylumbelliferone per minute per milligram of protein.

2-5-2-3 Protein concentration

Protein concentration was measured by the Coomassie blue binding assay (Sedmak and Grossberg 1977) using bovine serum albumin as standard. Equal volumes of sample or serial diluted sample and 0.06 % Coomassie blue G-250 in 3 % perchloric acid (twice filtered with Whatman No.1 filter) was mixed together and immediately read by spectrophotometer at A_{640}.

2-5-2-4 Substrate specificity assay

Activity against 1 % CMC, 1 % lichenan, 1 % cellobiose and 1 % xylan was tested by measuring reducing sugar as described above using appropriate standard. The assay for degradation of Avicel (PH101) and PMP were tested as follows. After preinfection at 37 °C, the phage-infected bacteria solution was added into M9 medium (Table 2-3) containing 0.1 % Avicel or 0.5 % PMP as sole carbohydrate and incubated at 37 °C for 1 week. One volume of mixture was withdrawn, centrifuged and the supernatant was transferred to another tube. Two volumes of HBAH reagent were then added to the supernatant and the release of reducing sugar measured as described above.

2-5-2-5 Inhibition and activation assay

Partially purified endoglucanase, exoglucanase and β-glucosidase were assayed for their optimal pH, temperature and for inhibition by surfactants and other inhibitors. The optimal pH was determined by mixing 1 μg of partially purified enzyme with 1 ml of PC buffer over a range of pH (4.0-7.2). The

optimal temperature was assayed with the same conditions as above at temperatures from 15 °C - 45 °C. Solutions of glucose, cellobiose, SDS, Triton X-100, Tween 80, iodoacetate, EDTA and EGTA were mixed with enzyme to determine their effect on its activity.

2-5-3 Cellular localization of enzymes in *E. coli*

The pGem3-blue(CMC$^+$) transformants were cultured in LB broth containing 50 μg/ml ampicillin for 3 h at 37 °C and were harvested by centrifugation at 5,000 x g for 10 min. The supernatants were mixed with 0.5% PMP solution as described above and the partially purified cellulase protein was obtained. Intracellular protein was released based on the method of Ames et al. (1984). After centrifugation, the pellet was resuspended in 1/100 volume of chloroform, vortexed well and left at room temperature for a further 15 min. One tenth initial volume of the 0.01 M phosphate buffer pH 6.0 was added and centrifuged again at 5,000 x g for 20 min. The protein released in supernatant was from the periplasmic space of the cell. The pellet was extracted with phenol-chloroform once and the aqueous layer was harvested as the sample of intracellular protein.

INDEX

CHAPTER III

CONSTRUCTION OF GENOMIC LIBRARY AND THE IDENTIFICATION OF CELLULASE GENES FROM *RUMINOCOCCUS FLAVEFACIENS*

Table of contents

3-1 Introduction

Following the development of methods for cloning foreign DNA into plasmids, researchers began to investigate means of preparing gene banks representative of the entire genomic DNA of both prokaryotes and eukaryotes. Plasmid vectors have the theoretical capacity to accommodate fragments of unlimited size, but the transformation efficiency of plasmids containing greater than 10 kb of insert DNA is so low that their use in constructing genomic libraries is not practical.

Lambda vectors were constructed to overcome this problem, being able to incorporate up to 20 kb of insert DNA into the region non-essential for replication and infection. An additional advantage of cloning larger inserts is the ease of identification of overlapping fragments from the original genome, as well as the isolation of intact genes and operons.

Two examples of lambda vectors are λL47 and λNM1149. λL47 is a replacement vector which is able to insert up to 18.9 kb of foreign DNA. This vector contains three unique restriction enzyme sites (*Bam*HI, *Eco*RI and *Hin*dIII) for the insertion by ligation of *Sau*3A digested DNA and allows recombinant selection by their Spi$^-$ (sensitive to P2 interference) phenotype, in which the central fragment contains the *red* and *gam* genes (the product of which inactivates the potent exonuclease coded by the host *rec*BC system) is replaced by the insert, but the *chi* gene, in the remaining arms, stimulates growth of recombinant derivatives on a Rec$^+$ host (Fig. 3-1).

λNM1149 is an insertion vector which allows the cloning of fragments generated by either *Eco*RI or *Hin*dIII directly into the *c*I gene of a λ*imm*434 hybrid phage. Space for the foreign DNA is made available largely by the *b*538 deletion that eliminates the phage attachment site and the *int* and *xis* genes. Because of its enlarged genomes (up to 11 kb insertion) and its

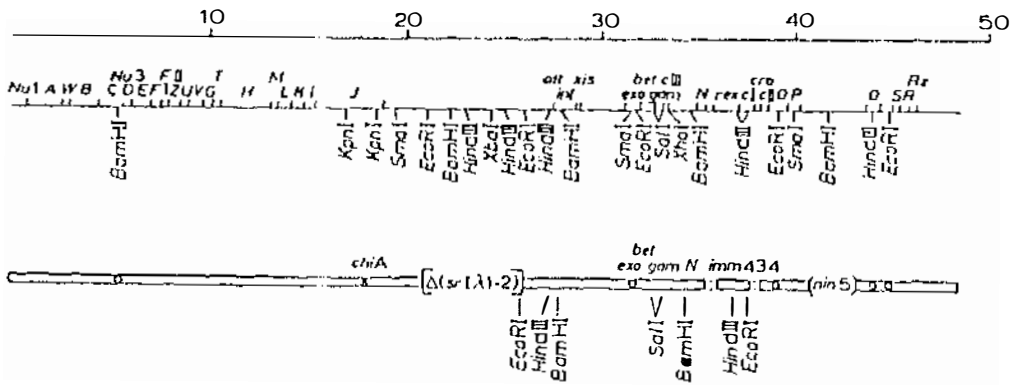


Figure 3-1 λL47 bacteriophage vector

clear plaque morphology, this vector has the advantage of easy and efficient use as recombinant phages. The recombinants can be selected on a Hfl⁻ host strain, which is defective in a proteimxthat normally antagonizes to phage-repressor synthesis and on which the native vector establishes repression so effectively as to reduce the efficiency of plating (Fig. 3-2).

After the genomic library was constructed using DNA from *R. flavefaciens*, the next step was to screen the library for clones containing the cellulase genes, identifying and isolating the positive recombinants.

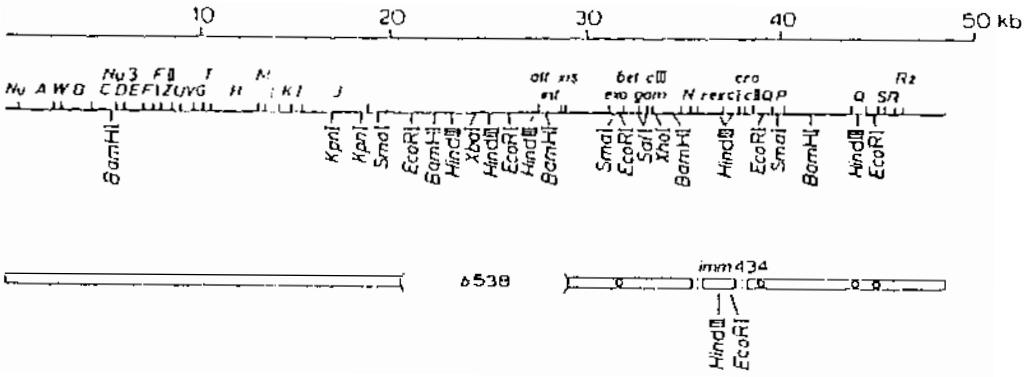


Figure 3-2 λNM1149 bacteriophage vector

The screening strategy was to first screen for the CMCase (endoglucanase) gene using the Congo red method. This method is based on the strong interaction between polysaccharides which contain contiguous β-(1,4)-linked D-glucopyranosyl units and Congo red (Teather and Wood, 1982). The other two enzymes involved in the degradation (exo-glucanase and β-glucosidase) were tested using the methylumbelliferone method. This method used the specific substrate (methylumbelliferyl-cellobiose (MUC) or methylumbelliferyl-pyroglucoside (MUG)) which linked the methylumbelliferone (fluorescent compound under UV) and cellobiose or glucose with a β-(1,4)-glycosydic bond.

In this section, the construction of a genomic library which represented the whole genome of *R. flavefaciens* 186, and the identification of recombinants encoding cellulase genes are described.

3-2 Results

3-2-1 Construction of genomic library

3-2-1-1 Preparation of partially digested chromosomal DNA

As the first step in the construction of a genomic library, the chromosomal DNA of *R. flavefaciens* 186 was prepared as described in Materials and Methods. One μg of chromosomal DNA was digested with restriction enzyme *Sau*3A or *Eco*RI (1U) at 37 C and 200 ng were withdrawn at 15 min intervals. The partially digested DNA was immediately heated at 65 C for 10 min to stop the reaction. After cooling in ice and mixing with gel-loading buffer, the DNA was electrophoresed on a 0.8% preparative gel. The incubation time for partial digestion was determined as 30 min and 45 min, and large scale preparation of partially digested DNA was performed by digesting fifty μg chromosomal DNA under the same conditions. After gel electrophoresis, the regions which corresponded to the 2-10 kb size fragments were cut out with a scalpel under long-wave UV light and the DNA was electroeluted and purified

by phenol-chloroform extraction procedures as described in Materials and Methods.

3-2-1-2 Ligation with vector DNA and packaging with lambda packaging extract

The first vector used was λL47. The vector DNA was purchased from Promega (Australia). One μg vector DNA was digested with the restriction enzyme *Bam*HI (1U) for 1 hr at 37 C and immediately heated at 65 C for 10 min to inactivate the enzyme. After phenol-chloroform extraction and ethanol precipitation, the digested DNA was dephosphorylated (refer to Materials and Methods).

The *Sau*3A treated chromosomal DNA was ligated with vector DNA using T4 DNA ligase and packaged into lambda packaging extract (Promega). *E. coli* Q359 was transfected with the packaged phage by plating onto double-layered LB plates. After incubation at 37 C overnight, recombinant clones were found on the plate at a titer of 3×10^2 plaque forming unit (PFU)/ml. Using the same conditions with *E. coli* genomic DNA instead of ruminococcal DNA gave a high titer recombinant library (3×10^5 PFU/ml). It was thus decided that λL47 was not an effective vector for cloning ruminococcal DNA.

DNA was prepared from another vector λNM1149 as described in Materials and Methods. Ten μg DNA was digested with *Eco*RI (10U) for 1 hr at 37 C and then heated at 65 C for 10 min. After phenol-chloroform extraction and ethanol precipitation, the digested DNA was dephosphorylated.

Purified chromosomal fragments were ligated with *Eco*RI digested vector DNA in a ratio of 1:2 (w/w) using T4 DNA ligase and packaged using lambda packaging extract (Promega, Australia). The packaged lambda was transfected into *E. coli* host POP-13 by plating in a double-layered LB plate and the resulting

phage lysate (2.1×10^4 PFU/ml) was stored at 4 C or -70 C as a gene library. The efficiency of packaging was slightly less than that achieved with an *E. coli* genomic library (5×10^4 PFU/ml), prepared in an identical manner. The amplification of the genomic library was performed by the method described in Materials and Methods. A phage titre of 6.1×10^{10} PFU/ml of the amplified library was achieved.

3-2-2 The genomic library of *R. flavefaciens*

Packaging into the bacteriophage heads imposed a size selection on the recombinant inserts. The average insert size, using λNM1149 was 5.5 kb. As the actual size of the *R. flavefaciens* 186 chromosome was not known, an arbitrary value of 3,000 kb was set in order to determine the number of λNM1149 clones required for the desired probability of a given sequence being present in the library. This was based on the known chromosome size of *Streptococcus lactis* (3,000 kb). The desired probability (0.99) of a unique DNA sequence being represented in the genomic library, was calculated using the equation of Clark and Carbon (1979):

$$N = \frac{\ln (1-P)}{\ln (1-1/n)}$$

where P = the probability of a given sequence being in a genomic library (0.99); N = the number of clones required in the clone bank; and n = the size of the genome relative to the average size of the cloned fragment.

$$N = \frac{\ln (1-0.99)}{\ln (1-1/545.45)} = \frac{-4.605}{-0.0018} = 2556$$

Therefore, based on the above formula, the 2.1×10^4 recombinant clones achieved should be representative of the entire genome of *R. flavefaciens* 186. If the genome is as large as *E. coli* (4,100 kb), 4812 clones would be required to maintain 99% probability of a given sequence being present.

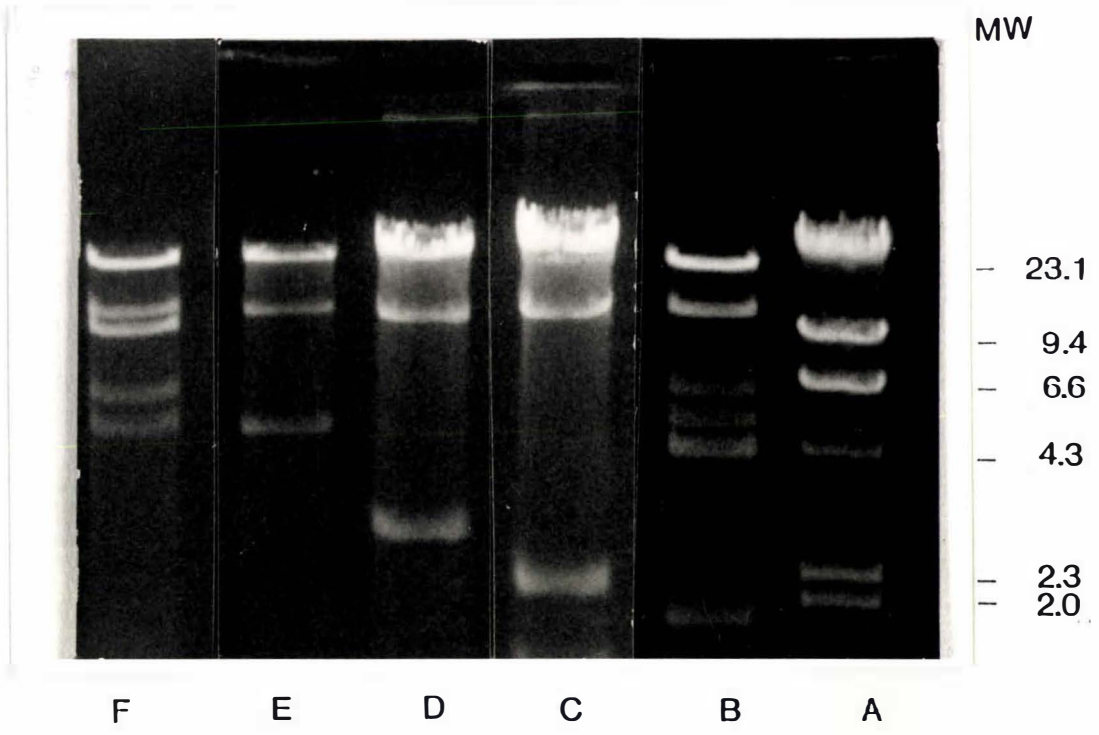
Several of the recombinant clones from the genomic library were chosen at random, and the phage DNA was isolated, digested with EcoRI, and analyzed by gel electrophoresis to determine the percentage and size of insert DNA. In all cases, a fragment of 21.6 kb, corresponding to the size of λNM1149 was present, and in 4 out of 12 cases additional fragments, presumably insert DNA, were also produced (Fig. 3-3). The total length of insert DNA in the 12 recombinant clones was measured, and found to vary from 2 to 9 kb, the mean being 5.5 kb. The total length of DNA being packaged appears to be in the range of 23.6 to 30.9 kb. This corresponds well with the reported values for size selectivity in lambda packaging in vitro (Hohn, 1979).

3-2-3 Screening the library for cellulase genes expressed in *E. coli*

In order to establish whether the *R. flavefaciens* insert DNA in the recombinant clones was able to show expression of cellulases in *E. coli*, a direct screening method (Congo red method) was used. The genomic library was serially diluted, preinfected *E. coli* POP-13 and plated onto a double-layered LB plate in which the top thin layer contained 0.5% carboxylmethylcellulose (CMC). The plates were incubated at 37 C until plaques appeared. Then the Congo red staining method was applied as described in Materials and Methods.

In about 2500 recombinant clones, 26 clones were detected with yellow-halo zone surrounding the plaque (Fig. 3-4), showing the degradation of CMC. After more than 3 rounds of single plaque purification, the clones were considered pure.

Figure 3-3 Gel electrophoresis of recombinants



A: Lambda DNA (HindIII digestion)

B: Mixture culture from library (EcoRI digestion)

C: λCM201 (EcoRI digestion)

D: λCM301 (EcoRI digestion)

E: λCM401 (EcoRI digestion)

F: λCM903 (EcoRI digestion)

Figure 3-4 Congo red method for detection of CMCase clones

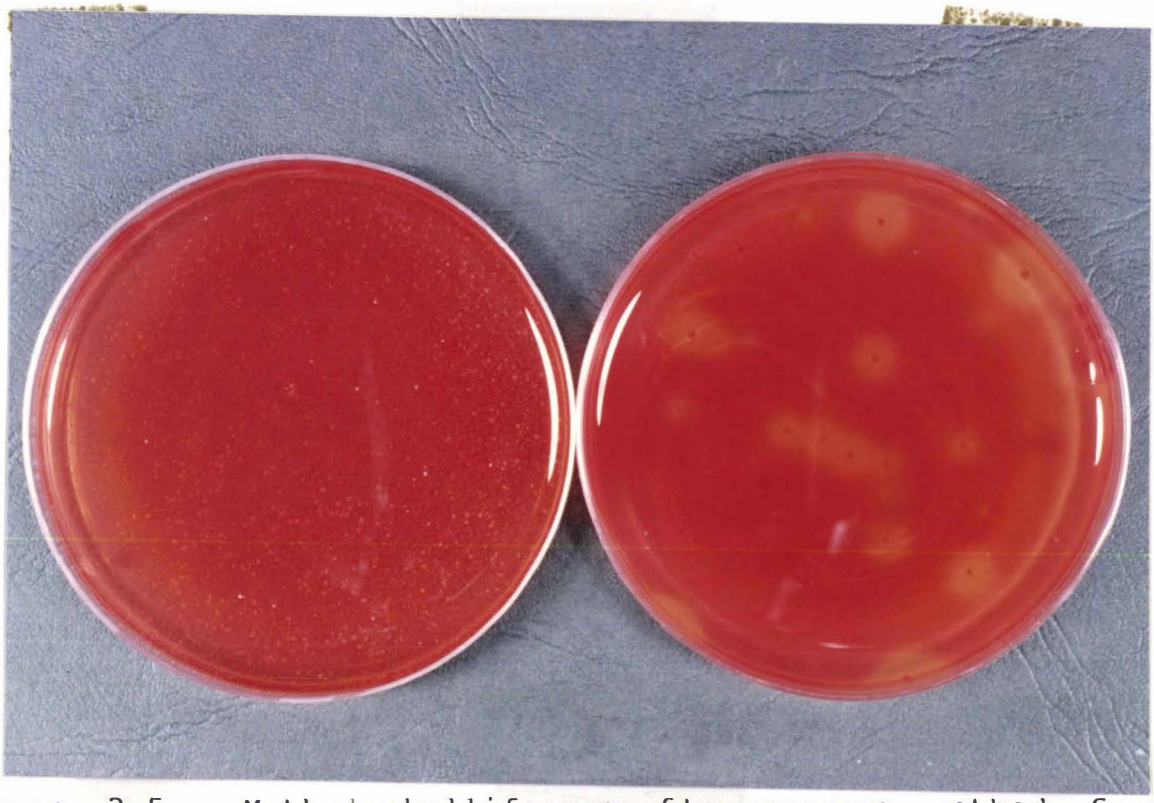
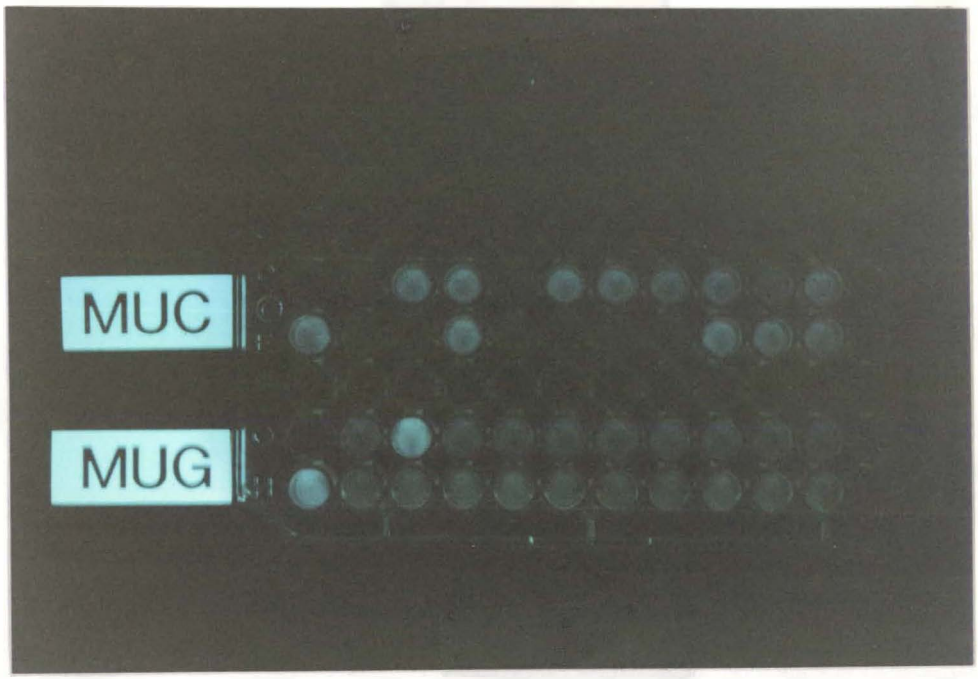


Figure 3-5 Methylumbelliferone fluorescence method for detection of MUCase and MUGase clones



Phage DNA was extracted from these clones, digested with *Eco*RI, and analyzed by agarose gel electrophoresis as described in Materials and Methods. All clones shared common DNA fragments of 21.6 kb, and the four unique digestion patterns observed are shown (Fig. 3-3).

To detect other cellulase enzyme activities, the genomic library was also screened for exo-glucanase and β-glucosidase using the methylumbelliferone method as described in Materials and Methods. The genomic library was diluted, preinfected into *E. coli* POP-13 and plated onto LB plates containing 1 mM of methylumbelliferylcellobiose (MUC) or methylumbelliferyl-glucoside (MUG). The plate was incubated at 37 C until plaques were formed and the fluorescence was observed immediately under long-wave length UV light in the dark. Large regions of fluorescence were detected from both MUC or MUG containing plates. It showed the positive reaction of degradation of MUC or MUG into methylumbelliferone(MU), confirmed as exo-glucanase or β-glucosidase activities. Because of the rapid diffusion of MU, the intensities of positive clones in these plate assay was rather low.

Another approach was based on the knowledge that the combined action of endo- and exo-glucanase degrades cellulose and thus the genes of endo- and exo-glucanase may be closely linked. The 26 CMCase positive clones were further assayed for MUCase and MUGase activities (Fig. 3-5). The lysate of 26 clones was added to 0.1% MUC or MUG in PC buffer (pH 6.0) and incubated at 37 C for 30 min. Four MUCase positive and two MUGase clones were detected from the 26 CMCase positive clones (Table 3-1).

The phage DNA of the 26 clones was isolated and the size of insertion was determined by digesting with *Eco*RI and separating on an agarose gel. The clones fell into size classes of 2, 3, 4 and 9 kb and these were named λCM200 series, λCM300

Table 3-1. Cellulase activities of λCM clones.

Clones	Specific activities (mU)				
	CMCase	MUCase	pNPCase	MUGase	pNPGase
λCM201	183	-	-	-	-
λCM301	176	-	-	-	-
λCM401	733	-	-	-	-
λCM404	438	1.6	1.0	-	-
λCM407	455	3.6	2.6	-	-
λCM901	245	1.4	1.3	-	-
λCM902	336	-	-	0.15	0.18
λCM903	337	3.1	3.0	0.40	0.35

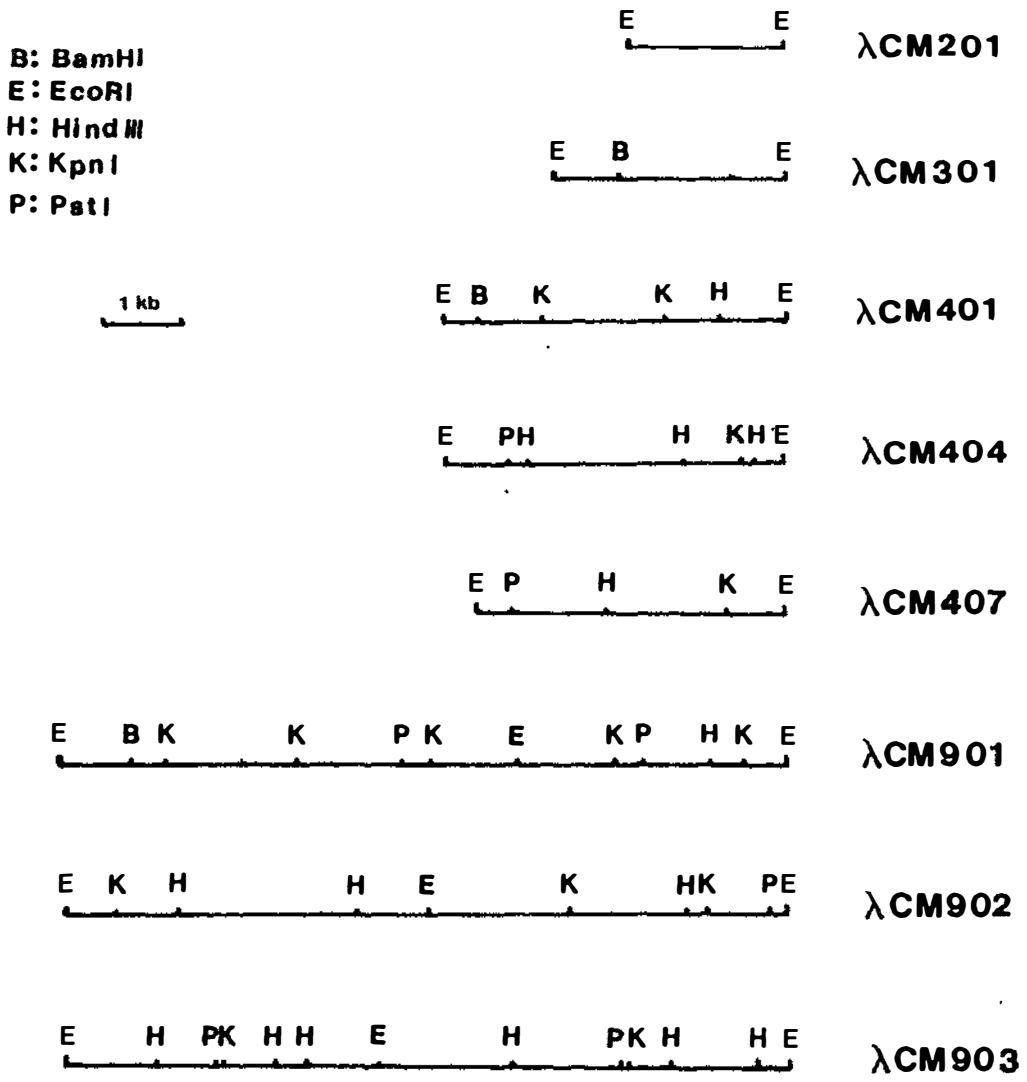
series, λCM400 series and λCM900 series, respectively.

3-2-4 Biochemical assay of cellulase activities

To prepare crude enzyme extract, the recombinant phage lysate was treated with PMP solution, centrifuged and resuspended in H_2O. After brief treatment at 42 C to release the cellulase enzyme, PMP was removed by centrifugation and the enzyme solution was concentrated by freeze-drying (re: Materials and Methods).

Each sample was then assayed for cellulase activity by incubation at 37 C for 3 h in the presence of either 0.5% CMC, 0.5 mM MUC, 1 mM pNPC, 0.5 mM MUG or 1 mM pNPG. The results are shown in Table 3-1. λCM200 and λCM300 series clones showed no enzyme activity other than CMCase. In λCM400 series clones, λCM404 and λCM407 showed both CMCase and MUCase activities. In λCM900 series, three clones showed different patterns of enzyme expression: λCM901 had both CMCase and MUCase activities; λCM902 showed CMCase and MUGase activities; and λCM903 expressed all three types of cellulase enzyme (CMCase, MUCase and MUGase).

Figure 3-6 Restriction maps of eight representative λCM-clones



3-2-5 The restriction mapping of recombinant clones

Restriction enzyme mapping provided a means to locate the gene and also helped to determine which fragment to subclone. After complete or partial digestion with *Eco*RI, the fragment of inserted foreign DNA was separated with vector DNA on an electrophoresis gel and purified. A set of restriction enzymes (*Hin*dIII, *Bam*HI, *Bgl*II, *Cla*I, *Sma*I, *Kpn*I, *Pst*I and *Eco*RI) were used to digest the DNA using the condition described in Materials and Methods. The restriction maps of eight representative clones is shown in Fig. 3-6. λCM201 had no internal restriction sites for the restriction enzymes used while λCM301 had an internal *Pst*I site. The clones λCM401, λCM404 and λCM407 showed different restriction patterns as did the clones of λCM901, λCM902 and λCM903.

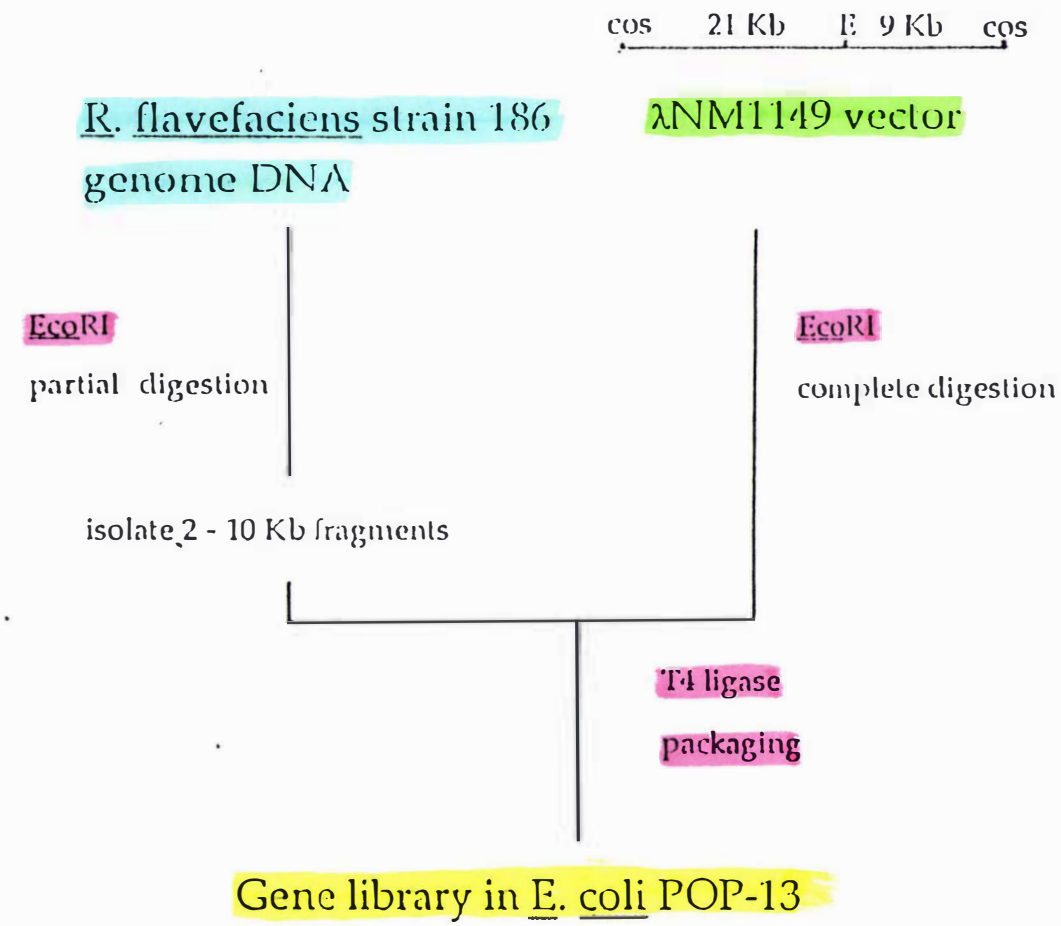
3-3 Discussion

3-3-1 Library construction

The production of a genomic library is an important step in the investigation and analysis of the genetic structure of *R. flavefaciens*. An understanding of its genetics will help to enable alteration of regulation of gene expression which may lead to the increase of the production of cellulase enzyme or understanding of the biochemical characteristics of cellulase.

Based on the benefit of easy storage, large insert size capacity and easy detection of the expression of genomic recombinants by lysis of the host, the lambda vectors (replace vector (λL47) and insertion vector (λNM1149)) were chosen to construct a gene library of *R. flavefaciens* 186. The protocol for construction of the genomic library has been summarized in Fig. 3-7. The library constructed with λL47 has a titer of 3 x 10^2 PFU/ml which was not large enough to represent the entire genome of the bacteria. The reason for the low titer in this library is still unknown. The library constructed with λNM1149

Figure 3-7 A summary in construction of genomic library of *R. flavefaciens 186*



has an average insert size of 5.5 kb and a PFU of 6.1 x 10^{10}/ml in the amplified library. According to the calculation shown above, the number of clones in this genomic library should be representative of the entire *R. flavefaciens* genome.

3-3-2 High copy number of cellulase genes and cellulase families

Twenty six cellulase positive clones were isolated after screening 2500 clones from the library. A frequency of more than 1 % cellulase-positive clones indicated the high number of cellulase gene copies in the *R. flavefaciens* 186 genome. A high level of cellulase positive clones was also reported in the cloning of *R. albus* 8 in λDASH with 0.33% (CMC^+), 0.37% (MUC^+) and 2.78% (OBR-HEC^+, ostazin brilliant red-hydroxyethyl cellulose) (Howard and White 1988). So far more than 20 different peptides have been found in *C. thermocellum* (Hazlewood et al, 1988). Different peptides or genes which encode similar cellulase activities were also found in *R. albus* (Kawai et al, 1987; Honda et al, 1988; Ohmiya et al, 1988; Romaniec et al, 1989; Flint et al, 1989; Ware et al, 1989), *R. flavefaciens* (Howard and White 1988), *Ce. fimi* (Beguin and Eisen, 1978; Whittle et al, 1982; Gilkes et al, 1984; Owolabi et al; 1988; Akhtar et al, 1988; Moser et al, 1989) and *Bacillus* spp. (Sashihara et al, 1984; Hinchliffe, 1984; Park and Pack, 1986; Fukumori et al, 1986; Koide et al, 1986; Robson and Chambliss, 1987; Sharma et al, 1987; Kim et al, 1987; Fukumori et al, 1989). Data from this work (Huang et al, 1989) showed eight different representative CMCase clones with different restriction enzyme patterns. This is in agreement with the proposition that there is more than one gene coding for the degradation of cellulose. Published data has also indicated cellulase is coded as gene families (Millet et al, 1985; Hazlewood et al, 1988).

Howard and White (1988) reported the detection of multiple enzymatic activities in single clones from *R. albus*. A similar

result with λCM903 which encoded CMC^+, MUC^+ and β-glu^+ activities has been shown. The three types of cellulase gene are closely linked within a 9 kb DNA fragment. Subcloning data showed that the β-glucosidase gene might be located at the central *Eco*RI site since digestion of λCM903 with *Eco*RI resulted in the disappearance of this enzyme activity. In addition, no β-glucosidase activity was detected at either end of the λCM903 insert. The cellulase genes might be expressed as a unit under a type of operon control.

3-4 Short summary

A functional genomic library of *Ruminococcus flavefaciens* 186 was constructed using the lambda vector λNM1149. It constituted 2.1×10^4 recombinant clones which was large enough to represent the entire genome of this bacterium.

From the library, 26 CMC^+ clones had been identified after screening from about 2500 recombinant clones. These CMC^+ clones were divided into four groups according to the size of their inserts (2, 3, 4 and 9 kb). The partial restriction maps of these clones have been achieved. Eight representative clones showed different mapping patterns. One of the 9 kb insert clones (λCM903) was selected for further study based on its expression of three type of cellulase activities.

Chapter IV

CHARACTERIZATION OF *R. FLAVEFACIENS* CELLULASE GENES AND THEIR GENE PRODUCTS

Table of contents

4-1 Introduction:

The λCM903 clone had been shown to express all three types of cellulase enzyme within a 9 kb fragment. The next step was to localize the genes within this fragment by subcloning. The first chosen vector for subcloning was pUC19, which is a high copy number vector in *E. coli*. Also different function vectors, such as the temperature controlled low copy number vector pOU71 and the shuttle vector pCK17, were tried.

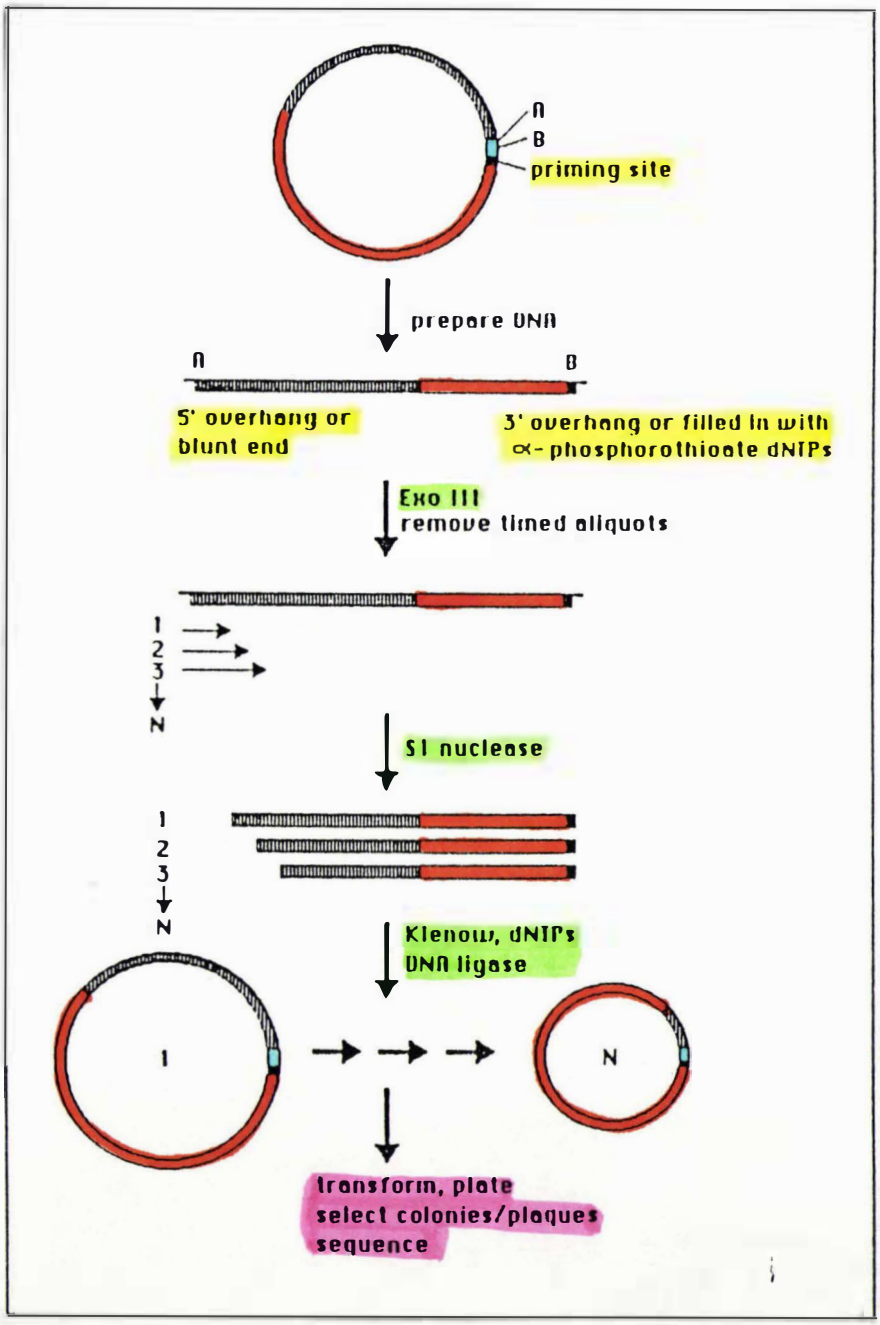
Restriction analysis of the fragments revealed too few sites to enable accurate identification of a smaller region carrying the cellulase genes and therefore deletion was chosen as the method to physically map the exact location of the genes. At the same time, the deletion clones could be used directly in sequencing experiments.

The deletion method chosen was the *Exo*III method. This enzyme catalyses the stepwise 3′-> 5′ removal of 5′ mononucleotides from double stranded DNA carrying a 3′-OH end or blunt end, while leaving a 3′ protruding end or α-phosphorothioate filled end intact (Fig. 4-1; Weiss, 1976). The uniform rate of digestion by the enzyme allowed deletions to be made of predetermined lengths by removing timed aliquots from the reaction.

The gene products of λCM903 in *E. coli* were characterized through a partial purification using gel filtration chromatography with several columns. The Bio-Gmx P-series resin was chosen based on the size fractions and to avoid the binding factor with which cellulase binds to cellulose based resin.

The aim of this section of study was to localize the cellulase genes, partially purify and characterize the gene products and prepare the subclones by deletion for sequencing work.

Figure 4-1 A diagram of *Exo*III deletion method



4-2 Results:

4-2-1 Localization of cellulase genes in λCM903

4-2-1-1 The difficulties of subcloning the cellulase genes of λCM903 into plasmid vectors

Since the cloning of cellulase genes into λNM1149 showed the expression of the cellulase activities, it was important to specify the location of the cellulase genes within these inserts to enable further study of their structures. The insert of λCM903 was first subcloned into the plasmid vector pUC19 after *Eco*RI digestion and transformation into *E. coli* JM109 (re: Materials and Methods). Various subclones could be isolated from a 100 μg/ml ampicillin selective medium. However, after 26 or 22 generations with or without antibiotic selection, respectively, all the subclones lost the insert plasmid.

The insert of the λCM903 was then subcloned into other types of plasmid vectors: pGem3-blue, pOU71 and pCK17 after *Eco*RI digestion and these were transformed into *E. coli* JM109 (re: Materials and Methods). After incubation for 15, 27 and 30 generations respectively without antibiotic selection, all recombinant clones of pGem3-blue, pOU71 and pCK17 also lost their plasmids. The results of adding antibiotic selection during incubation showed similar results as those with non-selection. Similarly, the subclones of pCK17 and pBD64 which were transformed into *B. subtilis* lost their recombinant plasmids and their cellulase activities after cultivation.

4-2-1-2 Localization of CMCase and MUCase genes in λCM903

Due to the difficulty of subcloning the *R. flavefaciens* DNA into plasmid vectors, the subcloning of λCM903 insert fragments and the localization of the cellulase genes were done in lambda vector. λCM903 DNA was digested with *Eco*RI and *Hin*dIII, and the resulting *Hin*dIII-*Hin*dIII or *Eco*RI-*Hin*dIII fragments were subcloned to the λGem-11 vector through transfection (re:

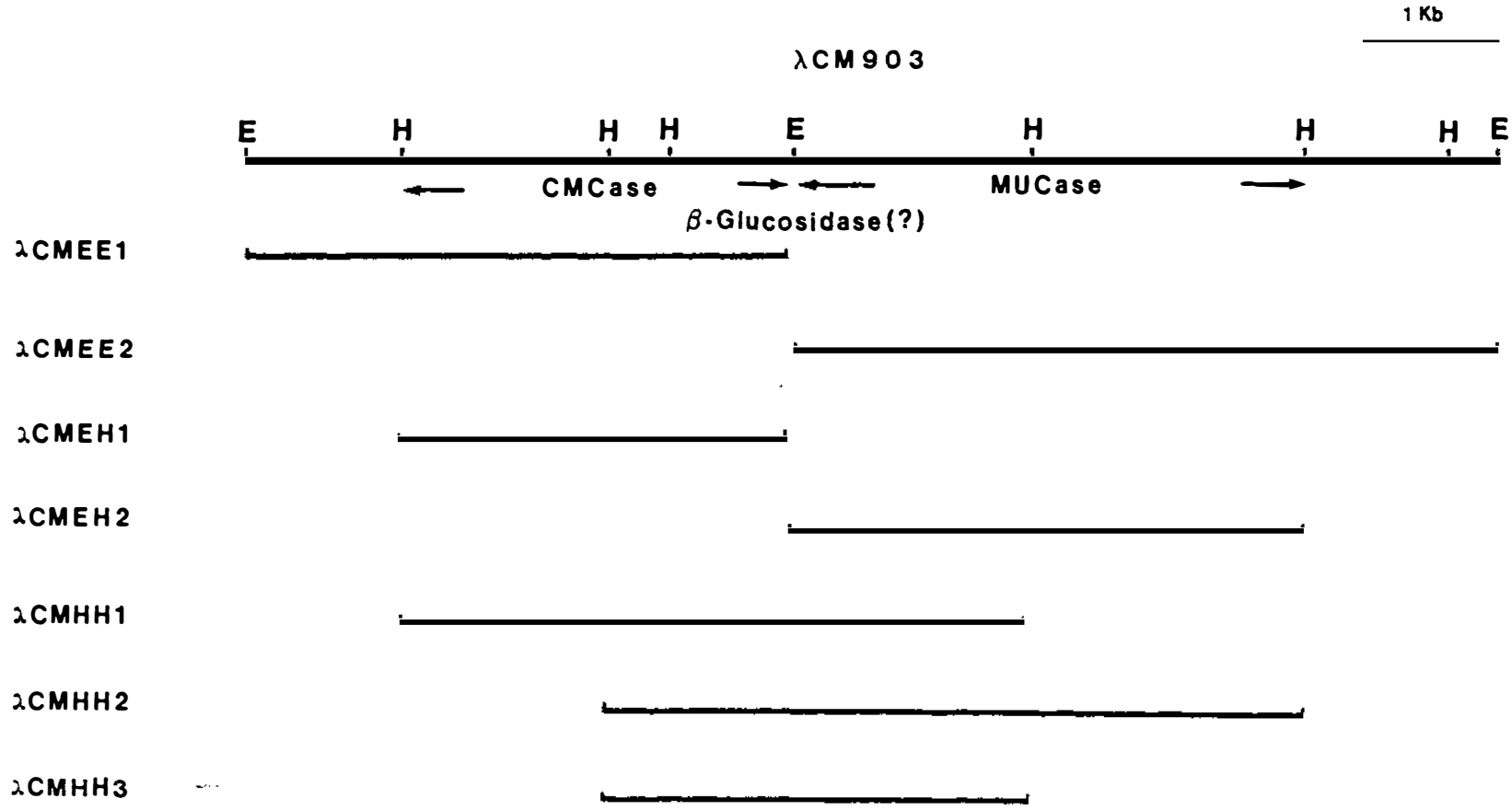


Figure 4-2 Subclones of λCM903 and genes location

Materials and Methods). After analysis of the expression of each subclone (λCMEE1, λCMEE2, λCMEH1, λCMEH2, λCMHH1, λCMHH2 and λCMHH3), the MUCase gene and CMCase gene were localized as shown (Fig. 4-2). No subclones were found that would express β-glucosidase activity except in the subclone of 4.5 kb *Hin*dIII-*Hin*dIII fragment (λCMHH2).

4-2-2 Homology studies of λCM- and *celA* gene of *C. thermocellum*

The hybridization was performed using the Chemiprobe kits (re: Materials and Methods)- a system for non-radioactive tagging of DNA probes and for the visualization of the tagged probes following hybridization reactions. The hybridization data showed that the CMCase gene of λCM903 (λCMEH1) has varying degrees of homology to the other CMCase positive clones. Homology was also detected when the MUCase gene of λCM903 (λCMEH2) was hybridized to λCM404, λCM407 and λCM901. The results of cross hybridization of all eight representative clones are shown in Table 4-1. The homology between *C. thermocellum* celA gene (a gift from J.P. Aubert) and the clones of *R. flavefaciens* was compared. No homology was detected using the fragment of celA gene as probe (Table 4-1).

4-2-3 Localization of enzymes in *E. coli* cells

To determine the difference in the expression of cellulases by *R. flavefaciens* and *E. coli*, the cellular locations of the cellulase enzymes cloned in *E. coli* were determined. The insert of the above λCM903 was subcloned into pGem3-blue after *Eco*RI digestion and transformed into *E. coli* JM109. Both CMCase and MUCase were expressed in the periplasmic space and cytoplasm using recombinant pGem3-blue (CMC^+) or pGem3-blue (MUC^+) in *E. coli* cells (Table 4-2). It showed a completely different expression from *R. flavefaciens* which secretes the enzymes to an extracellular space.

4-2-4 *Exo*III deletion of λCM903

To enable further localization and sequencing of the *R. flavefaciens* 186 cellulase genes on the 7.3 kb *Hin*dIII-*Hin*dIII fragment of λCM903, deletion of this DNA was carried out using *Exo*III. The DNA of λCMEH- series and λCMHH- series was prepared using the DEAE method and digested with *Eco*RI or *Hin*dIII as described in Materials and Methods. After protection of protruding 5′ ends by filling in with α-phosphorothioate dNTP, a second digestion with *Eco*RI or *Hin*dIII was performed. The treated 20 μg DNA was deleted with *Exo*III enzyme at the deletion rate of approximately 250 bp per minute from each end (digestion at 37 C with 150 U of *Exo*III and sampled at 30 second intervals). The deleted fragment was religated to lambda vector DNA with T4 DNA ligase and transfected into *E. coli* LE392. The deletion strategy is shown in Fig. 4-3.

4-2-5 Examination of endoglucanase, exoglucanase and β-glucosidase gene expression

The deletion subclones (λdGCM- series) were examined for CMCase, MUCase and MUGase activities as described previously. The resulting localization of the endoglucanase, exoglucanase and β-glucosidase genes were determined within 250 bp accuracy (Fig. 4-4).

4-2-6 Partial purification of cellulases from λCM903

The crude enzyme extract was prepared from λCM903 as described in Materials and Methods using P-2 and P-6 columns to exclude small peptides and salt. The protein eluent of the P-6 column was passed through a P-60 column with 3.8 cm I.D. x 25 cm at a flow rate of 1 ml/min. The resulting two peaks of activity were named peak A and B (Fig. 4-5). After assaying for CMCase, MUCase and MUGase, peak B showed only CMCase. Peak A had both MUCase and MUGase activities and weak CMCase activity.

Table 4-1. Cross hybridization of λCM clones and *celA* of *C. thermocellum*.

Clones	Probes				
	λCMEH1	λCMEH2	λCMHH1	λCM201	*celA*
λCM201	-	-	-	+++	-
λCM301	-	+	-	-	-
λCM401	-	++	-	+	-
λCM404	+		-	-	-
λCM407	-	+	-		-
λCM901	-	++	-	+	-
λCM902	++	-	+		-
λCM903	+++	+++	+++	+	-
celA	-	-	-	-	+++

+, ++, +++, different degree of homology; -, negative;
, different results from duplicate experiment.

Table 4-2. Cellular locations of cellulase enzyme in *E. coli*.

Locations	Enzyme activity (mU)				
	CMCase	MUCase	pNPCase	MUGase	pNPGase
Intracellular	58	0.15	0.04	0.45	0.27
Periplasmic space	93	1.2	0.89	0.05	0.01
Extracellular	ND	ND	ND	ND	ND

ND, not detectable.

Figure 4-3 A deletion strategy of λCM903

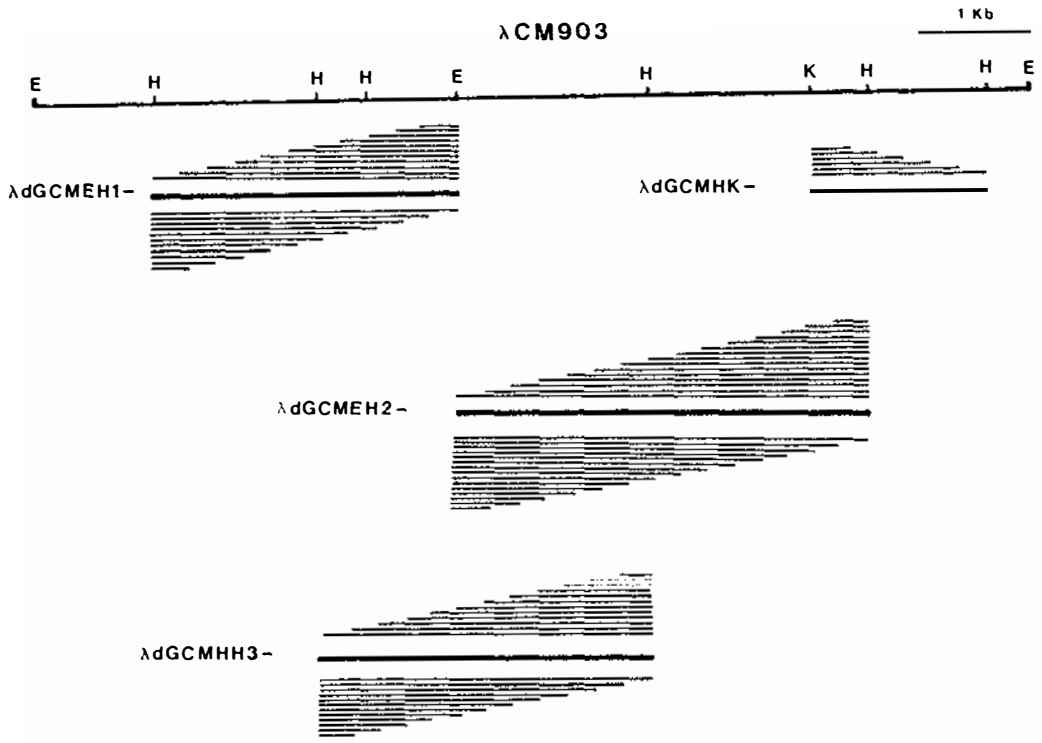
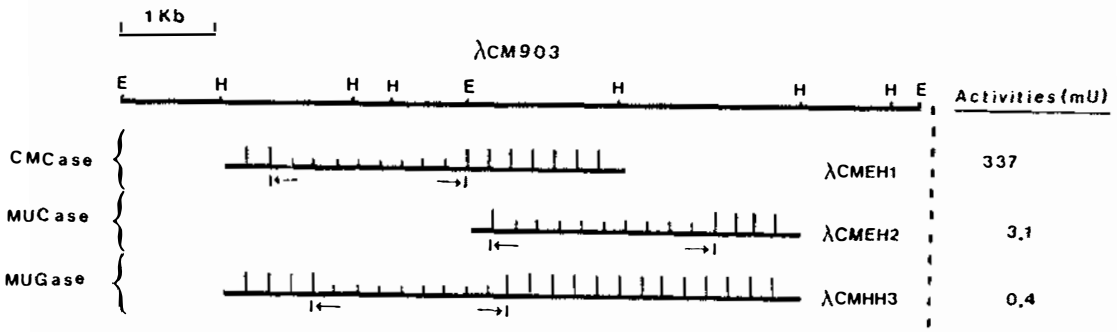


Figure 4-4. Gene location of λCM903.



Peak A was passed through a P-100 column of 3.8 cm I.D. x 60 cm at a flow rate of 0.5 ml/min. The resulting three active peaks were named peak C, D and E (Fig. 4-5). Peak C had both MUCase and CMCase activities. Peak D had only MUGase activity. Peak E showed both MUCase and MUGase activities. The fractions of peak C were introduced to a P-300 column of 2.5 cm I.D.x 75 cm with a flow rate of 0.1 ml/hr. Two active peaks (F and G) were recovered (Fig. 4-5). Peak F had CMCase and MUCase activities. Peak G had MUCase activity and very weak CMCase activity. The summary of partial purification is shown in Table 4-3.

4-2-7 Characterization of cellulase gene products

The protein sample partially purified from λCM903 clones in *E. coli* by column chromatography was used for biochemical characterization of endoglucanase (fraction B), exoglucanase (fraction D) and β-glucosidase (fraction G).

4-2-7-1 Endoglucanase

With zymogram staining techniques, the molecular weight of endoglucanase was estimated as approximately 45 kDa and two small subunits of 10-13 kDa. The 45 kDa protein seems to be identical to the purified enzyme from *R. flavefaciens* (approximately 42 kDa). The optimal pH and temperature of endoglucanase were 5.5 and 30 C (Fig. 4-6, 4-7), respectively. More than fifty percent inhibition of activity was shown by 20 mM EDTA, 15 mM EGTA, 0.5 mM iodoacetate, 10 mM cellobiose or 25 mM glucose in phosphate buffer pH 5.5 at 30 C. Also, more than fifty percent inhibition was given by the following surfactants: 0.05 % SDS, 2 % Triton X-100 or 2.5 % Tween 80 with the same conditions as above (Table 4-4).

Figure 4-5 Partial purification of cellulases from culture of λCM903 with serial Bio-Gel P columns

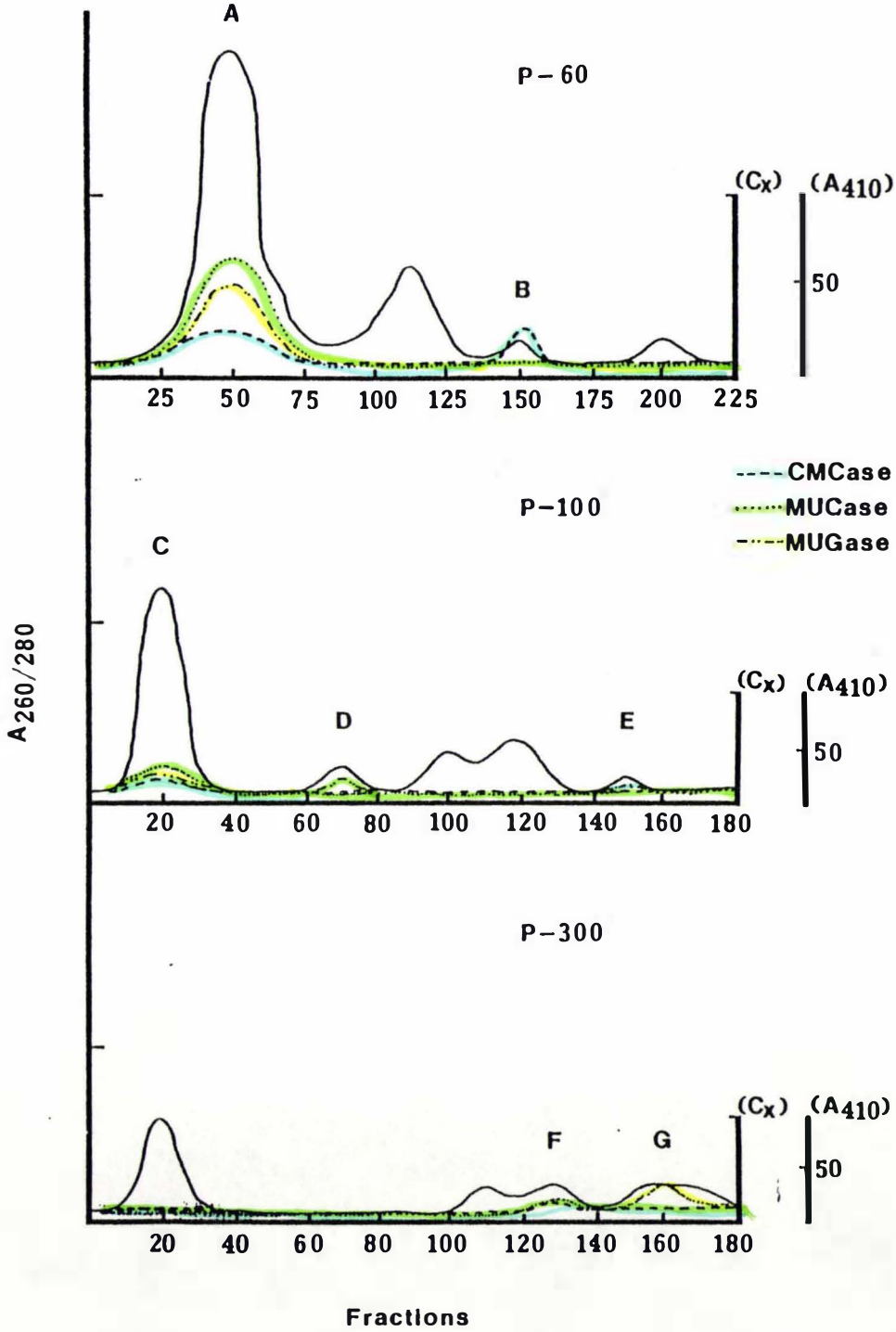


Table 4-3. Summary of partial purification of endoglucanase, exoglucanase and β-glucosidase.

Purification step	Total activity (μM/min)	Total protein (mg)	Specific activity (U)	Recovery of activity* (%)	Purification# (-fold)
Endoglucanase					
Crude extract	127.4	47.2	2.7	100	-
P-2, P-6	103.7	31.1	3.3	81	1.2
P-60 (peak B)	48.8	5.6	8.7	38	3.2
Exoglucanase					
Crude extract	61.36	47.2	1.3	100	-
P-2, P-6	55.98	31.1	1.8	91	1.4
P-60 (peak A)	39.2	19.6	2.0	64	1.5
P-100 (peak C)	18.6	9.3	2.0	30	1.5
P-300 (peak G)	3.7	1.2	3.1	6	2.4
β-glucosidase					
Crude extract	28.32	47.2	0.6	100	-
P-2, P-6	24.9	31.1	0.8	88	1.3
P-60 (peak A)	15.7	19.6	0.8	55	1.3
P-100 (peak B)	3.4	2.6	1.3	12	2.2

*, as a % total (crude extract as 100%).
#, as a ratio at specific acitvity in crude extract.

4-2-7-2 Exoglucanase

The estimate of the polypeptide molecular weight from SDS-PAGE zymograms was 90 kDa. The optimal pH and temperature of exoglucanase were 5.2 and 36 C (Fig. 4-6, 4-7), respectively. More than fifty percent inhibition of activity was shown by 30 mM EDTA, 20 mM EGTA, 0.1 mM iodoacetate, 10 mM cellobiose or 15 mM glucose in phosphate buffer pH 5.2 at 36 C. Also, more than fifty percent inhibition was given by the following surfactants: 0.075 % SDS, 1.5 % Triton X-100 or 2 % Tween 80 with the same conditions as above (Table 4-4).

4-2-7-3 β-Glucosidase

The molecular weight of the enzyme in the presence of SDS-PAGE was estimated to be 76 kDa. The optimal pH and temperature of β-glucosidase were 6.3 and 33 C (Fig. 4-6, 4-7), respectively. More than fifty percent inhibition of activity was shown by 20 mM EDTA, 20 mM EGTA, 0.25 mM iodoacetate or 30 mM glucose in phosphate buffer, pH 6.3, at 36 C. Also, more than

Figure 4-6 A pH profile of cellulases of λCM903

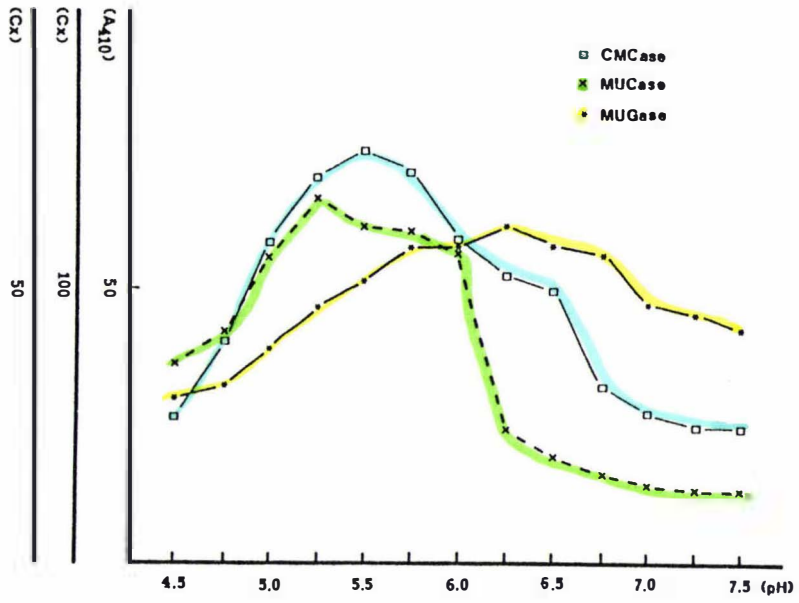
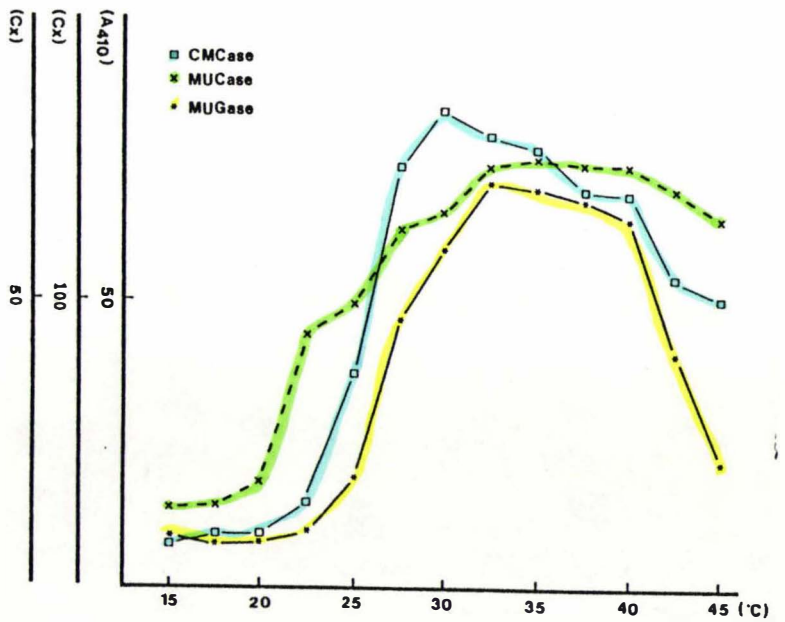


Figure 4-7 A temperature profile of cellulases of λCM903



fifty percent inhibition was given by the following surfactants: 0.05 % SDS, 2.5 % Triton X-100 or 2.5 % Tween 80 with the same conditions as above (Table 4-4).

4-2-8 Substrate specificity assay

Activity against 1% CMC, 1% lichenan, 1% cellobiose, 0.1% Avicel (PH101), 0.5% PMP and 1% xylan was tested by measuring the level of reducing sugar as described in Materials and Methods. The activities against specific substrates are shown in Table 4-5. Lichenan is a substrate for assaying degradation of β-(1,4)- and β-(1,3)-glycosidic bonds. Five out of eight representative clones showed a positive reaction. Avicel and PMP have different degrees of crystallization. Particularly, Avicel is the specific substrate for assaying exo-glucanase due to its high degree of crystallinity. Three out of the eight clones showed hydrolysis of Avicel, in agreement with the assay of MUCase. Also, cellobiose hydrolysis confirmed the reaction with MUGase of λCM902 and λCM903. Three out of eight clones had xylanase activity, again supporting the proposition that the genes coding for enzymes involved in biomass degradation arelocated closely in the genome.

4-3 Discussion

4-3-1 Substrate degradation

In detection of enzymatic degradation of different substrates (CMC, lichenan, cellobiose, Avicel and PMP), it was shown that the clones λCM404, λCM407 and λCM903 could degrade Avicel. Since these three clones showed both CMCase and MUCase activities, it has been suggested that the combined reaction of endoglucanase and exoglucanase could be important in the degradation of Avicel. However, λCM901 which has both CMC^+ and MUC^+ activities did not degrade Avicel. Also, all the clones did not show the ability to degrade PMP which had a different degree of crystallinity to that of Avicel. Further investiga

Table 4-4. Inhibition factors of cellulase enzyme.

Inhibitors*	Endoglucanase	Exoglucanase	β-glucosidase
Surfactants (mM)			
SDS	0.5	0.075	0.05
Triton-100	2.0	1.5	2.5
Tween 80	2.5	2.0	2.5
Chelate agents (%)			
EDTA	20	30	20
EGTA	15	20	20
Iodoacetate	0.5	0.1	0.25
Feed-back products (mM)			
glucose	25	15	30
cellobiose	10	10	-

*, more than 50 % inhibition when compared with non-treatment sample

Table 4-5. Specific substrate degradation by λCM clones

Clones	Substrate					
	CMC	Lichenan	Cellobiose	Avicel	PMP	Xylan
λCM201	+	-	-	-	-	-
λCM301	+	-	-	-	-	-
λCM401	+	+	-	-	-	-
λCM404	+	+	-	+	-	+
λCM407	+	+	-	+	-	-
λCM901	+	+	-	-	-	+
λCM902	+	-	+	-	-	-
λCM903	+	+	+	+	+	+

tion in the sequences and their expression of these genes and the structural differences between the two type of substrate may help to explain this discrepancy.

The degradation of lichenan by λCM401, λCM404, λCM407, λCM901 and λCM903 showed that the cloned cellulase not only hydrolyzed the β-(1,4)- linkage but also the β-(1,3)- linkage. In *C. thermocellum* some CMC^+ clones showed degradation of xylan (Hazlewood et al. 1988). It is possible that the cloned cellulase might have a non-specific reaction with lichenan and xylan or the clones might contain more than one gene in the fragment.

4-3-2 Plasmid instability

Subcloning and expression of the cellulase genes from *R. flavefaciens* 186 using plasmid vectors in *E. coli* and *B. subtilis* were unstable but lambda vectors were stable. We experienced difficulties in recovering the recombinant plasmids after a short period of cultivation or in long-term frozen storage (-70 C). The plasmid vectors chosen were pUC19 (high copy number vector), pGem3 (high copy number with primer sequence), pOU71 (low copy number with temperature control sequence) and pCK17 (shuttle vector of *E.coli* and *B. subtilis*).

The pUC19 and pGem3 vectors were chosen because of their high copy number, ease of transformation and their use in direct sequencing. The instability of these recombinants may have been caused by high level of expression of the cellulase gene in the cytoplasm and periplasmic space (O'Neill et al, 1986; Joliff et al, 1986; Schwarz, et al, 1987).

The low copy number vector pOU71 was chosen to overcome the problem of overproduction. Below 37 C the plasmid is present at one copy per chromosome, whereas at 42 C the copy number increases to more than 1000 plasmid molecules per cell (Larsen

et al., 1984). If the clones in this vector were stable at 30 C, a temperature change could be used to regulate the expression of target gene. Unfortunately, the subclones in this vector also showed instability indicating that clones were unstable even at low copy number. This indicated that the problem may be lie in the expression of gram positive genes in *E. coli*.

A possible test of this thought was to use a shuttle vector which is functional in both Gram-positive and Gram-negative hosts. A vector, pCK17, which replicated in both *E. coli* and *B. subtilis* was chosen for the cloning experiment. The plasmid pCK17 combined the replication regions of a *Streptococcus* cryptic plasmid with the antibiotic resistance genes (kanamycin and chloramphenicol) from *Bacillus* vector pBD64. Again, instability of recombinant plasmids was found in both hosts (*E. coli* and *B. subtilis*). The last attempt at subcloning in plasmid vectors was the use of plasmid pBD64. Plasmid pBD64 can be used for cloning DNA fragments in *B. subtilis* and has selectable phenotype of chloramphenicol resistance and kanamycin resistance in *Bacillus* (Gryczan et al., 1980). This was to test if the cloning of genomic DNA from Gram-positive to Gram-positive host would be easier than that of Gram-negative host. However, no stable recombinant clones could be isolated. The results indicated that the gene regulation systems or enzyme restriction modification systems of *Ruminococcus* and *Bacillus* were different. The problem of plasmid instability has been investigated recently with respect to that of difference in G+C ratio (Paradis et al, 1987), ion factor (Ohmiya et al, 1988), nutritional factor (Ohmiya et al, 1988) and temperature factor (Son et al, 1987). Further study of the expression of cellulases in foreign hosts will reveal the complexity of regulation of these genes.

4-3-3 Enzyme expression

From the data of *R. flavefaciens* (unpublished data, Asmundson), cellulase enzymes were secreted into medium, while in *E. coli* with plasmid vector, the cellulase enzymes were found mainly in the periplasmic space and cytoplasm. It was very difficult to detect any cellulase activities in the medium. Due to the instability of recombinant plasmids which contained the cellulase genes has made the enzyme location a difficult task. Normally, after transforming the recombinant plasmid and plating onto selective plates, the positive recombinants which showed as small colonies on plates were immediately transferred into Eppendorf tube and extracted the enzymes from different fractions.

The endoglucanases and exoglucanases of *Ce. fimi* (Langsford, et al, 1987; Paradis, et al, 1987; Guo et al, 1988; Curry, et al, 1988; Wong et al, 1988; Gilkes, et al, 1989), *C. thermocellum* (Cornet et al, 1983; Sacco et al, 1984; Schwarz et al, 1986; Joliff et al, 1986) and *Bacillus* spp. (Robson and Chambliss, 1986; Lee and Pack, 1987; Lo et al, 1988; Kawai et al, 1988) were glycoproteins but the proteins which were expressed from recombinants in *E. coli* were non-glycosylated. Since the endoglucanases purified from *R. albus* (Kawai et al, 1987; Ohmiya et al, 1988) and *R. flavefaciens* FD1 (Gardner et al, 1987) were glycosylated, it is suggested that glycosylation is important for efficient secretion and gene expression. Glycosylation could be involved in the protection of enzyme against protease (Langsford et al, 1987). So far, glycosylation was not detected from cellulases which were purified from culture of *R. flavefaciens* strain 186 (unpublish data, Asmundson). Further study on the glycosylation of cellulases will improve the stability of the cloned cellulase genes in foreign hosts.

4-3-4 Choice of deletion method

Although site specific mutagenesis with transposon is the common way of localization of genes, other considerations (eg. the size of the fragment, the screening method and the method of sequencing) may affect the use of this method. The method of choice for the study of cellulase genes was the use of *Exo*III deletion which has the benefit of locating the gene and also allowing the use of the deleted clones for sequencing analysis.

The growing use of bacteriophage lambda as a cloning vector both for genomic and cDNA has led to the development of novel techniques in the manipulation of lambda phages for cloning and sequencing of DNAs. New method such as the one described by Manfioletti and Schneider (1988), had opened a way to do direct sequencing of DNA in lambda vector, which avoided laborious mapping and subcloning steps and was of particular benefit for the problematic subcloning of DNA from Gram positive bacteria into the Gram negative microorganism *E. coli*.

4-3-5 Enzyme purification and characterization of gene products

CMCase, MUCase and MUGase from lysate of λCM903 culture were partially purified. The MW of CMCase, MUCase and MUGase was estimated as 45 kDa, 90 kDa and 76 kDa, respectively. Several native cellulases from ruminococci were also purified (Pettipher and Latham, 1979; Ohmiya et al, 1982; 1985, 1987). The endoglucanase, exoglucanase and β-glucosidase purified from *R. albus* strain F40 had estimated MW of 50 kDa, 110 kDa and 82 kDa, respectively (Ohmiya et al, 1982; 1985; 1987). In addition, endoglucanase purified from strain 67 of *R. flavefaciens* was 89 kDa (Pettipher and Latham, 1979), whereas in strain FD-1 the exoglucanase was about 118 kDa (Gardner et al, 1987). So far, the only reported recombinant gene product of

CMCase from *R. albus* SY3 strain was 56 kDa (Romaniec et al, 1989). When compared with the partially purified CMCase of strain 186 of *R. flavefaciens* (62 kDa; unpublish data, Asmundson), the recombinant gene products was about 7 kDa shorter than the native protein. The size of cellulases from recombinant clone in *E. coli* correlated to the size of native proteins.

4-4 Short summary

The locations of the endoglucanase, exoglucanase and β-glucosidase genes in λCM903 were determined. Two separate fragments were subcloned and named λCMEH1 (CMC^+) and λCMEH2 (MUC^+). Cross hybridization between *cel*A gene of *C. thermocellum* and 26 recombinant clones of *R. flavefaciens* was performed and showed no significant homology. However, different degrees of homology were found among the 26 ruminococcal clones.

The difficulty of subcloning the ruminococcal DNA fragments into plasmid vectors was their stability in the new hosts. Different vectors with different functions were tried but all recombinant plasmid clones showed instability in *E. coli* and *B. subtilis*.

The cellular location of enzymes in *E. coli* cell was determined. Most of the endo- and exo-glucanase were found in the periplasmic space. Partial purification of these cellulase enzymes from *E. coli* cells using chromatography was performed. The pH optimum, temperature optimum, inhibition and molecular weight characterization of these cellulase enzymes was achieved.

Chapter V

Nucleotide sequencing and analyses of a 7.3 kb cellulase gene fragment

Table of contents

5-1 Introduction

The development of improved sequencing methods for determining the order of nucleotide bases in deoxyribonucleic acid has led to the rapid increases in the number of genes sequenced and also there is a wide application of computer in the processing of information on biological systems contained in these sequences. Several sequencing procedures have been developed (Maxam and Gilbert, 1977, 1980; Sanger et al, 1977; Messing et al, 1981; Guo and Wu, 1982; Yanisch-Perron et al, 1985; Prober et al, 1987; Berchtold, 1989) which were based on two types of methods: the chemical degradation method and the dideoxy chain termination method. Recently, the M13 phage single-stranded vector system has become the most common protocol for cloning and sequencing using the chain termination method because of its simplicity and rapidity. Lambda sequencing is a newly improved method which overcomes the difficulty of subcloning DNA fragment by plasmid vectors. The genomic library constructed in lambda phage are sequenced directly using this technique.

In this section the sequence data of a 7.3 kb fragment from λCM903 is presented and analyzed with UWGCG package.

5-2 Results:

5-2-1 Nucleotide sequence results

The deletion clones (λdGCM- series) of the 7.3 kb fragment were sequenced by the dideoxy termination sequencing method using lambda vector. Both strands were sequenced completely using overlapping clones. Preliminary sequence data obtained for the 7.3 kb fragment was assembled using the computer programmes GELSTART, GELENTER and GELASSEMBLE (UWGCG; Staden, 1980). The resultant completed DNA sequence showing the main restriction enzyme sites is presented in Fig. 5-1.

In order to identify potential coding regions, the sequence data was analyzed using FRAMES (Devereux et al., 1984), CODONFREQUENCY, CODONPREFERENCE and TESTCODE (Fickett, 1982). From these results and those presented below (presence of Shine-Dalgarno sequences and promoter sequences), five separate open reading frames (ORFs) were predicted to be the coding regions (Fig. 5-2).

5-2-3-1 First open reading frame

The first ORF was identified as the structural gene for the *R. flavefaciens* endoglucanase gene (designated *ren* A : *Ruminococcus* endoglucanase) because it corresponded to the region of deletion map showing endoglucanase activities . The sequence of this ORF, the 5′ upstream sequences and the deduced amino acid sequence is shown in Fig. 5-3. The ORF has a coding region of 2157 bp, starting with an ATG codon at position 829 and ending with a TGA stop codon at position 2871.

The nucleotide sequences upstream of the 5′ end of ORF1 were screened for sequences complementary to the 3′ end of tmx 16S rRNA of *B. subtilis* and *E. coli* (Shine and Dalgarno, 1974; Moran et al., 1982; Stormo et al., 1982). A strong Shine-Dalgarno (SD) sequence (CGGGGA) containing 5 C/G's was located at position -10 to -5, 4 bp upstream of the ATG initiation codon (Fig. 5-3, Table 5-1). The region upstream of this presumptive translational start site was analyzed for potential promoter sequences homologous to *E. coli*, *Ce. fimi* and *C. thermocellum* consensus sequences (Hawley and McClure, 1983; Wong et al, 1986; Begiun et al, 1985) using BESTFIT (Smith and Waterman, 1981) and GAP (Needleman and Wunsch, 1970). Putative promoter sequences $^{-36}$TTTACA^{-31} and $^{-59}$TACGAA^{-54} were found as -10 and -35 regions, respectively. The distance (17bp) between -10 and -35 was similar to the consensus distance (17bp) of *E. coli* (Fig. 5-3, Table 5-1, Hawley and McClure, 1983). At the 3′-terminal end of the gene downstream of the stop translation codon, several inverted repeat sequences were identified (Fig. 5-3) that had the potential to form stem-loop structures which may be involved in rho-factor-independent transcriptional termination. The resulting DNA sequence of the ORF was analyzed using CODONFREQUENCY and CODONPREFERENCE (Devereux et al., 1984) revealing a strong bias towards codons containing dG+dC (71.7% G+C, Table 5-2; Fig. 5-4).

The ORF was sufficient in size to code for an unmodified protein of 680 amino acid residues, which consists of 65 acidic amino acid (Asp + Glu; 9.0%); 171 basic a.a. (Arg + Lys; 23.8%); 10 aromatic a.a. (Phe + Trp + Tyr; 1.4%) and 232 hydrophobic a.a. (aromatic + Ile + Leu + Met + Val; 32.5%) with a calculated MW of 75.8 kDa. This predicted molecular weight of the translated protein was far larger than the minimum size

Figure 5-1 Nucleotide sequence of 7.3 kb of λCM903

```
   1  AATTCGGAGG CCCTGGATTT TAGCTAGCTC CCCTTTTTTA AAGGAATCGT
  51  AGATCGCGCT AGCGATTCTC CAGAGCTAGC TCGATTTCGA TCTATTCGCT
 101  TGATCGCGCT TTAACTCTCG TATAGCTACG ATCGAGTCGA AGCTAGATAG
 151  CGAGAGATCG GCTTTCCCGA TCAATTATCG AATAGATCGC GGGATACATC
 201  CGGTATCGGC TCTCGATCGC CTACCTACGC TAGCAATATA TATTACCAAT
 251  CAATATATAA AGCTCTTTTA TACTTACGAT TTACGATTCC CTCTCGCTAC
 301  GATCGGCTAC AGCTCGATTT ACGCTCGCTT ATGAAAAGCT ACGCATCGGC
 351  ATCCGCTTAC GAAGCTCGCT TATACGCTTC AGCGGGCTAT CGAGAGAAAT
 401  CGGAAAAACG ATCGCTCGAT CGGGCCTCAG CGCTTATACT CAGCCGCGAT
 451  TACGCGCGCC TACGAGAAAA GGGGGCCCCC CCCCCTTTAT TCATCGGCTC
 501  GGCGCTGGAA TTATATTAGA GATAGATCGC TCAGGCTACG CTTACTTCGC
 551  GATTCCCGAT CGCGGCTCAG CGCGGCTTCG ATCGCTCGGT ACTTGAGCTC
 601  AAGCTACGAT CGGTACACAG CATAGATCGA CTATTCATCA ATCGACTACT
 651  ACGATCGGCC TTTTTGAAAA AAGCTCGATC GCTTACGCTC GATTCGAAGG
 701  CGCTCAGCTA GGCTACTATC TGGATATATT CAmxGAGAGA GCTCGCGCGT
                                    PstI
 751  CCGATACGAT CGTACGGAAT ACGAATTACC GCTGCAGAGG CATTTACAGG
 801  GCCTTTATAT ATACGATTCG GGGATACGAT GGTAACCAGC GAAAGCGCTC
          KpnI
 851  GATCGGGTAC CCCATCGAGC TCGCCTCCGA CTTCACCTAG TCCCAGCCCT
 901  ACGTCGCCCA CCCCGCCTTC GCCTTCGˆXxC GAATCCAGCC CTACTCCCTC
 951  GTCGCCAGCG TCACCGCGCT CGCCTACGGT GCGCTGCCGC GAAGCGGTAG
1001  CACCTGCGTG CCACGCAGAC GCACGTACCG CAGGTCGCGG GCGCGACCCG
1051  CGGCACCAAT GCGACGCCCG CCTCCGGGGC CGCCACCGCC CGCGGCTGGG
1101  CCACGCGGTG GTACGCGACG ACGGCCTGGG TCCACACCAA CGCGACTGTG
1151  TCCGTGGCGC TTGTCACGGG CCATCTGTTC ACCGCGCCGA TTGGCACCAA
1201  CCACGGCGTG CTGCACTTGA ACTTGTGGGA CAACTCGTCC ATCCGCAGTC
1251  CGGGCTCGTT CCGGTGGTTC GAGGGGTTCC GGCGCCGCCC CCCCGTGATC
1301  GTCGGCGAGA CCAATCCCGG CCGCGACGGG CACTGGTTCA GCTCCCCCTT
1351  CCGCGGGACC GGGAACAACC AGGCGGTCAC CAACCCCACG ACGCCGACGC
1401  TGTACACGGC GGGGGCCGAG GCGTACGTGG GCAACTCCAA TCCCGCCGTC
1451  CTCGGGATCG GCGTGTACGG GACCACCGTG ACGGCCCATC CGACGGCAGA
      HindIII
1501  AGCTTTGACC CCCCCGACCG ACAACGGGCT GTGCGACGCC AACGGGGACG
```

Figure 5-1 (continue)

```
 1551  AAGGCACCGA CTCGTGGACC GCCGGGTGCG GCCTCGGGAA CAACCGGTGC
 1601  CGCGGGTTCG CCAACGCGGT CCGGGGGTGC GGGCGCACGA ACACCGCCGA
 1651  CGGCGCGGCC CGCTTGCGGG TGACCTGGTG GAACGCGCGC AAGCCGCTCA
                                   ClaI
 1701  CCCGCGTTCC GTGCAGCGCG AACCGATCGA TCACGGTGGT CGGGAACTAC
 1751  ACCGCGTGGG TGCGCAGGGC CGTGGGGCAC GAGAAGATCG TGGCGACGGT
 1801  GCTGGTCGAG AACGCGGGCG TGCTCAACGA ATCCGGCGCG GTGGTCGGCT
 1851  ACACCGAGCG GAAGGCCGAC GGGAAGCATC CCTCGGCCTA CGAGCGCGGG
                                  HindIII
 1901  GCGTGCCTTC CGGTCTCCAG CGACGACAAG CTTCGGCGGT GGAACAACGG
 1951  CATCCGGGGC GGCGGCGGCC GCCGGAAGGA CGGGATCTCC AACGTCGCGC
            SmaI
 2001  CGTACCCCGGG TCCCGCGCGC CGCAAGCAGG CGCGTCCGGC CAACGGCGTG
 2051  TGGGGCCGCG GGGCGGCGGC GTCGGCAGGC TACGCGTCTC TGCGCTACGT
 2101  GTCGTCCCCC TCCGGCGCCG GTCTGGACGC GATCGTGCGG ACTGTCACGC
 2151  TCGGCCGGGG GTCTCCCACC GGGCTACTCC GCGCGGGCAC CCGCACCGGT
 2201  GTCGGGCTGT GCGGGGCTCC CGGCAAGATG ACCGTGCGCG AGACCTACGG
 2251  CGCCGTCACG TGGACGCGAC GGCGGGCCAG GGTCGGCGGG ACCGGCCCCA
 2301  GCTACGGGTA CTGGTACACC AACTGCGGCG GGAAGCTACG GGCCAGCGTC
 2351  CGGACGCTAT GGGAGCTCGC GCGCACCCGA TCCACTGGGG CGGTGTCAGG
 2401  CTACCCCGCG GGGCGCGGGC TAGCCAAGGA CACCGGCGGG ACCGCACTGC
 2451  TGTCCGGCTA CCCCGCGTCC GCCGCGACAG TCGTGCCCAA CGAGACACTC
 2501  GCGCCGACGT CGTCCGCGCA CGTCGGCTAC GTACGCGCCG GGGCGCTGCC
 2551  GGTACTGAGC GAGCTGAGCT ACGTCGCGTC CACGAGCGGC GCCGCCGTCG
 2601  CGCTACGGTG GGCCGCGTTA CGGTCCGAGG CTGCTTCAGG GAGCGCGTCC
 2651  GAAGCGACCC CCACCGCAGC CACGGTACGC GCGCCCAACA TGTCCACCGA
 2701  GCTACTCGCG ACGCTGCCGG GGTGCGAGCG GGCCTCGGCG AACAACGCGT
 2751  CCACAGGCGC GCCGCGGCGA CCCGAGGCAC CGGCCGACAA GAAACGCGCC
 2801  GATCGGGCTC GGGCGATACG TCCGGCGCGG GACGCGGGGG AGCCATCGCT
                                   EcoRI
 2851  AGCGGCCAAA CGACAGCCTG ATCTATATCG TACGAATTCT GGATATAAAG
 2901  CTCTAGCTCG CGTACGAAGT CTACGCGTAC AGCGATACGC GAGCGTAGCT
 2951  AGCGAGCGGA TCGTAGCTGG AGCTTCGCGA GCTGGCAGCA TCGAGCGCGA
 3001  TCGAGCGATA CGAGCGATCG TATCGACGTT CGATCCATCG AGTACGATCG
 3051  TATGCTCGAT CGTACGAGTC GTACGTATCG TTCGAGCGTA CTTACCCCTG
 3101  AGCACGCGTC AGGCAGCGAG CTACAGCAGA GCTATCGCGA GCAGCGATCG
```

Figure 5-1 (continue)

```
 3151  AGTCGAGCTA TACTATATAT CGCGCTATAT ATACGGCGGC CGTCGCGCGC
 3201  CTACTACGGC ATTACCCTAT TCAGCGAATT TGCGCAGCGT ATATTACGCT
 3251  AGCTACGCGC TATCGGTCGA GGCGATACAG AGAGGACTAC GCTACGTCTA
 3301  CGCGCGTACG CGCGTACTAT GGAGAGTACT CGACGATCGT ATATACAGCT
 3351  ACGCCCGTAC GGGCGGATCG GCGTAGCGCG CGATCGGATA GCGACTACTA
 3401  TATATACTAT CGGAGAGAGT CTCTCCCGAG CTACGCGACC GAGCTCAGCT
 3451  ACATTTATAT AAATGGTATT TACGGCGTAG CTCCAGCAAG CTACGACACT
 3501  CTTGCAATTT AAGCGTAGCA ACGTTAGCGA ATCGAAGGAA AGAATCGTAA
 3551  AGATTATATA CAGAGCGTGT AAAATCTATG GGGCGATGCG GCAAGACACC
 3601  CAGAGCGCTC AGCGTGAGCG GGTCCACGAC CTCGACCTAG ACCCGGTCAG
 3651  CGGGGTCCTC CGGCTCGCTT CGGACGAGCA GCACGGCCGA TCCGTTGACG
 3701  GTCTCGACCA GCACGTAGCC GTGGAGGGGC GGGACCACGA CCTCTCGGCC
 3751  CACGACCTGC TCGGCGTGGC GGCTCGCGTC TCGGACGAGC CGCTACTCTA
 3801  CCGGGACGAT CCCGCGGACA GGGAGCAGCC CCGCCCGGAC GGGCGACAGG
 3851  CTGTCCCGGA GGAGGTCGAC CAGGGGGCAG CGAGCGACGG TCACGACCTC
 3901  CGCGTCGACG CGCTCCAGCA CTCCGGACTG TCCCAACAGC ACGACCGCGG
 3951  GCGCCCGGAG GTCGCGCGAC ACCTTGACGA GGACGGCAGG ACGGTCCTCA
 4001  GGGCCGACTC GCAAGACGAC ACACAGGACT CGAGACTCCA GCTTCTAGCG
 4051  TGGTGCCGGC TCCTCACGGT CCGCGGCCAT GTCGACCCGG TGGTGGCTCA
 4101  CACAGGCCGG CAGGCCGACG AACAGGACGC CGGCGACGAG CTGCCCAGCG
 4151  TGCACCAAGC CGAAGACGGC GGCGCCCTTT CCCGTCAGCC CGATCACCTG
 4201  GACCTCAACG TCTCCCGCGA CGATCATTCG GTCGAGCGGC GACTGGACCA
 4251  GGACTCGCAC GACCTCCAGC GTGCCCTCCA AGACCAGGAC CGGCTCGCGG
 4301  CAGTCCTGAG ACTCGACGGC GACCACCTCG ACGACCCGGT GGGCGTCCAG
 4351  CTGCCGCGAC AAGACCACGA CGACTCCGCA CAGGATCACC GATCGCAGGG
 4401  CGAGCGATTG GCATCGGGAC ACCACCCGGA CCTCCGTTCG GAGCAGGTCA
 4451  CCGCAGCGCT CCTGGAAGAC GAGGCACACG TGGTGGGGCA CCAAGCATCC
 4501  TCAACGAGGG TGGACTCGCC GGACCAGGAG AGGGGTGACC ACCCGGATCG
                HindIII
 4551  AAGCTTCGCA AGGCTCGAGG ACGTCCAGGG GGACGAAGCA CGGCTCGACG
 4601  TGGACCAATC ACAGCTCGAC CCGCAGGACG GGCTCAGGGT CGCGGACCCT
 4651  GACCTGTCGG TGGAGGACCA CGGGCAGAGG GACGCCCTGT GCCAGTCAGA
 4701  CCCCCAGCGG GTCGAGCGCG CTGACGGGCT TGACGTCTGC CTCCGACACT
```

Figure 5-1 (continue)

4751 GTTCGACCGC ACCCAGCTCG TCGCCTCCAT CTACTCCTAC GCCCAGCAGC

4801 ACGACCCCCA CCCCTAGCTC CTCAGTATCT TCCGTACCTA CACCGAGCTC

4851 GCCCGATTCA CCGTCGCCTA CGGACCACGC CCGCTCGGGG GACCAGCATG

4901 AGGTGGACCT GCACGACCAG CGCGGCGACG CGGCCGGTCT CCACGACCAG

4951 GTGGATCACT CGCTGGGCGA GGACGTGCGC CTCGCCGACG GAGACGGAGA

5001 CCAGTCCCTC CGGCAGCAGG AGCTCAGCCA CGACGAGCTC AGGGTGGGAG

5051 AGCAGCTTGA GGCTGTCGAC CCGGATCGCT CAGCGCAGGG TGTGGAGCAG

5101 GACGCACCTC TGTCGCAGCT CCAGGCTGTG GCACAATCGG ACGCAGCTGA

5151 CCTACAGCAG GTGCGCGACG TCGATCACGA CGACGTCGTG GCTGGTGAGG

5201 ACGGCGGCCG CGACCACGAG CAGCTCCACC GCTTGGACAC GATCCCAGTC

5251 TTCGATGAGG CACACCCCCA GCCAGGAGCA ATCAGCGATC GGCTACGATC

5301 GTAAGGCTAA GCGACTTATC AGGCGCTAGC GTCGATTAGC GATGCGATAC

5351 GCATCGGTAT ATTAGCGCGC GAATTAAATT TCGGGCCCTA TTTTAATCGG

5401 CGCTAGCCCT ATTACGTACG GCTATACGGC GATTAGCGAT TCGCAGCGTA

5451 TTTCGCGTAC GCGAGCGCGT ACGCTATTAC AGATGCGATT ACAATCGGCT

5501 ATATTACGCG TATATTTACA GCTATACCTA CGGATTCGAT CGTAGCGCGT

5551 ACGCTCATCG GCTACTCAGC GCTACGCGCA GCAAGCGCTA TATTCGGCGA

5601 TCGGCTACTA GCCAGCAGAG TTCATACATA TACGGCGCCA TTCGATCGTA

5651 CGTACGCAGC TAGCATACGT ACGATCGGCT ACGCGCTATC GGCATGCGCG

5701 ATATGCAGCG CTTACGGCGA TCGGCTACGT ATCGGATCGT GCGCGATCGG

5751 TATCGCGTTA TACTCGCGTA TATTTCGCTT CGGCATATAT TATCGCGATA

5801 TTATATATAT ATTCGCATTA TCGAACGGCC GTTATATACG CATCTGGTCT

5851 ACGATCTGCG ATCGAGTATA AAGCGCATTA TATTTCGCTA AAAGCGATCG

5901 GCTTTACTAT CGGCTACGCG TTTCCCCCGT TACGTACTCG TATACGGGTC

5951 ATACTACGCT ATATTACAGC TATCTACAGC GAGCGATATT CGATCGTATA

PstI

6001 TCGGCTATAT ATGGCATACA GCATTATATA CGTATCTGCA GTGCGTCGTA

KpnI

6051 TACGATCGGA GCGATCGGCA TACTACGCGC ATTTATATAC GTCGTCAGGT

6101 ACCTTACAAA AGCTCAGTTC ATACGCGTTA CCGCGCGCGT CGGGTATGGA

6151 GCGTCTTCAG CGCTATCGCG CGGTCTACGT CTACGCGTTA AACGCGCGTA

6201 TCGCGTAGCG CATACAGCAA TTAGTCAGAT TTCACTTAAG TGCGAGTACG

6251 TTCTATTGGC GGCAGATGCG CTCAGTCCGT CCTCGGGGTC CACGACGGCA

6301 GCACTCCGGC CAACCAGGGG GCGAAAGGCG CCAACTCGGC GACCGGCCTA

Figure 5-1 (continue)

```
 6351  GACGCCAGCG CGCTACCGCT CAAGACGCGG GGCGCGCCGT CCGGGCACGT
 6401  ACCGACGATC GTGCTCGGGG TCACCAACCA GGGCGTCCTT CCCGGCGTCG
 6451  CGGCTACGGT CGGCTGCGGC GACACGCGCG GCCCGGCCGA CGCGCGGACG
 6501  CGCGCGATCG GCCGCGGGAC GGCCTCCGCG ATCGGCCACG CGCAGCGCAC
 6551  CCAGGCGCAG GCGAGCCTCC CCGACACGGG CGCCAGCGCC GGGATCGCCC
 6601  GGTCGCTGAA TGCGGCGTCC AACACGACGG CTCTCGCGAC GGAGCGCAAC
                                HindIII
 6651  GGCGCGGGCA ACGAGGTCCA CGTCTCGAAG CTTCCCACGT CGCAGGCCCA
 6701  GCACGAGCGG ACGCTGGCGC TGGCGACAGC TGCGGCCGCT GCGGCTGGCG
 6751  CGGCCGGCCA CGCGACCTCC GGCAAGACCG GGCCGCTGAC GCTCGGCGCC
 6801  GCGGCCTCGG CGACGTGGCA CAGGTGGGGC GAGCCGCTCG CGGGGCGCGT
 6851  CCGCGTCGGG TCGGTCGTGA ACGACCCGAG CGACCTCTGC ACCAACGGCA
 6901  CTCAGGGGAT CTCGTGGGCC TCGACGCACA CGGGAGGGCT GGAGGAGTGG
 6951  CTCGCGGCCG CCCTCGGCCA GGACACGGGC TCGGCCGTGG ATGGCTCGGT
 7001  CCCCGCGCGC GAGTGCGAGC ACGCGGTCGA GGCGCGCGTG TGCTCGGCGT
 7051  CCCCCGGCCT GTGGCTCTCC AGCGGCGCGC TCTCGCCCGA GAGGCCCCAC
 7101  GAGTCCCGCG TCGGCACCGG GTGCTCGGCC TGCGCGGCTC GACCCCATCG
 7151  AGCTCGCCTC CGACTTCACC TAGTCCCAGC CCTACGTCGC CCACCCCTCC
 7201  TTCGCCTTCG GTCAACGACC GGCCTCGGCC TCGGGCCCCC TCGATCGGCA
 7251  CCGGCGGTGC CGTGCCGCTG CGTCGATACG CTCGGTCGTA GCATACGCTA
 7301  CGAT
```

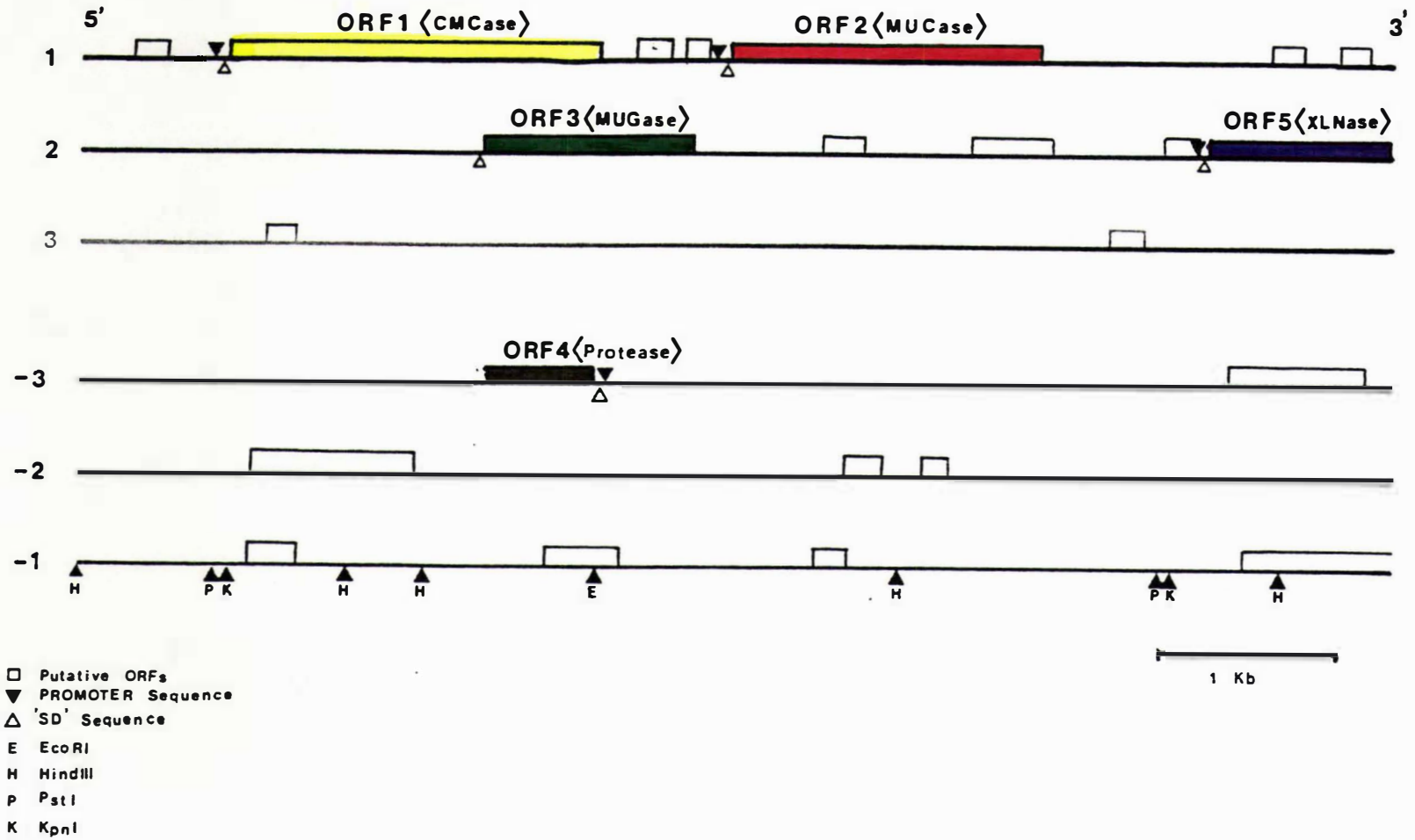


Figure 5-2 Open reading frame identification from 7.3 kb of λCM903 using "Frames" of UWGCG

Fig. 5-3 Nucleotide and deduced amino acid sequences of ORF1

```
712   GCTACTATCTGGATATATTCAAGGAGAGAGCTCGCGCGTCCGATACGATCGTACGGAATA   771
                -35          -10                    SD
772   CGAATTACCGCTGCAGAGGCATTTACAGGGCCTTTATATATACGATTCGGGGATACGATG   831
                                                                M
832   GTAACCAGCGAAAGCGCTCGATCGGGTACCCCATCGAGCTCGCCTCCGACTTCACCTAGT   891
      V  T  S  E  S  A  R  S  G  T  P  S  S  S  P  P  T  S  P  S
892   CCCAGCCCTACGTCGCCCACCCCGCCTTCGCCTTCGACCGAATCCAGCCCTACTCCCTCG   951
      P  S  P  T  S  P  T  P  P  S  P  S  T  E  S  S  P  T  P  S
952   TCGCCAGCGTCACCGCGCTCGCCTACGGTGCGCTGCCGCGAAGCGGTAGCACCTGCGTGC   1011
      S  P  A  S  P  R  S  P  T  V  R  C  R  E  A  V  A  P  A  C
1012  CACGCAGACGCACGTACCGCAGGTCGCGGGCGCGACCCGCGGCACCAATGCGACGCCCGC   1071
      H  A  D  A  R  T  A  G  R  G  R  D  P  R  H  Q  C  D  A  R
1072  CTCCGGGGCCGCCACCGCCCGCGGCTGGGCCACGCGGTGGTACGCGACGACGGCCTGGGT   1131
      L  R  G  R  H  R  P  R  L  G  H  A  V  V  R  D  D  G  L  G
1132  CCACACCAACGCGACTGTGTCCGTGGCGCTTGTCACGGGCCATCTGTTCACCGCGCCGAT   1191
      P  H  Q  R  D  C  V  R  G  A  C  H  G  P  S  V  H  R  A  D
1192  TGGCACCAACCACGGCGTGCTGCACTTGAACTTGTGGGACAACTCGTCCATCCGCAGTCC   1251
      W  H  Q  P  R  R  A  A  L  E  L  V  G  Q  L  V  H  P  Q  S
1252  GGGCTCGTTCCGGTGGTTCGAGGGGTTCCGGCGCCGCCCCCCCGTGATCGTCGGCGAGAC   1311
      G  L  V  P  V  V  R  G  V  P  A  P  P  P  R  D  R  R  R  D
1312  CAATCCCGGCCGCGACGGGCACTGGTTCAGCTCCCCCTTCCGCGGGACCGGGAACAACCA   1371
      Q  S  R  P  R  R  A  L  V  Q  L  P  L  P  R  D  R  E  Q  P
1372  GGCGGTCACCAACCCCACGACGCCGACGCTGTACACGGCGGGGGCCGAGGCGTACGTGGG   1431
      G  G  H  Q  P  H  D  A  D  A  V  H  G  G  G  R  G  V  R  G
1432  CAACTCCAATCCCGCCGTCCTCGGGATCGGCGTGTACGGGACCACCGTGACGGCCCATCC   1491
      Q  L  Q  S  R  R  P  R  D  R  R  V  R  D  H  R  D  G  P  S
1492  GACGGCAGAAGCTTTGACCCCCCCGACCGACAACGGGCTGTGCGACGCCAACGGGGACGA   1551
      D  G  R  S  F  D  P  P  D  R  Q  R  A  V  R  R  Q  R  G  R
1552  AGGCACCGACTCGTGGACCGCCGGGTGCGGCCTCGGGAACAACCGGTGCCGCGGGTTCGC   1611
      R  H  R  L  V  D  R  R  V  R  P  R  E  Q  P  V  P  R  V  R
1612  CAACGCGGTCCGGGGGTGCGGGCGCACGAACACCGCCGACGGCGCGGCCCGCTTGCGGGT   1671
      Q  R  G  P  G  V  R  A  H  E  H  R  R  R  R  G  P  L  A  G
1672  GACCTGGTGGAACGCGCGCAAGCCGCTCACCCGCGTTCCGTGCAGCGCGAACCGATCGAT   1731
      D  L  V  E  R  A  Q  A  A  H  P  R  S  V  Q  R  E  P  I  D
1732  CACGGTGGTCGGGAACTACACCGCGTGGGTGCGCAGGGCCGTGGGGCACGAGAAGATCGT   1791
      H  G  G  R  E  L  H  R  V  G  A  Q  G  R  G  A  R  E  D  R
1792  GGCGACGGTGCTGGTCGAGAACGCGGGCGTGCTCAACGAATCCGGCGCGGTGGTCGGCTA   1851
      G  D  G  A  G  R  E  R  G  R  A  Q  R  I  R  R  G  G  R  L
1852  CACCGAGCGGAAGGCCGACGGGAAGCATCCCTCGGCCTACGAGCGCGGGGCGTGCCTTCC   1911
      H  R  A  E  G  R  R  E  A  S  L  G  L  R  A  R  G  V  P  S
1912  GGTCTCCAGCGACGACAAGCTTCGGCGGTGGAACAACGGCATCCGGGGCGGCGGCGGCCG   1971
      G  L  Q  R  R  Q  A  S  A  V  E  Q  R  H  P  G  R  R  R  P
```

Figure 5-2 (continue)

```
1972  CCGGAAGGACGGGATCTCCAACGTCGCGCCGTACCCGGGTCCCGCGCGCCGCAAGCAGGC  2031
      P  E  G  R  D  L  Q  R  R  A  V  P  G  S  R  A  P  Q  A  G

2032  GCGTCCGGCCAACGGCGTGTGGGGCCGCGGGGCGGCGGCGTCGGCAGGCTACGCGTCTCT  2091
      A  S  G  Q  R  R  V  G  P  R  G  G  G  V  G  R  L  R  V  S

2092  GCGCTACGTGTCGTCCCCCTCCGGCGCCGGTCTGGACGCGATCGTGCGGACTGTCACGCT  2151
      A  L  R  V  V  P  L  R  R  R  S  G  R  D  R  A  D  C  H  A

2152  CGGCCGGGGGTCTCCCACCGGGCTACTCCGCGCGGGCACCCGCACCGGTGTCGGGCTGTG  2211
      R  P  G  V  S  H  R  A  T  P  R  G  H  P  H  R  C  R  A  V

2212  CGGGGCTCCCGGCAAGATGACCGTGCGCGAGACCTACGGCGCCGTCACGTGGACGCGACG  2271
      R  G  S  R  Q  D  D  R  A  R  D  L  R  R  R  H  V  D  A  T

2272  GCGGGCCAGGGTCGGCGGGACCGGCCCCAGCTACGGGTACTGGTACACCAACTGCGGCGG  2331
      A  G  Q  G  R  R  D  R  P  Q  L  R  V  L  V  H  Q  L  R  R

2332  GAAGCTACGGGCCAGCGTCCGGACGCTATGGGAGCTCGCGCGCACCCGATCCACTGGGGC  2391
      E  A  T  G  Q  R  P  D  A  M  G  A  R  A  H  P  I  H  W  G

2392  GGTGTCAGGCTACCCCGCGGGGCGCGGGCTAGCCAAGGACACCGGCGGGACCGCACTGCT  2451
      G  V  R  L  P  R  G  A  R  A  S  Q  G  H  R  R  D  R  T  A

2452  GTCCGGCTACCCCGCGTCCGCCGCGACAGTCGTGCCCAACGAGACACTCGCGCCGACGTC  2511
      V  R  L  P  R  V  R  R  D  S  R  A  Q  R  D  T  R  A  D  V

2512  GTCCGCGCACGTCGGCTACGTACGCGCCGGGGCGCTGCCGGTACTGAGCGAGCTGAGCTA  2571
      V  R  A  R  R  L  R  T  R  R  G  A  A  G  T  E  R  A  E  L

2572  CGTCGCGTCCACGAGCGGCGCCGCCGTCGCGCTACGGTGGGCCGCGTTACGGTCCGAGGC  2631
      R  R  V  H  E  R  R  R  R  R  A  T  V  G  R  V  T  V  R  G

2632  TGCTTCAGGGAGCGCGTCCGAAGCGACCCCCACCGCAGCCACGGTACGCGCGCCCAACAT  2691
      C  F  R  E  R  V  R  S  D  P  H  R  S  H  G  T  R  A  Q  H

2692  GTCCACCGAGCTACTCGCGACGCTGCCGGGGTGCGAGCGGGCCTCGGCGAACAACGCGTC  2751
      V  H  R  A  T  R  D  A  A  G  V  R  A  G  L  G  E  Q  R  V

2752  CACAGGCGCGCCGCGGCGACCCGAGGCACCGGCCGACAAGAAACGCGCCGATCGGGCTCG  2811
      H  R  R  A  A  A  T  R  G  T  G  R  Q  E  T  R  R  S  G  S

2812  GGCGATACGTCCGGCGCGGGACGCGGGGGAGCCATCGCTAGCGGCCAAACGACAGCCTGA  2871
      G  D  T  S  G  A  G  R  G  G  A  I  A  S  G  Q  T  T  A  *

2872  TCTATATCGTACGAATTCTGGATATAAAGCTCTAGCTCGCGTACGAAGTCTACGCGTACA  2931

2932  GCGATACGCGAGCGTAGCTAGCGAGCGGATCGTAGCTGGAGCTTCGCGAGCTGGCAGCAT  2991

2992  CGAGCGCGA  3000
```

Table 5-1 Comparison of consensus of promoter sequence and SD sequence

Bacteria	Gene	SD sequence	Promoter sequence		
			-10	Distance	-35
E. coli/B. subtilis[1]	cons*	AAGAAG	TATAAT	17 bp	TTGACA
Ce. fimi	*cen*A[2]	$^{-13}$AGGGAGCT^{-6}	$^{-58}$TTTCCT^{-53}	16 bp	$^{-82}$TCGCGCCG^{-75}
	cex[3]	$^{-12}$AGGAGGA^{-6}	$^{-105}$TTGGCT^{-100}	16 bp	$^{-127}$TATCGA^{-122}
B. subtilis PAP115[4]	EG*	$^{-19}$AAGGAGG^{-13}	$^{-65}$TACAAT^{-60}	17 bp	$^{-88}$TAGACA^{-83}
C. thermocellum[5]	*cel*A	$^{-12}$AGGAGG^{-7}	$^{-52}$TATAAT^{-47}	22 bp	$^{-77}$TTG^{-75}
F. succinogenes S85[6]	*cel*-3	$^{-11}$AAGAGG^{-5}	$^{-43}$TATA^{-40}	22 bp	$^{-71}$TTTACA^{-66}
Streptomyces sp.[7]	*cas*A	$^{-13}$AGGAAGGA^{-6}	$^{-12}$TTACCGT^{-66}	16 bp	$^{-94}$TTCACC^{-89}
R. flavefaciens	*ren*	$^{-10}$CGGGGA^{-5}	$^{-36}$TTTACA^{-31}	17 bp	$^{-59}$TACGAA^{-54}
	rex	$^{-8}$ATGGGG^{-8}	$^{-31}$TATATA^{-26}	17 bp	$^{-54}$TCGAAG^{-49}
	rbg	$^{-9}$TCCCGG^{-4}	NF		NF

cons, consensus; EG, endoglucanase
From: 1. Hawley and McClure, 1983. 2. Wong et al, 1986. 3. O'Neill et al, 1986.
4. MacKay et al, 1986. 5. Beguin et al, 1985. 6. McGavin et al 1989.
7. Nakai et al, 1988.

estimated on the basis of SDS-PAGE zymograms (42 kDa + 10-13 kDa).

The available nucleotide sequences of other endoglucanase genes were compared to the ORF1 using the programmes FASTA (Pearson and Lipman, 1988). Results of these comparisons showed that there were no obvious homologies to known endoglucanase sequences (*C. thermocellum* celA, B, C, D and E (Beguin et al, 1985; Grepinet and Beguin, 1986; Schwarz et al, 1988; Joliff et al, 1986; Hall et al, 1988), *C. acetobutylicum* (Zappe et al, 1988), *Ce. fimi* cenA (Wong et al, 1986), *Bacillus* sp. (Fukumori et al, 1986; MacKay et al, 1986; Nakamura et al, 1987; Fukumori et al, 1989), *Pseudomonas fluorescens* subsp. *cellulosa* (Hall and Gilbert, 1988), *Erwinia chrysanthemi* (Aymeric et al, 1988), *Agrobacterium* sp. (Wakarchuk et al, 1988), *T. reesei* (Chen et al, 1987; Shoemaker et al, 1983; Teeri et al, 1987) and Avocado (Tucker et al, 1987)). Since the *Streptomyces* casA (Nakai et al, 1988) and *Caldocellum saccharolyticum* cellulase (CMCase and MUCase, Love et al, 1988) nucleotide sequences were not in the data bases, these were added separately. No obvious homology

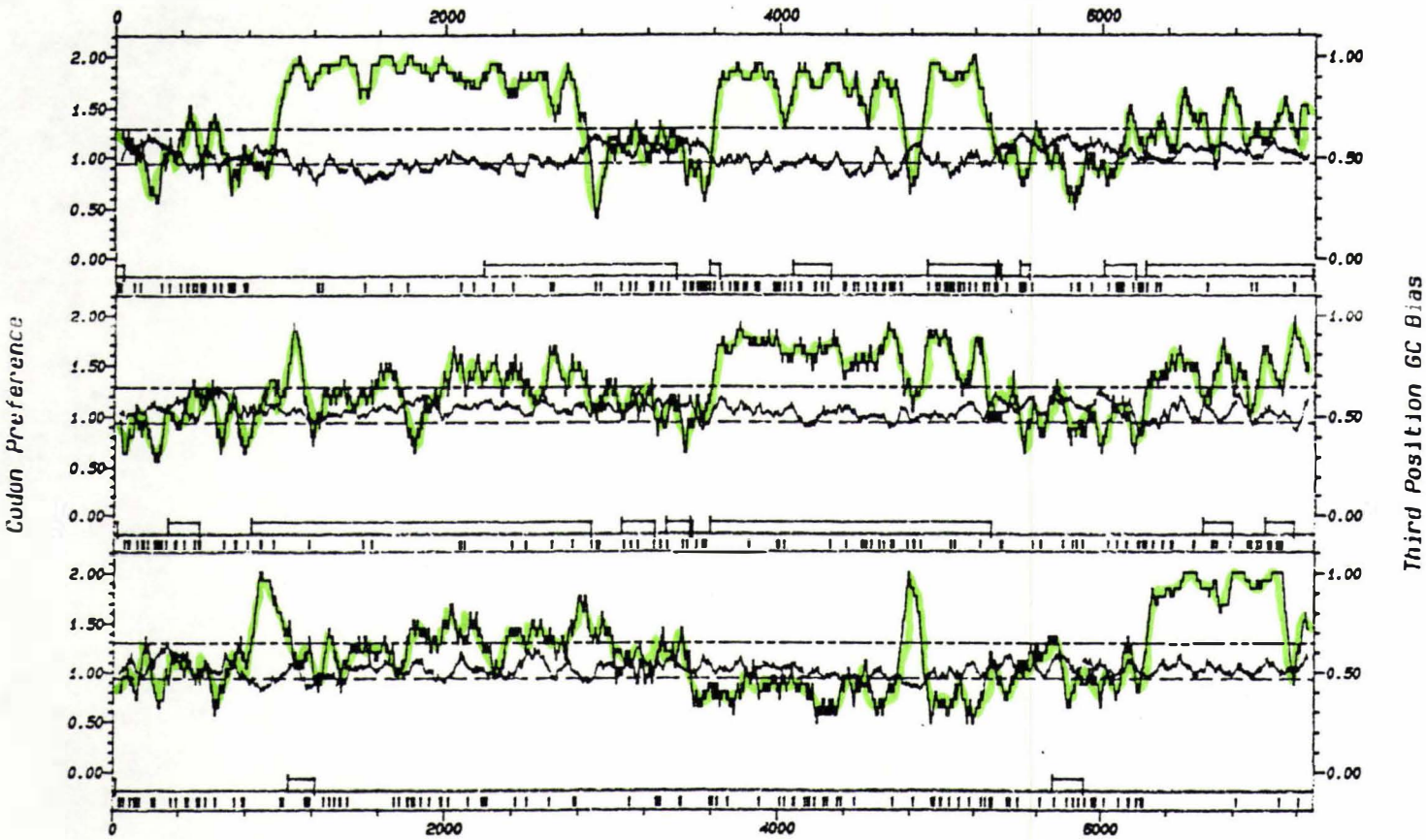


Figure 5-4 G+C bias of 7.3 kb fragment using "Codenpreference" of UWGCG

Table 5-2 codonfrequency of cellulases

AmAcid	Codon	Fraction							
		ORF1	ORF2	ORF3	ORF4	ORF5	CFCEN*	CFCEX*	CSCEL*
Gly	GGG	0.23	0.31	0.38	0.33	0.20	0.14	0.10	0.19
Gly	GGA	0.11	0.15	0.10	0.17	0.10	0.02	0.00	0.27
Gly	GGT	0.22	0.15	0.05	0.33	0.16	0.08	0.08	0.45
Gly	GGC	0.45	0.38	0.47	0.17	0.54	0.77	0.82	0.09
Glu	GAG	0.24	0.82	0.93	1.00	0.43	0.85	0.89	0.44
Glu	GAA	0.76	0.18	0.07	0.00	0.57	0.15	0.11	0.56
Asp	GAT	0.25	0.12	0.55	0.60	0.25	0.00	0.00	0.68
Asp	GAC	0.75	0.88	0.45	0.40	0.75	1.00	1.00	0.32
Val	GTG	0.37	0.39	0.15	0.16	0.47	0.26	0.33	0.09
Val	GTA	0.17	0.08	0.42	0.58	0.00	0.00	0.00	0.41
Val	GTT	0.13	0.04	0.03	0.11	0.00	0.00	0.00	0.40
Val	GTC	0.33	0.49	0.39	0.16	0.53	0.74	0.67	0.09
Ala	GCG	0.34	0.25	0.51	0.45	0.27	0.53	0.51	0.23
Ala	GCA	0.16	0.29	0.09	0.14	0.20	0.03	0.03	0.41
Ala	GCT	0.33	0.22	0.17	0.28	0.37	0.03	0.03	0.28
Ala	GCC	0.17	0.24	0.23	0.14	0.17	0.41	0.44	0.08
Arg	AGG	0.04	0.16	0.02	1.00	0.02	0.07	0.06	0.33
Arg	AGA	0.01	0.07	0.05	0.00	0.02	0.03	0.00	0.41
Ser	AGT	0.05	0.00	0.04	0.08	0.14	0.02	0.00	0.35
Ser	AGC	0.23	0.25	0.35	0.19	0.14	0.23	0.31	0.2
Lys	AAG	0.00	0.75	0.57	1.00	0.00	0.93	0.95	0.30
Lys	AAA	0.00	0.25	0.43	0.00	0.00	0.07	0.05	0.70
Asn	AAT	0.00	0.00	0.29	1.00	0.00	0.00	0.00	0.69
Asn	AAC	0.00	1.00	0.71	0.00	1.00	1.00	1.00	0.31
Met	ATG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Ile	ATA	0.40	0.17	0.43	0.40	0.50	0.00	0.00	0.45
Ile	ATT	0.10	0.17	0.00	0.00	0.00	0.08	0.09	0.49
Ile	ATC	0.50	0.67	0.57	0.60	0.50	0.92	0.91	0.06
Thr	ACG	0.40	0.44	0.44	0.50	0.25	0.42	0.54	0.17
Thr	ACA	0.07	0.22	0.09	0.19	0.00	0.00	0.02	0.62
Thr	ACT	0.27	0.06	0.09	0.00	0.13	0.02	0.00	0.17
Thr	ACC	0.27	0.28	0.38	0.31	0.63	0.57	0.44	0.04
Trp	TGG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
End	Tmx	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.32
Cys	TGT	0.50	0.40	0.00	0.83	0.00	0.14	0.00	0.86
Cys	TGC	0.50	0.60	1.00	0.17	1.00	0.86	1.00	0.14
End	TAG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07
End	TAA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.61
Tyr	TAT	0.67	0.33	0.22	0.50	0.33	0.00	0.09	0.67
Tyr	TAC	0.33	0.67	0.78	0.50	0.67	1.00	0.91	0.33
Leu	TTG	0.00	0.05	0.03	0.00	0.00	0.03	0.00	0.26
Leu	TTA	0.03	0.00	0.09	0.00	0.11	0.00	0.00	0.33
Phe	TTT	0.50	0.33	0.00	0.00	0.50	0.00	0.00	0.84
Phe	TTC	0.50	0.67	1.00	1.00	0.50	1.00	1.00	0.16

Table 5-2 (continue)

AmAcid	Codon	Fraction							
		ORF1	ORF2	ORF3	ORF4	ORF5	CFCEN*	CFCEX*	CSCEL*
Ser	TCG	0.30	0.42	0.25	0.35	0.38	0.41	0.44	0.07
Ser	TCA	0.07	0.11	0.06	0.08	0.14	0.02	0.00	0.14
Ser	TCT	0.07	0.04	0.06	0.08	0.05	0.00	0.00	0.14
Ser	TCC	0.30	0.18	0.23	0.23	0.14	0.32	0.25	0.06
Arg	CGG	0.32	0.23	0.25	0.36	0.33	0.17	0.31	0.04
Arg	CGA	0.18	0.18	0.22	0.15	0.17	0.10	0.11	0.14
Arg	CGT	0.15	0.09	0.15	0.15	0.14	0.00	0.11	0.08
Arg	CGC	0.30	0.27	0.31	0.33	0.31	0.62	0.40	0.00
Gln	CAG	0.22	0.80	1.00	1.00	0.41	1.00	1.00	0.47
Gln	CAA	0.78	0.20	0.00	0.00	0.59	0.00	0.00	0.52
His	CAT	0.08	0.09	0.00	0.00	0.00	0.25	0.08	0.75
His	CAC	0.92	0.91	1.00	0.00	1.00	0.75	0.92	0.25
Leu	CTG	0.18	0.21	0.29	0.45	0.21	0.38	0.29	0.13
Leu	CTA	0.32	0.08	0.41	0.36	0.00	0.00	0.00	0.14
Leu	CTT	0.12	0.08	0.00	0.18	0.00	0.00	0.00	0.12
Leu	CTC	0.35	0.57	0.18	0.00	0.68	0.59	0.71	0.01
Pro	CCG	0.45	0.43	0.27	0.40	0.15	0.67	0.59	0.28
Pro	CCA	0.11	0.09	0.09	0.00	0.28	0.06	0.05	0.49
Pro	CCT	0.19	0.20	0.09	0.12	0.38	0.00	0.05	0.22
Pro	CCC	0.24	0.29	0.55	0.48	0.20	0.28	0.31	0.01

*, CFCEN, Ce. *fimi cen* A (Wong et al, 1986); CFCEX, Ce. *fimi cex* (O'Neill et al, 1986); CSCEL, Ca. *saccharolyticum* cellulase gene (personal contact, Soul).

was detected to the casA nucleotide sequence from *Streptomyces* or the cellulase gene of *Caldocellum*. Also, no obvious homology was detected to any DNA sequence in the Genbank (V. 60.0) or European Molecular Biology Laboratory (EMBL: V. 19.0) DNA sequence data or to the National Biochemical Research Foundation (NBRF) Nucleotide sequence Data Library (V. 35.0).

The deduced amino acid sequence of the ORF1 was also compared to the amino acid sequences of the other cellulase proteins and those within the NBRF resource data base (release 21.0) using FASTA. No obvious homology was detected at the amino acid level. A similar result was found when comparing the N-terminal amino acid sequence of Exo A of *R. flavefaciens* FD-1 (Gardner et al, 1987) with ORF1 using BESTFIT. The protein

with the greatest similarity to renA was that encoded by the cenA gene of *Ce. fimi* with 32.1 % similarity in 485 a.a. residues (Fig. 5-5). The similarity with other cellulase genes is summarized in Table 5-3.

An unusual sequence region from position 853 to position 978, 22 bp down stream of putative initiated codon (ATG), coded for a polypeptide about 42 amino acids long which was composed mainly of proline, threonine and serine (PTS; 38 out of 42 a.a.). This region was located close to the N-terminal end of ORF1 (Fig. 5-3).

5-2-3-2 Second open reading frame

The second ORF was identified as the structural gene for the *R. flavefaciens* exoglucanase gene (designated *rex*: *Ruminococcus* exoglucanase) since that region of the deletion map corresponded to exoglucanase activity. The sequence of this ORF, the 5′ upstream sequence and the deduced amino acid sequence is shown in Fig. 5-6. This ORF consisted of a coding sequence of 1821 bp, starting with an ATG codon at position 3586 and ending with a TAA stop codon at position 5304.

The nucleotide sequence upstream of the 5′ end of ORF2 was screened for SD and potential promoter sequences. A SD sequence (ATGGGG) containing 4 C/G's was located at position -8 to -3, 3 bp upstream of the ATG initiation codon (Fig. 5-6, Table 5-1). Also sequences TATATA and TCGAAG was the putative -10 and -35 promoter sequence (Fig. 5-6, Table 5-1). Several inverted repeat sequences were identified downstream of the stop codon (Fig. 5-6). The nucleotide sequence of the ORF contained 67.3% G+C showing a strong bias towards codons containing dG+dC (Table 5-2; Fig. 5-4).

Fig. 5-5 Comparison of ORF1 (*ren*A) with *cen*A of *Ce. fimi* using 'Bestfit' of UWGCG

```
renA    1 MVTSESARSGTPSSSPPTSPSPSPTSPTPPSPSTESSPTPSSPASPRSPT 50
          | |. .| .  :....:.:. .. |  |.:.....:.:... .....:...
cenA    1 MSTRRTAAALLAAAAVAVGGLTALT..TTAAQAAPGCRVDYAVTNQWPGG 48

renA   51 VRCREAVAPACHA.DARTAGRGRDPRHQCDARLRGRHRPRLGHAVVRDDG 99
          . .. .:.  ..: ..:. ::. .: :..:. :.|  ... |:. | . .
cenA   49 FGANVTITNLGDPVSSWKLDWTYTAGQRIQQLWNGTASTNGGQVSVTSLP 98

renA  100 LGPHQRDCVRGACHGPSVHRADWHQPRRAALELVGQLVHPQSGLVPVVRG 149
          :..  .... .|: | .. :|:.: | .|.:.| |  .   .| ||....
cenA   99 WNGSIPTGG.TASFGFNGSWAGSN.PTPASFSLNGTTC...TGTVPTTSP 143

renA  150 VPAPPPRDRRRDQSRPRRALVQLPLPRDRE...QPGGHQPHDADAVHGGG 196
          .|.|.| .. ..... : . :. ....:   :|..:. :. :|. |.:
cenA  144 TPTPTPTTPTPTPTPTPTPTVTPQPTSGFYVDPTTQGYRAWQAASGTD 193

renA  197 RGVRGQLQSRRPRDRRVRDHRDGPSDGRSFDPPDRQRAVRRQRGRRHRLV 246
          ::: :.: .  |.. :| :::.::|.:..  :.  .|||  . |: ..||
cenA  194 KALLEKI.ALTPQAYWV.GNWADASHAQAEVADYTGRAV..AAGKTPMLV 239

renA  247 DRRVRPREQPVPRVRQRGPGVRAHEHRRRRGPLAGDLVERAQAAHPRSVQ 296
             :..|:    ..:.|.||.. |. |: :.:|.:: :.: .  ....
cenA  240 VYAIPGRD....CGSHSGGGVSESEYARWVDTVAQGIKGNPIVILEPDAL 285

renA  297 REPIDHGGRELHRVGAQGRGAREDRGDGAG..RERGRAQRIR...RGGRL 341
            :  | :|.: .|||    :|:.   .||    : |:|.::.   ...||
cenA  286 AQLGDCSGQG.DRVGFLKYAAKSLTLKGARVYIDAGHAKWLSVDTPVNRL 334

renA  342 HRAEGRREASLGLRARGVPSGLQRRQ.ASAVEQRHPGRRRPPEGRDLQRR 390
          :..:   ..:::|.... ... :.:. : .:.|| .|:: . :..  ..
cenA  335 NQVGFEYAVGFALNTSNYQTTADSKAYGQQISQRLGGKKFVIDTSRNGNG 384

renA  391 AVPGSRAPQAGASGQRRVGPRGGGVGRLRVSALRVVPLRRRSGRDRADCH 440
          . .:. .|.: | |:|.|:..:|       |:| .: : : .| . :.|:
cenA  385 SNGEWCNPRGRALGERPVAVNDG.......SGLDALLWVKLPGESDGACN 427

renA  441 ARPGVSHRATPRGHPHRCRAVRGSRQDDRARDLRR 475
          : |:.:::             :.. . :.||: |:
cenA  428 GGPAAGQW.............WQEIALEMARNARW 449
```

|,:,., different similarity score of amino acid, from high to low.

Table 5-3 Summary of homology percentage of *ren* with other cellulase gene

Name	Code	Length*	Homology (%)
Bacillus sp. (strain 1139)	endoglucanase	800	28.5
Bacillus sp. (pNK1)	endoglucanase	488	29.3
Bacillus sp. (pNK2)	endoglucanase	409	28.9
B. subtilis	endoglucanase	508	29.7
B. subtilis	endoglucanase	499	29.9
Clostridium thermocellum	*cel*A	477	29.0
C. thermocellum	*cel*B	563	27.7
C. thermocellum	*cel*D	649	26.9
C. acebutylicum	endoglucanase	448	27.3
Cellulomonas fimi	*cen*A	449	32.1
Trichoderma reesei	EGI	459	27.9
T. reesei	EGIII	471	30.9

*, amino acid length

The deduced protein size of this ORF was 572 amino acid residues with an estimated MW of 67 kDa. This predicted molecular weight of the translated protein was smaller than the size estimated on the basis of SDS-PAGE (90 kd).

The nucleotide sequences of other exoglucanase genes were compared to the ORF2 which showed no obvious homologies to the exoglucanase sequences *Ce. fimi* cex (O'Neill, et al, 1986), *Agrobacterium* sp. (Greenbert et al, 1988), *T. reesei* CBHI (Shoemaker et al, 1983) and *Caldocellum saccharolyticum* cellulase (CMCase and MUCase, Love et al, 1988). Also no obvious homology was detected to any DNA sequence in the Genbank, EMBL DNA sequence data or NBRF Nucleotide sequence Data Library.

When compared with other exoglucanase proteins, the ORF2 deduced amino acid sequence showed no obvious homologies. A similar result was found when comparing the N-terminal amino acid sequence of Exo A of *R. flavefaciens* FD-1 (Gardner et al, 1987) with ORF2. The best similarity of exoglucanase gene was

Fig. 5-6 Nucleotide and deduced amino acid sequences of ORF2

```
                                                                  -35
3481  CTCCAGCAAGCTACGACACTCTTGCAATTTAAGCGTAGCAACGTTAGCGAATCGAAGGAAAGAATCGTAA  3550
                -10                    SD
3551  AGATTATATACAGAGCGTGTAAAATCTATGGGGCGATGCGGCAAGACACCCAGAGCGCTCAGCGTGAGCG  3620
                                        M  R  Q  D  T  Q  S  A  Q  R  E  R

3621  GGTCCACGACCTCGACCTAGACCCGGTCAGCGGGGTCCTCCGGCTCGCTTCGGACGAGCAGCACGGCCGA  3690
      V  H  D  L  D  L  D  P  V  S  G  V  L  R  L  A  S  D  E  Q  H  G  R

3691  TCCGTTGACGGTCTCGACCAGCACGTAGCCGTGGAGGGGCGGGACCACGACCTCTCGGCCCACGACCTGC  3760
      S  V  D  G  L  D  Q  H  V  A  V  E  G  R  D  H  D  L  S  A  H  D  L  L

3761  TCGGCGTGGCGGCTCGCGTCTCGGACGAGCCGCTACTCTACCGGGACGATCCCGCGGACAGGGAGCAGCC  3830
        G  V  A  A  R  V  S  D  E  P  L  L  Y  R  D  D  P  A  D  R  E  Q  P

3831  CCGCCCGGACGGGCGACAGGCTGTCCCGGAGGAGGTCGACCAGGGGGCAGCGAGCGACGGTCACGACCTC  3900
      R  P  D  G  R  Q  A  V  P  E  E  V  D  Q  G  A  A  S  D  G  H  D  L

3901  CGCGTCGACGCGCTCCAGCACTCCGGACTGTCCCAACAGCACGACCGCGGGCGCCCGGAGGTCGCGCGAC  3970
      R  V  D  A  L  Q  H  S  G  L  S  Q  Q  H  D  R  G  R  P  E  V  A  R  H

3971  ACCTTGACGAGGACGGCAGGACGGTCCTCAGGGCCGACTCGCAAGACGACACACAGGACTCGAGACTCCA  4040
        L  D  E  D  G  R  T  V  L  R  A  D  S  Q  D  D  T  Q  D  S  R  L  Q

4041  GCTTCTAGCGTGGTGCCGGCTCCTCACGGTCCGCGGCCATGTCGACCCGGTGGTGGCTCACACAGGCCGG  4110
      L  L  A  W  C  R  L  L  T  V  R  G  H  V  D  P  V  V  A  H  T  G  R

4111  CAGGCCGACGAACAGGACGCCGGCGACGAGCTGCCCAGCGTGCACCAAGCCGAAGACGGCGGCGCCCTTT  4180
      Q  A  D  E  Q  D  A  G  D  E  L  P  S  V  H  Q  A  E  D  G  G  A  L  S

4181  CCCGTCAGCCCGATCACCTGGACCTCAACGTCTCCCGCGACGATCATTCGGTCGAGCGGCGACTGGACCA  4250
        R  Q  P  D  H  L  D  L  N  V  S  R  D  D  H  S  V  E  R  R  L  D  Q

4251  GGACTCGCACGACCTCCAGCGTGCCCTCCAAGACCAGGACCGGCTCGCGGCAGTCCTGAGACTCGACGGC  4320
      D  S  H  D  L  Q  R  A  L  Q  D  Q  D  R  L  A  A  V  L  R  L  D  G

4321  GACCACCTCGACGACCCGGTGGGCGTCCAGCTGCCGCGACAAGACCACGACGACTCCGCACAGGATCACC  4390
      D  H  L  D  D  P  V  G  V  Q  L  P  R  Q  D  H  D  D  S  A  Q  D  H  R

4391  GATCGCAGGGCGAGCGATTGGCATCGGGACACCACCCGGACCTCCGTTCGGAGCAGGTCACCGCAGCGCT  4460
        S  Q  G  E  R  L  A  S  G  H  H  P  D  L  R  S  E  Q  V  T  A  A  L

4461  CCTGGAAGACGAGGCACACGTGGTGGGGCACCAAGCATCCTCAACGAGGGTGGACTCGCCGGACCAGGAG  4530
      L  E  D  E  A  H  V  V  G  H  Q  A  S  S  T  R  V  D  S  P  D  Q  E

4531  AGGGGTGACCACCCGGATCGAAGCTTCGCAAGGCTCGAGGACGTCCAGGGGGACGAAGCACGGCTCGACG  4600
      R  G  D  H  P  D  R  S  F  A  R  L  E  D  V  Q  G  D  E  A  R  L  D  V

4601  TGGACCAATCACAGCTCGACCCGCAGGACGGGCTCAGGGTCGCGGACCCTGACCTGTCGGTGGAGGACCA  4670
        D  Q  S  Q  L  D  P  Q  D  G  L  R  V  A  D  P  D  L  S  V  E  D  H

4671  CGGGCAGAGGGACGCCCTGTGCCAGTCAGACCCCCAGCGGGTCGAGCGCGCTGACGGGCTTGACGTCTGC  4740
      G  Q  R  D  A  L  C  Q  S  D  P  Q  R  V  E  R  A  D  G  L  D  V  C

4741  CTCCGACACTGTTCGACCGCACCCAGCTCGTCGCCTCCATCTACTCCTACGCCCAGCAGCACGACCCCCA  4810
      L  R  H  C  S  T  A  P  S  S  S  P  P  S  T  P  T  P  S  S  T  T  P  T

4811  CCCCTAGCTCCTCAGTATCTTCCGTACCTACACCGAGCTCGCCCGATTCACCGTCGCCTACGGACCACGC  4880
        P  S  S  S  V  S  S  V  P  T  P  S  S  P  D  S  P  S  P  T  D  H  A

4881  CCGCTCGGGGGACCAGCATGAGGTGGACCTGCACGACCAGCGCGGCGACGCGGCCGGTCTCCACGACCAG  4950
      R  S  G  D  Q  H  E  V  D  L  H  D  Q  R  G  D  A  A  G  L  H  D  Q

4951  GTGGATCACTCGCTGGGCGAGGACGTGCGCCTCGCCGACGGAGACGGAGACCAGTCCCTCCGGCAGCAGG  5020
      V  D  H  S  L  G  E  D  V  R  L  A  D  G  D  G  D  Q  S  L  R  Q  Q  E

5021  AGCTCAGCCACGACGAGCTCAGGGTGGGAGAGCAGCTTGAGGCTGTCGACCCGGATCGCTCAGCGCAGGG  5090
        L  S  H  D  E  L  R  V  G  E  Q  L  E  A  V  D  P  D  R  S  A  Q  G

5091  TGTGGAGCAGGACGCACCTCTGTCGCAGCTCCAGGCTGTGGCACAATCGGACGCAGCTGACCTACAGCAG  5160
      V  E  Q  D  A  P  L  S  Q  L  Q  A  V  A  Q  S  D  A  A  D  L  Q  Q

5161  GTGCGCGACGTCGATCACGACGACGTCGTGGCTGGTGAGGACGGCGGCCGCGACCACGAGCAGCTCCACC  5230
      V  R  D  V  D  H  D  D  V  V  A  G  E  D  G  G  R  D  H  E  Q  L  H  R

5231  GCTTGGACACGATCCCAGTCTTCGATGAGGCACACCCCCAGCCAGGAGCAATCAGCGATCGGCTACGATC  5300
        L  D  T  I  P  V  F  D  E  A  H  P  Q  P  G  A  I  S  D  R  L  R  S
```

Figure 5-6 (continue)

```
5301  GTAAGGCTAAGCGACTTATCAGGCGCTAGCGTCGATTAGCGATGCGATACGCATCGGTATATTAGCGCGC  5370
                                                                    *

5371  GAATTAAATTTCGGGCCCTATTTTAATCGGCGCTA  5405
```

Fig. 5-7 Comparison of ORF2 (*rex*) with other exoglucanase genes

```
          2                                                            51
 cons  R..ta.aq.a r.....l... ..p..t.... as.....q.. ..l...a..a
  Cex  RTTPAPGHPA RGARTALRTT RRRAATLVVG ATVVLPAQAA TTLKEAADGA
 Agro                                                  MTDPNTLA
  Tri    ATARAQSA CTLQSET... .HPPLTWQKC SSGGTCTQQT GSVVIDANWR
 Orf2  RQDTQSAQRE RVHDLDL... .DPVSGVLRL ASDEQHGRSV DGLDQHVAVE

          52                                                           101
 cons  grd..fs.d. ..gvatr.s. ...l..d..t .a....pdg. .aa..e....
  Cex  GRDFGFALD. .....PNRLSE AQY....... .........K AIADSEFNLV
 Agro  AR...FPGDF LFGVAT.... .ASFQIEGST KADGRKPSIW DAFCNMPGHV
  Tri  WTHATNSSTN CYDGNTWSST ...LCPDNET CAKNCCLDG. ..AAYASTYG
 Orf2  GRDHDLSAHD LLGVAARVSD EPLLYRDDPA DREQPRPDGR QAVPEEVDQG

          102                                                          151
 cons  va..g.dl.. .a...s..s. ...rg...vA s.....g.tl l..d..gdt.
  Cex  VAENA..MKW DATEPSQNSF SFGAGDR.VA SYAADTGKEL Y.....GHTL
 Agro  FGRHNGDIAC DHYNRWEEDL DLIKEMGVEA YRFSL.AWPR IIPDGFGPIN
  Tri  VTTSGNSLSI GFVTQSAQKN VGARLY.LMA SDTTYQEFTL LGNEFSFDVD
 Orf2  AASDGHDLRV DALQHSGLSQ QHDRGRPEVA RHLDEDGRTV LRADSQDDTQ

          152                                                          201
 cons  vs.lql.dw. .ll.vs..vd g.vah..k.a .......d.. ..v..a.dgg
  Cex  VWHSQLPDWA RNLNGSAFES AMVNHVTKVA DHFEGKVASW DVVNEAFADG
 Agro  EKGLDFYDR. .......LVD GCKARGIKTY AT.LYHWD.. LPLTLMGDGG
  Tri  VSQLPCGLNG ALYFVSMDAD GGVSKYPTNT AG........ .....AKYGT
 Orf2  DSRLQLLAWC RLLTVRGHVD PVVAHTGRQA DE.QDAGDEL PSVHQAEDGG

          202                                                          251
 cons  ..srs..h.. ..f.r.d..v e.rlq..s.. ........dr l.avl.i...
  Cex  .......... .DGPPQDSAF QQKLGNGYIE TAFRAARAAD PTAKLCI.ND
 Agro  WASRSTAH.. .AFQRYAKTV MARLG..... ........DR LDAV......
  Tri  GYCDSQCPRD LKFINGQANV EGW.EPSSNN ANTGIGGHGS CCSEMDIWEA
 Orf2  ALSRQPDHLD LNVSRDDHSV ERRLDQDSHD LQRALQDQDR LAAVLRL.DG

          252                                                          301
 cons  ....ep.g.. ..s....d.a ...r..g... ..rla..ghh p...dl...q
  Cex  YNVE...GIN AKSNSLYDLV KDFKARG... .VPLDCVGFQ S...HLIVGQ
 Agro  ATFNEPWCAV WLSHLYGVHA PGERNME... .AALAA.MHH I...NL....
  Tri  NSISEALTPH PCTTVGQEIC EGDGCGGTYS DNRYGG.TCD PDGCDWNPYR
 Orf2  DHLDDPVGVQ LPRQDHDDSA QDHRSQG... .ERLAS.GHH P...DLRSEQ

          302                                                          351
 cons  v........a h...fgvdas .t..d....e .sda..ra.. ......q..e
  Cex  VPGDFRQNLQ RFADLGVDVR ITELDIRMRT PSDATKLATQ AAD.YKKVVQ
 Agro  .........A HG..FGVEAS RHVAPKVPVG LVLNAHSAIP .....ASDGE
  Tri  LGNTSFYGPG SS..FTLDTT KKLTVVTQFE TSGAINRYYV QNGVTFQQPN
 Orf2  VTAALLEDEA HV..VGHQAS STRVDSPDQE RGDHPDRSFA RLE.DVQGDE

          352                                                          401
 cons  A.l.v...q. ...dg..... ..v.dp.fgg e......dal .q..pa.v..
  Cex  ACMQVTRCQG VTVWGITDKY SWVPDVFPGE GAALV.WDAS YAKKPAYAA.
 Agro  ADLKAAERAF QFHNGA.... ..FFDPVFKG EYPAEMMEAL ..GDRMPVV.
  Tri  AELGSYSGNE INDDYC.... .TAEEAEFGG SSFSD.KGGL TQFKKATSGG
 Orf2  ARLDVDQSQL DPQDGL.... .RVADPDLSV EDHGQ.RDAL CQSDPQRVE.

          402                                                          451
 cons  ........a. .lq..l.... ......stap ...tpsTP.p ..ttptpss.
  Cex  .......VME AFG..... ........AS PTPTPTTPTP TPTTPT....
 Agro  .......EAE DLGIISQKLD WWGLNYYTPM RVADDATPGV EFPATMPAPA
  Tri  MVLVMSLWDD YYANMLW... ....LDSTYP TNETSSTPGA VRGSCSTSSG
 Orf2  .......RAD GLDVCLR... ....HCSTAP SSSPPSTPTP SSTTPTPSSS
```

Figure 5-7 (continue)

```
          452                                                  501
cons      v...ptp.s. esp.p..h.. ..nq.e..dl .........a g...gvd.g.
 Cex      ....PTPTS. .GPAGCQVLW GVNQWN.TGF TA..NVTVKN TSSAPVD.GW
Agro      VSDVKTDIGW EVYAPALHTL VETLYERYDL PECYITENGA CYNMGVENGE
 Tri      V.....PAQV ESQSPNAKVT FSNIKF.GPI GS..TGNPSG GNPPGGNRGT
Orf2      VSSVPTPSSP DSPSPTDHAR SGDQHE.VDL HD..QRGDAA GLHDQVDHSL

          502                                                  551
cons      t...r.adg. g..s.....l shdg..... ....v....e .....rs..g
 Cex      TLTFSFPSGQ ..........  ..........  ...QVTQAWS STVTQSGSAV
Agro      VNDQPRLDYY AEHLGIVADL IRDGYPMRGY FAWSLMDNFE WAEGYRMRFG
 Tri      TTTRRPATTT G..SSPGPTQ SHYG..... .........Q CGGIGYSGPT
Orf2      GEDVRLADGD GDQSLRQQEL SHDE..... ..LRVGEQLE AVDPDRSAQG

          552                                                  601
cons      vv..ap..qv ..v..sg... a.l..... ..gv..g..G .........L
 Cex      TVRNAPWN.. GSIPAGGT.. AQF..... ..GFNGSHTG TNAAPTAFSL
Agro      LVHVDYETQV RTVKNSGKWY SALASGFPKG NHGVAKG
 Tri      VCASGTTCQV LNPYYSQC
Orf2      VEQDAPLSQL QAVAQSDA.. ADLQQVRDVD HDDVVAGEDG GRDHEQLHRL

          602
cons      ...P.
 Cex      NGTPC
Orf2      DTIPV
```

Cex, *Ce. fimi*; Agro, *Agrobacterium* sp.; Tri, *Trichoderma reesei* CBHI

Ce. fimi cex gene with 33.8% homology and 567 a.a. residues (Fig. 5-7). The percent similarity with other exoglucanases was *T. reesei*, 29.2% and *Agrobacterium* sp., 31.8%.

A proline, threonine and serine rich region similar to the PTS rich region of *ren*A, from position 389 to position 429, about 40 amino acids, was also found. This region was 433 bp away from the C-terminal and formed a highly conserved sequencewhen compared with the corresponding regions of ORF1 and ORF2 (Fig. 5-8). This ORF could be divided into three domains: N-terminal domain, PTS domain and C-terminal domain.

5-2-3-3 Third open reading frame

The third ORF was an overlapping reading frame with ORF1, transcribed using the second reading frame and identified as the structural gene for the *R. flavefaciens* β-glucosidase (designated *rbg* : *Ruminococcus* β-glucosidase) because of its correspondence to activity of the deletion subclones. The sequence of this ORF together with the 5′ upstream sequences and the deduced amino acid sequence is shown in Fig.5-9. This

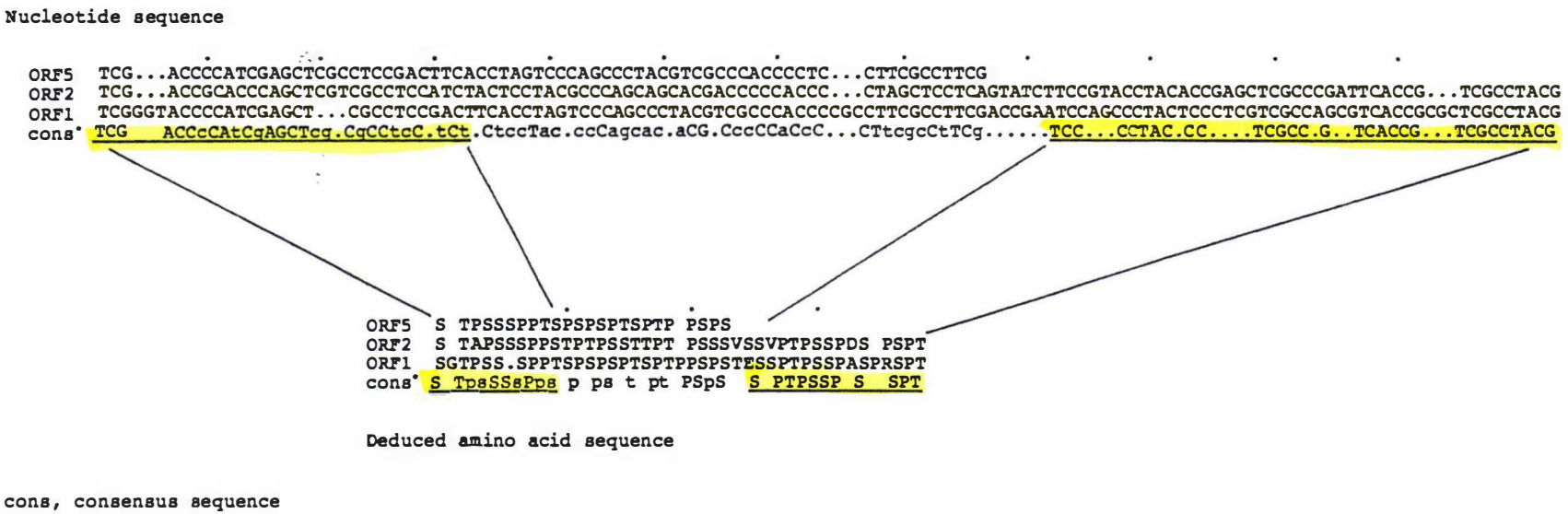


Figure 5-8 Alignment with PTS rich regions of ORF1, ORF2 and ORF5

Fig. 5-9 Nucleotide and deduced amino acid sequences of ORF3

```
2072  TCGGCAGGCTACGCGTCTCTGCGCTACGTGTCGTCCCCTCCGGCGCCGGTCTGGACGCGATCGTGCGGA  2141

2142  CTGTCACGCTCGGCCGGGGGTCTCCCACCGGGCTACTCCGCGCGGCACCCGCACCGGTGTCGGGCTGTG  2211

            SD
2212  CGGGGCTCCCGGCAAGATGACCGTGCGCGAGACCTACGGCGCCGTCACGTGGACGCGACGGCGGGCCAGG  2281
                        M  T  V  R  E  T  Y  G  A  V  T  W  T  R  R  R  A  R

2282  GTCGGCGGGACCGGCCCCAGCTACGGGTACTGGTACACCAACTGCGGCGGGAAGCTACGGGCCAGCGTCC  2351
      V  G  G  T  G  P  S  Y  G  Y  W  Y  T  N  C  G  G  K  L  R  A  S  V  R

2352  GGACGCTATGGGAGCTCGCGCGCACCCGATCCACTGGGGCGGTGTCAGGCTACCCCGCGGGGCGCGGGCT  2421
        T  L  W  E  L  A  R  T  R  S  T  G  A  V  S  G  Y  P  A  G  R  G  L

2422  AGCCAAGGACACCGGCGGGACCGCACTGCTGTCCGGCTACCCCGCGTCCGCCGCGACAGTCGTGCCCAAC  2491
       A  K  D  T  G  G  T  A  L  L  S  G  Y  P  A  S  A  A  T  V  V  P  N

2492  GAGACACTCGCGCCGACGTCGTCCGCGCACGTCGGCTACGTACGCGCCGGGGCGCTGCCGGTACTGAGCG  2561
      E  T  L  A  P  T  S  S  A  H  V  G  Y  V  R  A  G  A  L  P  V  L  S  E

2562  AGCTGAGCTACGTCGCGTCCACGAGCGGCGCCGCCGTCGCGCTACGGTGGGCCGCGTTACGGTCCGAGGC  2631
        L  S  Y  V  A  S  T  S  G  A  A  V  A  L  R  W  A  A  L  R  S  E  A

2632  TGCTTCAGGGAGCGCGTCCGAAGCGACCCCCACCGCAGCCACGGTACGCGCGCCCAACATGTCCACCGAG  2701
       A  S  G  S  A  S  E  A  T  P  T  A  A  T  V  R  A  P  N  M  S  T  E

2702  CTACTCGCGACGCTGCCGGGGTGCGAGCGGGCCTCGGCGAACAACGCGTCCACAGGCGCGCCGCGGCGAC  2771
      L  L  A  T  L  P  G  C  E  R  A  S  A  N  N  A  S  T  G  A  P  R  R  P

2772  CCGAGGCACCGGCCGACAAGAAACGCGCCGATCGGGCTCGGGCGATACGTCCGGCGCGGGACGCGGGGA  2841
        E  A  P  A  D  K  K  R  A  D  R  A  R  A  I  R  P  A  R  D  A  G  E

2842  GCCATCGCTAGCGGCCAAACGACAGCCTGATCTATATCGTACGAATTCTGGATATAAAGCTCTAGCTCGC  2911
       P  S  L  A  A  K  R  Q  P  D  L  Y  R  T  N  S  G  Y  K  A  L  A  R

2912  GTACGAAGTCTACGCGTACAGCGATACGCGAGCGTAGCTAGCGAGCGGATCGTAGCTGGAGCTTCGCGAG  2981
      V  R  S  L  R  V  Q  R  Y  A  S  V  A  S  E  R  I  V  A  G  A  S  R  A

2982  CTGGCAGCATCGAGCGCGATCGAGCGATACGAGCGATCGTATCGACGTTCGATCCATCGAGTACGATCGT  3051
        G  S  I  E  R  D  R  A  I  R  A  I  V  S  T  F  D  P  S  S  T  I  V

3052  ATGCTCGATCGTACGAGTCGTACGTATCGTTCGAGCGTACTTACCCCTGAGCACGCGTCAGGCAGCGAGC  3121
       C  S  I  V  R  V  V  R  I  V  R  A  Y  L  P  L  S  T  R  Q  A  A  S

3122  TACAGCAGAGCTATCGCGAGCAGCGATCGAGTCGAGCTATACTATATATCGCGCTATATATACGGCGGCC  3191
      Y  S  R  A  I  A  S  S  D  R  V  E  L  Y  Y  I  S  R  Y  I  Y  G  G  R

3192  GTCGCGCGCCTACTACGGCATTACCCTATTCAGCGAATTTGCGCAGCGTATATTACGCTAGCTACGCGCT  3261
        R  A  P  T  T  A  L  P  Y  S  A  N  L  R  S  V  Y  Y  A  S  Y  A  L

3262  ATCGGTCGAGGCGATACAGAGAGGACTACGCTACGTCTACGCGCGTACGCGCGTACTATGGAGAGTACTC  3331
       S  V  E  A  I  Q  R  G  L  R  Y  V  Y  A  R  T  R  V  L  W  R  V  L

3332  GACGATCGTATATACAGCTACGCCCGTACGGGCGGATCGGCGTAGCGCGCGATCGGATAGCGACTACTAT  3401
      D  D  R  I  Y  S  Y  A  R  T  G  G  S  A  *

3402  ATATACTATCGGAGAGAGTCTCTCCCGAGCTACGCGACCGAGCTCAGCT  3450
```

ORF started with an ATG codon at position 2072 and ended with a TAG codon at position 3376 and consisted of a coding sequence of 1300 bp.

After screening upstream of the 5′ end of ORF3 for SD and promoter sequences, a SD sequence (TCCCGG) was located at position -10 to -5, 5 bp upstream of the ATG initiation codon (Fig. 5-9) but no promoter sequence was found. Several inverted repeat sequences were identified downstream of the stop codon (Fig. 5-9). The nucleotide sequence of the ORF contained 66.6 % G+C revealing a strong bias towards codons containing dG+dC (Table 5-2; Fig. 5-4).

The available nucleotide and amino acid sequences of other β-glucosidase genes were compared to the ORF3 which showed no obvious homologies. Sequence comparison showed a similarity of 29.2%, 45.3% and 32.6% to the β-glucosidase gene of *E. coli* bgl (Schnetz et al, 1987), *Aspergillus wentii* (Bause and Legler, 1980) and *Candida pelliculosa* (Kohchi and Toh-e, 1985)(Fig. 5-10).

The ORF coded for a polypeptide of 382 amino acid residues with a calculated MW of 44.8 kDa. The size was smaller than the peptide estimated from PAGE of 76 kDa.

5-2-3-4 Fourth open reading frame

The fourth open reading frame was found in the reverse transcription frame which was 690 bp long and overlapped with the first ORF (Fig. 5-11). This ORF had putative SD sequence of $^{-8}$CGCTAG^{-3} and promoter sequences of $^{-31}$TATAGA^{-26} and $^{-59}$TAGAGC^{-54}.

Two of the enzymes this ORF may code for either a transferase or a protease which was presumed to be coproduced with the cellulase complex. When compared with the known nucleotide and amino acid sequences of transferase and protease from GenBank and EMBL library, no significant homologies were found. However, a region from 104 to 125 of the deduced peptide sequence of this ORF was highly conserved with the signal

Fig. 5-10 Comparison of ORF3 (*rbg*) with other β-glucosidase genes.

```
        1                                                  50
cons    l......... t.....ar.. .....v.p.. .s..fy.t.. g.....gk.q
 Eco    LFAEMGFTCL RISIAWARIF PQGDEVEPNE AGLAFYDRLF D.....EMAQ
 Clh    LVLSQAALVN TSAPQASN.. .....DDPFN HSPSFYPTPQ GGRINDGKWQ
Orf3    MTVRETYGAV TWTRRRAR.. .....VGGTG PSYGYWYTNC G.....GKLR

        51                                                 100
cons    A....l.el. ...s....s. y..g.Glak. ....ntGg.. ..a.l.....
 Eco    AGIKPL.... .....VTLSH YEMPYGLVK. ....NYGGWA NRAVIG....
 Clh    AAFYRARELV DQMSIAEKVN LTTGVGSASG PCSGNTGS.. .VPRLNISSI
Orf3    ASVRTLWELA RTRSTGAVSG YPAGRGLAK. ....DTGG.. .TALLS....

        101                                                150
cons    .......... a.a.Tvv... ..a..ssa.. g.i...a.av .se.....a.
 Asp                                AZL GF...ZGF.V MSDWAAHHAG
 Eco    .......HFE HYARTVFTRY QHKVALWLTF NEINMSLHAP FTGV.GLAEE
 Clh    CVQDGPLSVR AADLTDVFPC GMAASSSFNK QLIYDRAVAI GSEF.K..GK
Orf3    .......GYP ASAATVVPNE TLAPTSSAHV GYVRAGALPV LSEL.SYVAS

        151                                                200
cons    .sgA.vgl.y gam...aa.g sa.e...... .l...n.... lla.l....t
 Asp    VSGALAGLBM GSMPGBVBYB S.......GT SYWGTNLTIS LWVNGT...V
 Eco    SGEAEV...Y QAIHHQLV.A SARAVKACHS LLPEAKIGNM LLGGLVYPLT
 Clh    GADAILGPVY GPMGVKAAGG RGWEGHGPDP YL......EG VIAYLQ...T
Orf3    TSGAAVALRW AALRSEAASG SASEATPTAA TVRAPNMSTE LLATLP...G

        201                                                250
cons    c..qs...as t...r..... ....rg..p. ...ra.rd.. i.......e.
 Asp    PZW
 Eco    CQPQDMLQAM EENRRWMFFG DVQARGQYPG YMQRFFRDHN I..TIEMTES
 Clh    IGIQSQGVVS TAKHL..... ....IGNEQE HFRFAKKDKH ..........
Orf3    CERASANNAS TGAPR..... ....RPEAPA DKKRADRARA IRPARDAGEP

        251                                                300
cons    ..a.k..p.. f.t.S.y... ...r...... .a..a..... ...s...S.e
 Eco    DAEDLKHTVD FISFSYYMTG CVSHDESINK NAQGNILNMI PNP.HLKSSE
 Clh    ..AGKIDPGM FNTSSSLSSE IDDRAMHEIY LWPFAEAVRG GVSSIMCSYN
Orf3    SLAAKRQPDL YRTNSGYKAL ARVRSLRVQR YASVASERIV AGASRAGSIE

        301                                                350
cons    ....i.a... .......... ......r... ....l.s... aA..s...d.
 Eco    WGWQIDPV.. ....GLRVLL NTLW.DRYQK PLFIV.ENGL GAKDSVEADG
 Clh    KLNGSHACQN SYLLNYLLKE ELGFQGFVMT DWGALYSGID AANAGLDMDM
Orf3    RDRAIRAIVS TFDPSSTIVC SIVRVVRIVR AYLPL.STRQ AASYSRAIAS

        351                                                400
cons    s....y.... .......... .......... ...sr.iY.g .h.p......
 Eco    SIQDDY.... .......... .......... ....RIAYLN DHLVQVN...
 Clh    PCEAQYFGGN LTTAVLNGTL PQDRLDDMAT RILSALIYSG VHNPDGPNYN
Orf3    SDRVELY... .......... .......... .YISRYIYGG RRAPTTA...

        401                                                450
cons    .......... .......... .l.....vrs ..y.s.a... .l..sve...
 Eco    .......... .......... .EAIADGVDI MGYTSWGPID LV..SASHS.
 Clh    AQTFLTEGHE YFKQQEGDIV VLNKHVDVRS DINRAVA... .LRSAVEGVV
Orf3    .......... .......... .LPYSANLRS VYYASYA... .L..SVEAI.

        451                                                500
cons    ...q...ry. .......... .......R.k v.r.....vi ..r..s....
 Eco    ...QMSKRYG FIYVDRDDNG EGSLTRTRKK SFRMVCAEVI KTRGLSLKKI
 Clh    LLKNEHETLP L......... ......GREK VKRI....SI LGQAAGDDSK
Orf3    ...QRGLRYV Y......... ......ARTR VLWR....VL DDRIYSYART

        501
cons    g.S
 Eco    T
 Clh    GTS
Orf3    GGS
```

cons, consensus sequence; Eco, *E coli* K-12 *bgl*B gene (Schnetz et al, 1987); Clh, *Candida pelliculosa* (Kohchi and Toh-e, 1985).

Fig. 5-11 Nucleotide and deduced amino acid sequences of ORF4

```
                                                        -35
4356 GCTACGCTCGCGTATCGCTGTACGCGTAGACTTCGTACGCGAGCTAGAGCTTTATATCCAGAATTCGTAC 4425
          -10                    SD
4426 GATATAGATCAGGCTGTCGTTTGGCCGCTAGCGATGGCTCCCCCGCGTCCCGCGCCGGACGTATCGCCCG 4495
                                  M  A  P  P  R  P  A  P  D  V  S  P  E
4496 AGCCCGATCGGCGCGTTTCTTGTCGGCCGGTGCCTCGGGTCGCCGCGGCGCGCCTGTGGACGCGTTGTTC 4565
       P  D  R  R  V  S  C  R  P  V  P  R  V  A  A  A  R  L  W  T  R  C  S
4566 GCCGAGGCCCGCTCGCACCCCGGCAGCGTCGCGAGTAGCTCGGTGGACATGTTGGGCGCGCGTACCGTGG 4635
       P  R  P  A  R  T  P  A  A  S  R  V  A  R  W  T  C  W  A  R  V  P  W
4636 CTGCGGTGGGGGTCGCTTCGGACGCGCTCCCTGAAGCAGCCTCGGACCGTAACGCGGCCCACCGTAGCGC 4705
       L  R  W  G  S  L  R  T  R  S  L  K  Q  P  R  T  V  T  R  P  T  V  A  R
4706 GACGGCGGCGCCGCTCGTGGACGCGACGTAGCTCAGCTCGCTCAGTACCGGCAGCGCCCCGGCGCGTACG 4775
       R  R  R  R  S  W  T  R  R  S  S  A  R  S  V  P  A  A  P  R  R  V  R
4776 TAGCCGACGTGCGCGGACGACGTCGGCGCGAGTGTCTCGTTGGGCACGACTGTCGCGGCGGACGCGGGGT 4845
       S  R  R  A  R  T  T  S  A  R  V  S  R  W  A  R  L  S  R  R  T  R  G
4846 AGCCGGACAGCAGTGCGGTCCCGCCGGTGTCCTTGGCTAGCCCGCGCCCCGCGGGGTAGCCTGACACCGC 4915
       S  R  T  A  V  R  S  R  R  C  P  W  L  A  R  A  P  R  G  S  L  T  P  P
4916 CCCAGTGGATCGGGTGCGCGCGAGCTCCCATAGCGTCCGGACGCTGGCCCGTAGCTTCCCGCCGCAGTTG 4985
         Q  W  I  G  C  A  R  A  P  I  A  S  G  R  W  P  V  A  S  R  R  S  W
4986 GTGTACCAGTACCCGTAGCTGGGGCCGGTCCCGCCGACCCTGGCCCGCCGTCGCGTCCACGTGACGGCGC 5055
       C  T  S  T  R  S  W  G  R  S  R  R  P  W  P  A  V  A  S  T  *
5056 CGTAGGTCTCGCGCACGGTCATCTTGCCCGGGAGCCCCGCACAGCCCGACACCGGTGCGGGTGCCCGCGCG 5125
5126 GAGTAGCCCGGTGGGAGACCCCCGGCCGAGCGTGACAGTCCGCA 5170
```

sequence of protease III gene of *E. coli* (Claverre-Martin etal, 1987) and the 5' terminal sequence of the *E. coli capR* (ATP-dependent protease) gene (Gayda et al, 1988) (Fig. 5-12). Also, several regions were found with homology to the signal sequences of *Bacillus* sp. protease gene (Henner et al, 1988; Sloma et al, 1988; Wong et al, 1984), protease enhancer gene (Tanaka and Kawata, 1988) and *Streptomyces griseus* protease A gene (Henderson et al, 1987).

5-2-3-5 Fifth open reading frame

The fifth ORF was found near to the 3' end of the 7.3 kb fragments using the second reading frame. The ORF was identified as the structural gene for the *R. flavefaciens* xylanase gene (designated *rxy*: *Ruminococcus xy*lanase) based on the deletion clone assay. The sequence of this ORF together with the 5' upstream sequences and the deduced amino acid sequence was shown in Fig.5-13. This ORF consisted of a coding sequence of 1080 bp, starting with an ATG codon at position 6266 and ending with a TAG codon at position 7291.A SD sequence (GGCGGC) containing 6 C/G's was located at position -8 to -3, 3 bp upstream of the ATG initiation codon (Fig. 5-13). Also a putative promoter sequence was found (Fig.5-13). Several inverted repeat sequences were identified downstream of the stop codon (Fig.5-13). The nucleotide sequence of the ORF contained 74.3% G+C. The resulting DNA sequence revealed a strong dG+dC bias. (Table 5-2; Fig. 5-4).

The ORF was sufficient in size to code for an unmodified protein of 341 amino acid residues with a calculated MW of 40 kDa. The available nucleotide and peptide sequences of other xylanase genes were compared to this ORF. Results of these comparisons showed that there were no obvious homology to the xylanase sequences of *C. thermocellum xynZ* (Grepinet et al, 1988), *Bacillus* sp. (Hamamoto et al, 1987; Fukusaki et al,1984; Yang et al, 1988) and *Cryptococcus albidus* partial nucleotide

Fig. 5-12 Comparison of ORF4 with other signal sequences of proteases

E. coli (K12) *ptr* gene (Claverre-Martin et al, 1987)

```
ORF4 104 RVRSRRARTTSARVSRWARLSR 125
         ..|| : :.    |. ||.||.
       1 MPRSTWFKALLLLVALWAPLSQ 22
```

E. coli capR gene (Gayda et al, 1988)

```
ORF4  10 VSPEPDRRVSCRPVPRVAAARLWTRCSP..RPARTPAASRVARWTCWARV 57
         :.||.. |:....:| .    ::|::||  ..::. .. :  |||   ::
       1 MNPERSERIEIPVLPCAMWWFIRTWSSPYLSGGKNLSVVWKRRWTMIKKL 50

ORF4  58 PWLRWGSL 65
         .| |: .|
      51 CWSRRKKL 58
```

B. subtilis alkaline protease (*apr*E) gene (Henner et al, 1988)

```
ORF4 105 VRSRRART 112
         |||:: :.
       1 VRSKKLWI 8
```

B. subtilis iep gene (Tanaka and Kawata, 1988)

```
ORF4  56 RVPWLRWG 63
         |:::.  |
       1 RLAFY*AG 8
```

B. subtilis extracellular protease (*epr*) gene (Sloma et al, 1988)

```
ORF4  16 RRVSCRPVPRVAAARLWTRCSPRP 39
         :.:||: |..|.    :  .:| :
       2 KNMSCKLVVSVTLFFSFLTIGPLA 25
```

*B. subtilis spr*E gene (Wong et al, 1984)

```
ORF4  50 RWTCWARVPWLRWGSLRTRSLKQPRTVTRPTVARRRRRSWTRRSSARSVP 99
         |:. :.| ..|: : | . .|                  :|.  |..|..
       2 RRG*RVRSKKLWISLLFALTL.................IFTMAFSNMSAQ 34

ORF4 100 A 100
         |
      35 A 35
```

*B. amyloliquefaciens sac*Q gene (Yang et al, 1986)

```
ORF4 107 SRRARTTSARVSRWARLSRRTRGSRTAVRSRRCPWLARAPRGSLTPPQWI 156
         .::  .... : |:..  | | :|             |. . |:.. :.:
       2 EKKLEEVKQLLFRLENDIRETTDS............LRNINKSIDQLDKF 39

ORF4 157 GCAR 160
         ::|.
      40 SYAM 43
```

S. griseus protease A (*spr*A) gene (Henderson et al, 1987)

```
ORF4 120 WARLSRRTRGSRTAVRSRRCPWLARAPRGSLTPPQWIGCARAPIASGRWP 169
         : |:|. ...|| |                   .::::.|.: :|.: ::
       3 FKRFSPLSSTSRYA...................RLLAVASGLVAAAALA 32

ORF4 170 VAS 172
         .:|
      33 TPS 35
```

|, :, ., different digree of similarity of amino acid.

Fig. 5-13 Nucleotide and deduced amino acid sequences of ORF5

```
                  -35                  -10                  SD
6209  CGCATACAGCAATTAGTCAGATTTCACTTAAGTGCGAGTACGTTCTATTGGCGGCAGATGCGCTCAGTCC  6278
                                                            M  R  S  V  R

6279  GTCCTCGGGGTCCACGACGGCAGCACTCCGGCCAACCAGGGGGCGAAAGGCGCCAACTCGGCGACCGGCC  6348
        P  R  G  P  R  R  Q  H  S  G  Q  P  G  G  E  R  R  Q  L  G  D  R  P

6349  TAGACGCCAGCGCGCTACCGCTCAAGACGCGGGGCGCGCCGTCCGGGCACGTACCGACGATCGTGCTCGG  6418
        R  R  Q  R  A  T  A  Q  D  A  G  R  A  V  R  A  R  T  D  D  R  A  R

6419  GGTCACCAACCAGGGCGTCCTTCCCGGCGTCGCGGCTACGGTCGGCTGCGGCGACACGCGCGGCCCGGCC  6488
      G  H  Q  P  G  R  P  S  R  R  R  G  Y  G  R  L  R  R  H  A  R  P  G  R

6489  GACGCGCGGACGCGCGCGATCGGCCGCGGGACGGCCTCCGCGATCGGCCACGCGCAGCGCACCCAGGCGC  6558
        R  A  D  A  R  D  R  P  R  D  G  L  R  D  R  P  R  A  A  H  P  G  A

6559  AGGCGAGCCTCCCCGACACGGGCGCCAGCGCCGGGATCGCCCGGTCGCTGAATGCGGCGTCCAACACGAC  6628
        G  E  P  P  R  H  G  R  Q  R  R  D  R  P  V  A  E  C  G  V  Q  H  D

6629  GGCTCTCGCGACGGAGCGCAACGGCGCGGGCAACGAGGTCCACGTCTCGAAGCTTCCCACGTCGCAGGCC  6698
      G  S  R  D  G  A  Q  R  R  G  Q  R  G  P  R  L  E  A  S  H  V  A  G  P

6699  CAGCACGAGCGGACGCTGGCGCTGGCGACAGCTGCGGCCGCTGCGGCTGGCGCGGCCGGCCACGCGACCT  6768
        A  R  A  D  A  G  A  G  D  S  C  G  R  C  G  W  R  G  R  P  R  D  L

6769  CCGGCAAGACCGGGCCGCTGACGCTCGGCGCCGCGGCCTCGGCGACGTGGCACAGGTGGGGCGAGCCGCT  6838
        R  Q  D  R  A  A  D  A  R  R  R  G  L  G  D  V  A  Q  V  G  R  A  A

6839  CGCGGGGCGCGTCCGCGTCGGGTCGGTCGTGAACGACCCGAGCGACCTCTGCACCAACGGCACTCAGGGG  6908
      R  G  A  R  P  R  R  V  G  R  E  R  P  E  R  P  L  H  Q  R  H  S  G  D

6909  ATCTCGTGGGCCTCGACGCACACGGGAGGGCTGGAGGAGTGGCTCGCGGCCGCCCTCGGCCAGGACACGG  6978
        L  V  G  L  D  A  H  G  R  A  G  G  V  A  R  G  R  P  R  P  G  H  G

6979  GCTCGGCCGTGGATGGCTCGGTCCCCGCGCGCGAGTGCGAGCACGCGGTCGAGGCGCGCGTGTGCTCGGC  7048
      L  G  R  G  W  L  G  P  R  A  R  V  R  A  R  G  R  G  A  R  V  L  G

7049  GTCCCCCGGCCTGTGGCTCTCCAGCGGCGCGCTCTCGCCCGAGAGGCCCCACGAGTCCCGCGTCGGCACC  7118
      V  P  R  P  V  A  L  Q  R  R  A  L  A  R  E  A  P  R  V  P  R  R  H  R

7119  GGGTGCTCGGCCTGCGCGGCTCGACCCCATCGAGCTCGCCTCCGACTTCACCTAGTCCCAGCCCTACGTC  7188
        V  L  G  L  R  G  S  T  P  S  S  S  P  P  T  S  P  S  P  S  P  T  S

7189  GCCCACCCCTCCTTCGCCTTCGGTCAACGACCGGCCTCGGCCTCGGGCCCCCTCGATCGGCACCGGCGGT  7258
      P  T  P  P  S  P  S  V  N  D  R  P  R  P  R  A  P  S  I  G  T  G  G

7259  GCCGTGCCGCTGCGTCGATACGCTCGGTCGTAGCATACGCTACGAT  7304
      A  V  P  L  R  R  Y  A  R  S  *
```

sequence. The best similarity of xylanase gene was xylA gene of alkaline *Bacillus* sp. strain C-125 (Yang et al, 1988) with 29.0% homology (Fig. 5-14).

A shorter Proline, Threonine and Serine rich region similar to the PTS regions in *ren*A and *rex*A was found, from the position 7138 to position 7209 of about 24 amino acids. The region was 82 bp upstream from the C-terminal. When comparing this region with the PTS sequence of *ren*A and *rex*A, a highly consensus sequence of these regions was found (Fig. 5-8). The ORF5 could be divided into three domains: N-terminal domain, PTS domain and C-terminal domain.

5-3 Discussion

The sequence results of the 7304 bp fragment had five largecomplete ORFs. The 5′ and 3′ noncoding regions were enriched in adenine and thymine residues, as opposed to the ORFs, which had high G+C content. The G+C content of the nucleotide in *ren*A, *rex* and *rbg* were 71.7 %, 67.3 % and 66.6%, respectively, which is higher than that of *C. thermocellum* (celA, celB, celC and celD, 39.4%, 39.8%, 35.5% and 40.3%, respectively; Beguin et al, 1985; Grepinet and Beguin, 1986; Schwarz et al, 1988; Joliff et al, 1986; Hall et al, 1988) and *Bacillus* sp. (1139, 37.9%; pNK1, 39.3%; pNK2, 39.7%; Fukumori et al, 1986; MacKay et al, 1986; Nakamura et al, 1987; Fukumori et al, 1989), but more closely related to that of *Ce. fimi* cenA and cex (74.3 % and 71.8 %; Wong et al, 1986; O'Neill et al, 1986) and *T. reesei* EGI and EGIII (61.4% and 51.2%; Chen et al, 1987; Shoemaker et al, 1983; Teeri et al, 1987).

Fig. 5-14 Comparison of ORF5 (*rxy*) with other xylanase genes

```
        2                                                          51
cons    .S........ .hSgqp.e.w .nwgd..... .na.a.sag. ..r..t....
Orf5    RSVRPRGPRR QHSGQPGGER RQLGDRPRR. QRATAQDAGR AVRARTDD..
 Bsp    NSLVAENAMK PESLQPREGE WNWEG.ADK. IVEFARKHNM ELRFHTLV..
 Bpu              NHSGYDYELW KDYGNTSMTL NNGGAFSAGW N.........
 Bci                SAASTDYW QNWTDGGGI. VNAVNGSGGN YSVNWSNTGN

        52                                                        101
cons    .....g..pg rpfr..dynr .v....p..r .......r.. l...y.a...
Orf5    ..RARGHQPG RPSRRRGYGR LRRHARPGRR ADARDRPRDG LRDRPRAAHP
 Bsp    ..W.HSQVPE WFFIDEDGNR MVDETDPDKR EANKQLLLER MENHIKTVVE
 Bpu    ...NIGNALF RKGKKFDSTR THHQL..... ........GN ISINYNASFN
 Bci    FVVGKGWTTG SPFRTINYNA GVWAPNGNGY LTLYGWTRSP LIEYYVVDSW

        102                                                       151
cons    g.g.p..... g.v.sp.aey gi....sw.. .......... ..tar..G.r
Orf5    GAGEPPRHGR QRRDRPVAEC GVQHDGSR.. .......... .DGAQRRGQR
 Bsp    RYKDDVT.SW DVVNEVIDDG GGLRESEWYQ ITGTDYIKVA FETARKYGGE
 Bpu    PSGNSYLCVY GWTQSPLAEY YI..VDSW.. .......... .GTYRPTGA.
 Bci    GTYRPTGTYK GTVKSDGGTY DIYTTTRY.. .......... .NAPSIDGDR

        152                                                       201
cons    ...l..s... .......... .......vaq g.r.d...t. s..qigw.g.
Orf5    GPRLEASH.. .......... .......VAG PARADAGAGD SCGRCGWRGR
 Bsp    EAKLYINDYN TEVPSKRDDL YNLVKDLLEQ GVPIDGVGHQ SHIQIGWPS.
 Bpu    ...YKGSF.. .......... .......YAD GGTYDIYETT RVNQPSIIGI
 Bci    TTFTQYWS.. .......... .......VRQ SKR....PTG SNATITFTNH

        202                                                       251
cons    ....kq.ra. ....r.lg.v aqv..... .g....ga.. r.v..e....
Orf5    PRDLRQDRAA DARRRGLGDV AQV..... .GRAARGARP RRVGRERPER
 Bsp    ...IEDTRAS FEKFTSLGLD NQVTELDMSL YGWPPTGAYT SYDDIPAELL
 Bpu    A.TFKQYWSV RQTKRTSGTV SV........ ......SAHF RK..WESLGM
 Bci    VNAWKSHGMN LGSNWAYQVM ATE..... .GYQSSGSSN VTV...W*

        252                                                       301
cons    p........l ..l.y...a. .........g ......wl.g raR....G..
Orf5    PLHQRHSGDL VGLDAHGRAG GVARGRPRPG HGLGRGWLGP RARVRARGRG
 Bsp    QAQADRYDQL FEL.YEELAA DISSVTFW.G IADNHTWLDG RAREYNNG..
 Bpu    PMGKMYETAF TVEGYQSSGS .........A NVMTNQLFIG N*

        302                                                       351
cons    ....G..... .......... ..P....... .......... A.......R.
Orf5    ARVLGVP... .......... .RPVALQRRA LARE...... APRVPRRHRV
 Bsp    ...VGID... .......... .APFVFDHNY RVKP...... AYW.....RI

        352
cons    ..*
Orf5    LG
 Bsp    ID*
```

cons, consensus sequence; Bsp, alkalophilic *Bacillus* sp. *xyl*A (Hamamoto et al, 1987); Bpu, *B. pumilus xyn*A (Fukusaki et al, 1984); Bpu, *B. circulans xln*A (Yang et al, 1988).

The ORF1, which was identified as renA, started at nucleotide 712 of the DNA sequence. There is some evidence which supports the contention that this ATG is in fact used as an initiation codon. There is

(i) a Gram positive bacterial type of ribosome-binding site, CGGGGA, which starts at -10, the optimum distance for this ATG,

(ii) putative -10 and -35 promoter sequence was found upstream of this ATG

(iii) Application of programs (TESTCODE and CODENPREFERENCE) for prediction of a peptide coding region.

Similar structure were also found with ORF2, ORF3 and ORF4.

The codon used in ORF1 (*ren*) was very similar to that of *Ce. fimi cen*A which also had a high G+C content. The cenA gene showed the best similarity of the renA (32.1% with 485 a.a. residues. Three domains (binding domain, PT box and catalytic domain) were found in the *Ce. fimi* gene (Warren et al, 1986). A similar high pro-thr-ser rich region located near to the N-terminal of the peptide was found in renA and divided the protein into two domains. This suggested that the *ren*A gene of *R. flavefaciens* could be grouped into family A of cellulase genes (Ong et al, 1989).

The putative molecular weight of renA gene product was 75.8 Kd with 680 a.a. residues which was larger than the peptide that was isolated from culture of clone λCM903. It suggested that post-translational modification with protease was involved as in the case with *T. reesei*, *Ce. fimi* and *C. thermocellum* (Stahlberg et al, 1988; Gilkes et al, 1988; Lamed and Bayer, 1988).

The unusual structure of the pro-thr-ser rich region in *ren*A, *rex* and *rxy* was compared. The PTS of *ren*A and *rex* had 67.5% similarity. It showed the highly conserved sequence of PTS in *ren*A, and *rex* genes which is comparable to the PT boxes

of *Ce. fimi cen*A and *cex* genes (Warren et al, 1986). The PT region was a possible primary sites for protease or papain which were reported in *Ce. fimi* and *T. reesei* (Ong et al, 1989). However, the protease cleavage site has not been confirmed in the PTS regions of *ren*A, *rex* and *rxy*. Also in *Ce. fimi*, the PT box divided the cellulase gene into the binding and catalytic domain. In the case of *ren*A, the PTS was located near the N-terminal end of the gene. This might indicated that reshuffling of the gene would be required in *ren*A to form the general cellulase gene structure during the transcriptional step. The function of PTS region needs further study.

The ORF4 was compared with other protease genes. Although there is no obvious homology with the protease sequences from GenBank and EMBL library, the homology of this gene with the signal sequence of *E. coli* protease III and 5' end matured ATP-dependent protease gene strongly suggested that this ORF coded for a protease gene. Also homologies with other *Bacillus* sp. and *Streptomyces* sp. protease sequences, supports this suggestion. Further study to assay the protease activity and to purify this enzyme will help to identify the gene product.

The *rbg* gene was found in the second reading frame overlapping with *ren*A gene and this suggested that the two genes share some sequence perhaps important for gene regulation. The *ren*A, *rex* and *rbg* genes were found in close proximity along with ORF4, which has a protease gene characteristic. These suggested that the cellulase genes could be under some type of regulatory control as a whole, even though they were all found having individual putative promoter sequences.

The noncellulosic degradation enzyme against xylan were detected which had been described in the previous chapter. The *rxy* gene was sequenced and was located close to the cellulase genes. The structure of *rxy* gene showed some similarity to the cellulase genes which contained three domains. In particular,

a short sequence of PTS rich region was found in this gene. The study of xylanase gene from *R. flavefaciens* strain 17 has been reported (Flint et al, 1989). Although the sequence of xylanase gene from strain 17 is still unavailable, the finding of PTS rich region in *rxy* gene has not been reported in other known xylanase gene sequences.

5-4 Short summary

A fragment of 7.3 kb *Hin*dIII-*Hin*dIII fragment within the λCM903 insert was sequenced. Four ORFs were found and analyzed using computer software (UWGCG).

The first ORF corresponded to subclones showing endoglucanase activity and was designated as the structured gene for endoglucanase gene (*ren*A). It was 2157 bp long with a putative 680 residues amino acid. SD and promoter sequences were found. *cen*A gene of *Ce. fimi* gave the best fit from 20 known endoglucanase sequences. An unusual structure of Pro-Thr-Ser (proline, threonine or serine) rich region, in which 38 out of 42 a.a. residues, was found near the N-terminus of the *ren*A gene.

The second ORF which was 1821 bp long and was designated as coding for exoglucanase gene (*rex*) since subclones of this gene corresponded to MUCase activity. The putative protein sequence had 572 amino acid residues, and SD and promoter sequences have been found. A pro-thr-ser rich region, which has highly conserved region to the PTS sequence of *ren*A, was found near the C-terminus of the putative peptide. Also, the most similar cellulase gene was the *cex* gene of *Ce. fimi*. It suggested that the gene structure of cellulase in *R. flavefaciens* strain 186 belonged to the cellulase gene family A, according to the classification of Ong et al, 1989.

The third ORF was found overlapping with ORF1 using the second transcriptional reading frame. The ORF had a putative

SD sequence but lacked of any clear promoter sequence. This ORF coded for a β-glucosidase gene which was 1300 bp long (a putative protein sequence of 443 amino acid residues).

The fourth ORF overlapped with ORF1 but used the reverse transcriptional reading frame. This ORF was 690 bp long with SD and promoter sequences upstream of the ATG codon. This ORF ha the characteristic of a protease gene due to the homologies of this sequence to signal sequences and 5′ terminal sequence of certain protease genes.

The fifth ORF near the 3′ end of the 7.3 kb fragment was identified as a xylanase gene which used the second transcriptional reading frame. This ORF was 1080 bp long with SD and promoter sequences. A short PTS rich sequence was found and showed highly consensus with PTS rich region of ORF1 and ORF2.

Chapter VI

FINAL DISCUSSION AND CONCLUSIONS

Table of contents

Techniques of molecular biology are important tools in biological research. There has been a rapid development of these techniques over the last 20 years and this development has been responsible for a vast increase in the understanding of the structure of genes and proteins in the biological world. The aim of the work presented in this thesis was to use these techniques to isolate and characterize genes coding for the cellulases of *R. flavefaciens*.

6-1 Use of the *E. coli* lambda system in the study of rumen microorganisms

The discovery and construction of host/vector systems has caused a revolution in the understanding of genes at the molecular level. Construction of gene libraries, screening of recombinants for specific protein products, restriction endonuclease mapping of the genes, and subcloning fragments of a particular gene into vectors for expression or sequencing has been the usual procedure in the investigation of gene structure and function. The ease of isolation, wide range of selection markers and varied choice of vectors have made plasmids the most commonly used vectors. However, in some cases, instability of plasmid and low transforming frequency of plasmid systems can be problematic in practice. In this study, instability of ruminococcal DNA inserts in *E. coli* plasmid vectors has made cloning studies using these types of vectors impossible. Several attempts at cloning ruminococcal DNA by Gram positive and Gram negative plasmid vectors were unsuccessful. It is thus suggested that plasmid vectors are not suitable for cloning studies with ruminococcal DNA.

The increasing use of bacteriophage lambda as a cloning vector has led to recent development of techniques for its manipulation. Most of the classical problems which limited the use of lambda vector such as difficulties in isolation of recombinants, high cost of packaging, laborious purification

of DNA and lack of sequencing protocol, have been solved. In this study, the genomic library was constructed in the lambda vector NM1149, the insert fragments were subcloned into λGem11 and the subsequent deletion clones were sequenced using the dideoxy chain termination method. Ease of screening and stability of the clones have made the lambda cloning system the best system for this type of genetic study.

6-2 The cellulase of ruminococci

The cellulase of *C. thermocellum* is thought to occur as a complex of cellulase activities and this complex has been called a cellulosome (Lamed and Bayer, 1988). A similar structure may be presented with *R. flavefaciens* 186 growing on Avicel as shown by TEM (re: Fig. 1-1). The enzyme purified from *R. flavefaciens* 186 consistently showed about 15 polypeptides of which only a small number showed CMCase activity (four bands, the smaller of which are probably products of proteolysis), MUCase activity (1-2 bands) or MUGase activity (1 band) (unpublished data (Asmundson) and data from this study (not shown)). The cellulases of *R. flavefaciens* have been shown to exist in three different forms:mx high molecular weight complex with MW > 3 x 10^6 Da and smaller forms with MW about 8 x 10^5 Da and about 89 x 10^3 Da (Pettipher and Latham, 1979). The cellulase complex of *R. flavefaciens* is thus composed of a number of polypeptides, only a number of which are glycosidases.

The predicted molecular weights of the glycosidases sequenced in this work and the molecular weight of the enzymes determined from other studies are compared in Table 6-1.

The lack of agreement between the estimated molecular weight of isolated proteins and proteins predicted from DNA sequences may be due to non-correspondence of the genes cloned with the isolated form of glycosidase (since there are several

Table 6-1 Summary of the characterization of *R. flavefaciens* gene products and deduced peptides and comparison with other cellulase proteins.

	native protein	gene product	gene deduced	$T_{opt(°C)}$	pH_{opt}
Endoglucanase					
R. albus					
strain F40[1]	50 kd			44	6.8
strain SY3[2]	-	56 kd		-	-
R. flavefaciens					
strain 67[3]	89 kd			45	6.4
strain 186	62 kd[7]	45 kd	75.8 kd	30	5.5
Exoglucanase					
R. albus[4]	110 kd			37	6.8
R. flavefaciens					
strain FD1[5]	118 kd			39-45	5.0
strain 186		90 kd	67 kd	36	5.2
β-glucosidase					
R. albus					
strain F40[6]	82 kd			30-35	6.5
strain 186		76 kd	44.8 kd	33	6.3

1, Ohmiya et al, 1987; 2, Romaniec et al, 1989; 3, Pettipher and Latham, 1979; 4, Ohmiya et al, 1982; 5, Gardner et al, 1987; 6, Ohmiya et al, 1985; 7, Asmundson, unpublish data.

classes of non-homologous genes present) or due to post-translational processing, either intracellularly, during transport or due to extracellular proteolysis. The latter cannot explain the situation where the observed protein is larger than the predicted form, but may explain that where it is smaller, as with the endoglucanase. Also, the methods chosen to determine the molecular weight of active protein may not be the same as the result observed by SDS-PAGE. The molecular weight of gene product in this study was determined by native gel or SDS-PAGE at a single polymer concentration without boiling the samples, which can only give an approximate estimate of molecular weight.

6-3 The cellulase gene system in ruminococci

Although this is the first report on sequencing of more than one cellulase gene from a single clone of *R. flavefaciens* (Summary in Table 6-2), several cloning studies of endoglucanase genes from *R. albus* and *R. flavefaciens* have been published recently as described in the Introduction. Several general principles resulting from these studies are described in the following sections.

Table 6-2 Summary of ORFs of 7.3 kb fragment

	ORF1	ORF2	ORF3	ORF4	ORF5
Code	Endoglucanase *renA*	Exoglucanase *rex*	β-glucosidase *rbg*	protease[1]	xylanase *rxy*
NT Length (bp)	2157	1821	1300	690	1080
SD	CGGGGA	ATGGGG	TCCCGG	CGCTAG	TGGCGG
Promoter (-10)	TTTACA	TATATA	NF[2]	TATAGA	TTAAGT
(-35)	TACGAA	TCGAAG	NF[2]	TAGAGC	TACAGCA
G+C (%)	71.7	67.3	66.6	70.9	74.7
PTS[3]	YES	YES	NO	NO	YES
aa Length[4]	680	572	382	196	341

1, Putative coding
2, Not found
3, Proline-threonine-serine rich region
4, deduced amino acid length

(i) Multiple genes coding for the cellulase

This work showed that from about 2500 recombinants, 26 CMCase$^+$ clones were found, indicating a high frequency of CMCase genes in *R. flavefaciens* 186, assuming that the library presents a homogeneous distribution of the entire genome. Homology studies indicated that there were at least 8 different CMCase gene classes within the 26 positive clones. This is in agreement with the reports from *R. albus* strain 8 (Howard and White, 1988), F-40 (Ohmiya et al, 1988; Kawai et al, 1987), AR67 and AR68 (Ware et al, 1989) and *R. flavefaciens* FD-1 (Barros and Thomson, 1987). A high percentage of cellulase positive clones was also reported in *R. albus* 8 using λDASH

vector with 0.33% (CMC^+), 0.37% (MUC^+) and 2.78% ($OBR\text{-}HEC^+$) (Howard and White 1988). So far more than 20 different non-homologous cellulase genes have been found in *C. thermocellum* (Hazlewood et al, 1988;). Also, non-homologous cellulase genes were found in *R. albus* (Kawai et al, 1987; Honda et al, 1988; Ohmiya et al, 1988; Romaniec et al, 1989; Flint et al, 1989; Ware et al, 1989), *R. flavefaciens* (Howard and White 1988), *Ce. fimi* (Beguin and Eisen, 1978; Whittle et al, 1982; Gilkes et al, 1984; Owolabi et al; 1988; Akhtar et al, 1988; Moser et al, 1989) and *Bacillus* spp. (Sashihara et al, 1984; Hinchliffe, 1984; Park and Pack, 1986; Fukumori et al, 1986; Koide et al, 1986; Robson and Chambliss, 1987; Sharma et al, 1987; Kim et al, 1987; Fukumori et al, 1989) showed different genes coding for the similar function of cellulase activity. Homologous studies of *R. albus* clones have identified at least 10 different cellulase genes (Howard and White, 1988). Two CMC^+ clones and three CMC^+ clones which were isolated by different laboratories (Ohmiya et al, 1988; Kawai et al, 1987) from *R. albus* F-40 showed no homology. Four non-homologous recombinant CMC^+ phages were isolated from AR67 and AR68 (Ware et al, 1989), also in FD-1, two colonies with CMCase activity were isolated from 4400 transformants (Barros and Thomson, 1987).

Enzyme purification data and genetic data have both strongly indicated the presence of multigenes coding for the cellulase enzymes in ruminococci. This is also observed in other cellulolytic bacteria such as *Cellulomonas* sp., *Clostridium* sp., *Bacillus* sp. and *Fibrobacter* sp.(re: Introduction).

(ii) <u>The genes coding for cellulases form a gene cluster</u>

λCM900 series and λCM400 series from *R. flavefaciens* in this report and at least four clones of *R. albus* strain 8 were found to express more than one type of cellulase activity (Howard and White, 1988). Four clones of recombinant phage

isolated from strain 8 showed CMCase and pNPCase. Also another four clones showed CMCase, pNPCase and pNPGase. The 7.3 kb fragment of λCM903 which has been sequenced in this work, coded for all three types of cellulase genes within a close proximity (within 5 kb). In particular, the endoglucanase and β-glucosidase genes shared the same DNA at different reading frames. Although the regulation of these genes is still unclear, clustering of the genes suggests a coordinated regulation mechanism.

Similar results were found in *Clostridium thermocellum* and *Caldocellum saccharolyticum* (Soul, personal contact), in which a single gene fragment coded for both CMCase and MUCase. In *C. thermocellum*, several clones had both CMCase and MUCase (Millet et al, 1985) and one fragment of DNA, on which two ORFs were found, coded for endoglucanase and xylanase (Hall et al, 1988). This suggests that cellulases are clustered and hence, along with the cellulosome concept (a complex of activities), that there may occur some system for the coordinated expression of these enzymes to degrade cellulose.

(iii) The recombinant gene products are located in the periplasmic space and intracellularly

Cultures of ruminococci normally secreted their cellulases, whereas in *E. coli* the cellulases were detected intracellularly or in the periplasmic space (This study; Ohmiya et al, 1988; Kawai et al, 1987). Differences in secretion mechanisms or post-translational modification may explain these differences.

The relationship of glycosylation and post-translational modification by protease during secretion and the stability of cellulase had been studied in *Clostridium* (Lamed and Bayer, 1988) and *Cellulomonus* (Gilkes et al, 1989). In *Ce. fimi*, the binding domain was excised from CenA or Cex by proteolytic

cleavage immediately adjacent to the carboxyl terminus of the PT box. The amino-terminal fragment (p20) of CenA and carboxyl-terminal fragment (p8) of Cex were retained in cellulose columns due to binding to cellulose while the corresponding fragments, p30 from CenA and p35 from Cex, which were unable to bind to cellulose, contained catalytic domains (Gilkes et al, 1988). The p30 and p35 fragments were generated by papain or α-chymotrypsin and was used in the test for stability. The results showed no change of the fragments indicating that the catalytic domain adopts a tightly folded conformation affording protection from proteolytic attack (Gilkes et al 1989). Similar results of EG2 on *Bacteroides succinogenes*, in which a 118 kDa enzyme is composed of a 51 kDa catalytic domain and a highly antigenic 43 kDa substrate-binding domain after processing by protease (McGavin and Forberg, 1989) and an endoglucanase from *Bacillus subtilis*, in which a 52.2 kDa proenzyme was cleaved progressively to a product of about 32 kDa (Lo et al, 1988).

Also, glycosylated cellulases from *Ce. fimi* were compared with their non-glycosylated counterparts synthesized in *E. coli*. The glycosylated enzyme was protected from attack by a *Ce. fimi* protease when bound to cellulose, whereas the non-glycosylated enzymes were active, truncated products with greatly reduced affinity for cellulose (Langsford et al, 1987). Also, using various truncated forms of CenA, it has been shown that the C-terminal half of the binding domain, which is adjacent to glycosylated site, is more important in cellulose binding than the N-terminal half (Ong et al, 1989). This is in agreement with the observation in *Trichoderma reesei* that O-linked but not N-linked glycosylation is necessary for the secretion of endoglucanase I and II (Kubicek et al, 1987). Although there are no significant homologies between the cellulose binding domains of fungal and bacterial enzymes, there are similarities in the distribution of hydrophobic and hydroxyamino acids and low charged residues in both fungi and bacterial cellulose binding domains.

Generally, in *E. coli*, there is no glycosylation of the expressed gene products which might explain their instability in this host (Kane and Hartley, 1988). In addition, protease that cleaved cellulases into smaller peptides which would facilitate the action of transferase but this expression could be lethal to the *E. coli* host. Further study of the action of proteolytic processing and glycosylation of cellulases in *R. flavefaciens* is needed.

Several strategies have been used to overcome the problem of secretion of foreign protein in *E. coli*. Overproduction has been reported with recombinant exoglucanase of *Ce. fimi* in which the exoglucanase gene was fused to a synthetic ribosome-binding site and placed under the control of the leftward promoter of lambda phage. The overproduced exoglucanase can be isolated easily in an enriched form as insoluble aggregates and exoglucanase activity can be recovered by solubilization of the aggregates in 6 M urea or 5 M guanidine hydrochloride (O'Neill et al, 1986). Another experiment using a high copy number plasmid containing a *lac* promoter to replace the promoter sequence of cenA caused an increase of some 800-fold in expression. Accumulation of a protein to a high level in the periplasm may destabilize the outer membrane of *E. coli*, resulting in the leakage of periplasmic protein to the medium (Gatz and Hillen, 1986; Abrahmsen et al, 1986). In addition, high level expression of *Clostridium thermocellum* cellulase genes can be achieved in *E. coli* by subcloning in a temperature-regulated vector which contains the leftward promoter of lambda and using thermal inactivation of the heat-sensitive lambda *cI*857 repressor. Overexpression of the *cel*A gene would resulted in a decrease in cell viability concomitant with the accumulation of endoglucanase A in the membrane fraction.

(iv) Cellulase gene consists of different domains

A general structure of cellulase gene has been reported

which contains at least three domains (cellulose binding domain, catalytic domain and a connecting domain which is rich in proline, threonine and serine; Ong et al, 1989). In this report, the ORFs coding for endo- and exo-glucanase all had PTS rich regions separate from the sequences of the other two domains. However, this concept did not completely fit with the *ren*A gene where the PTS rich region was close to the start codon ATG (21 bp downstream of ATG). Also West et al (1989) reported that the homologous catalytic domains of *Ce. fimi* found in fungi and *Bacillus* sp. apparently evolved by reshuffling of catalytic domains and several substrate-binding domains. When comparing the binding domain of *cen*A and *cex* with ORF1 and ORF2 in this study, homologous regions of the N-terminal end of ORF1 and ORF2 were found. Also the catalytic domains were found by comparing the catalytic domains of *cen*A and *cex,* which indicated the region between PTS rich region and the binding region in ORF1 and the region near to the C-terminal end in ORF2 were the catalytic domains (summarized in Fig. 6-1). Wong et al (1988) suggested that the function of PT box of *cel*A and *cex* was to separate the binding and catalytic domains. The case in *ren*A, a different structure of cellulase gene is shown in which the PTS rich region is outside of the catalytic and binding domains. Further study of the function of PTS rich region is needed.

(v) High G+C content of cellulase genes

The G+C content of the nucleotide in *ren*A, *rex* and *rbg* were 71.7 %, 67.3 % and 66.6%, respectively, which is higher than that of *C. thermocellum* (celA, celB, celC and celD, 39.4%, 39.8%, 35.5% and 40.3%, respectively; Beguin et al, 1985; Grepinet and Beguin, 1986; Schwarz et al, 1988; Joliff et al, 1986; Hall et al, 1988) and *Bacillus* sp. (strain 1139, 37.9%; pNK1, 39.3%; pNK2, 39.7%; Fukumori et al, 1986; MacKay et al, 1986; Nakamura et al, 1987; Fukumori et al, 1989), but close to that of *Ce. fimi* cenA and cex (74.3 % and 71.8 %; Wong et

al, 1986; O'Neill et al, 1986) and *T. reesei* EGI and EGIII (61.4% and 51.2%; Chen et al, 1987; Shoemaker et al, 1983; Teeri et al, 1987).

6-4 Applications of cellulases and cellulase genes

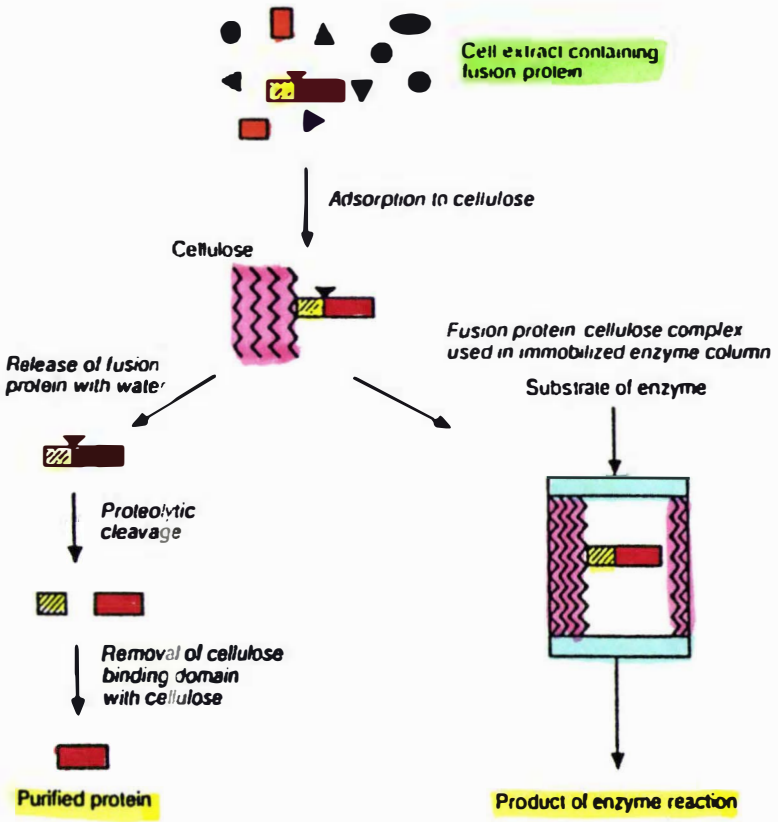
Several potential uses of cellulases have been reviewed (Coughlan 1985; Forsberg et al, 1986; Kennedy et al, 1987; Russell and Wilson, 1987; Hazlewood and Teather, 1988) particularly in the area of hydrolysis of polysaccharide during brewing and in the extraction of fruit juice and other products of plant origins. Alkalophilic cellulase has been used in detergent (Kawai et al, 1988). The specific activity of cellulase has been improved through molecular biology studies of cellulase producing microorganisms such as *Trichoderma* (Coughlan, 1985).

A novel usage of cellulase gene was based on the specialized function of the catalytic domain and the cellulose binding domain of cellulase gene. This property has been used to construct hybrids in which a cellulose binding domain is fused to the N- or C-terminus of a heterologous protein (Ong et al, 1989; Beldman et al, 1987; Greenwood et al, 1989). This concept has been further developed for broad applications in protein purification and enzyme immobilization (Fig. 6-1, Ong et al, 1989). This system used the cheap source of cellulose as ligand and under appropriate conditions the binding of fusion protein was strong enough to function as one form of enzyme immobilization. In addition, the enzyme or the product of enzyme reaction could be eluted from cellulose without specific reagents.

6-5 Study of cellulase genes of rumen microorganisms

Investigation of the molecular genetics of rumen microorganisms offers great potential for improvements in ruminant

Figure 6-1 A diagram in application of binding domain of cellulase gene (from: Ong et al, 1989)



Applications of cellulose-binding domains in protein purification and enzyme immobilization. The fusion protein comprising a CBD (▨) is linked to the desired protein by a specific protease site (▼).

nutrition. Increased rates of cellulose digestion and changes in amino acid composition of the microflora in rumen may be aspects of ruminant nutrition particularly sensitive to genetically altered bacteria. The study of the genetics of anaerobic rumen microorganisms would greatly facilitate the application of existing in vitro genetic techniques to these microorganisms and the understanding of the control mechanisms in these bacteria within this type of ecosystem.

6-6 Conclusion

This study, using a rumen microorganism to study the cellulase genes involved in cellulose degradation, has provided the first detailed molecular and genetical analysis of a cellulase gene cluster which contained all three types of cellulase enzymes. The cellulase gene structures of *R. flavefaciens* fitted well with the general concept that cellulase genes formed clusters. Further study of the regulatory system of these gene clusters is needed. The function of PTS rich region, the structure of binding domains and the function of the protease gene need to be further characterized also.

BIBLIOGRAPHY

Abuja PM, Pilz I, Claeyssens M and Tomme P, 1988. Domain structure of cellobiohydrolase II as studied by small angle-x-ray scattering: close resemblance to cellobiohydrolase I. Biochem. Biophys. Res. Commun. 156: 180-185.

Alberts B, Bray D, Lewis V, Raff M, Roberts K and Watson JD, 1983. Molecular Biology of the cell. Garland Publishers, New York

Alexander M, 1961. Introduction to soil microbiology. John Wiley. PP. 163-182.

Akhtar MW, Duffy M, Dowds BCA, Sheehan MC and McConnell DJ, 1988. Multigene families of *Cellulomonas flavigena* encoding endo-β-1,4-glucanases (CM-cellulases). Gene 74:549-553.

Ames GF, Prody C and Kustu S, 1984. Simple, rapid and quantitative release of periplasmic proteins by chloroform. J. Bacteriol. 160: 1181-1183.

Asmundson RV and Kelley WJ, 1987. Isolation and characterization of plasmid DNA from *Ruminococcus*. Curr. Microbiol. 16: 97-100.

Aymeric JL, Guiseppi A, Pascal MC and Chippaux M, 1988. Mapping and regulation of the *cel* genes in *Erwinia chrysanthemi*. Mol. Gen. Genet. 211: 95-101.

Bailey M, Enari TM and Linko M (eds), 1975. Symposium on enzymatic hydrolysis of cellulose. SITRA, Helsinki.

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

Bause E and Legler G, 1980. Isolation and structural of a tryptic glycopeptide from the active site of β-glucosidase A_3 from *Aspergillus wentii*. Biochim. Biophys. Acta. 626: 459-465.

Béguin P, Eisen H and Roupas A, 1977. Free and cellulose-bound cellulase in a Cellulomonas species. J. Gen. Micribiol. 101: 191-196.

Béguin P and Eisen H, 1978. Purification and partial characterization of three extracellular cellulases from *Cellulomonas* sp. Eur. J. Biochem. 87: 525-531.

Béguin P, Cornet P and Aubert JP, 1985. Sequence of a cellulase gene of the thermophilic bacterium *Clostridium thermocellum*. J. Bacteriol. 162: 102-105.

Berchtold MW, 1989. A simple method for direct cloning and sequencing cDNA by the use of a single specific oligonucleotide and oligo(dT) in a polymerase chain reaction (PCR). Nucleic Acid Res. 17: 453.

Bernier Jr. R, Driguez H and Desrochers M, 1983. Molecular cloning of a *Bacillus subtilis* xylanase gene in *Escherichia coli*. Gene 26: 59-65.

Bigger CH, Murray K and Murray NE. 1983. Recognition sequence of a restriction enzyme. Natl. New Biol. 244: 7.

Bisaria VS and Ghose TK, 1981. Microdegradation of cellulosic materials: substrates, microorganisms, enzymes and products. Enzyme Microb. Technol. 3: 90-104.

Boyer MH, Cami B, Chambost JP, Magnan M and Cattaneo J, 1987a. Characterization of a new endoglucanase from *Erwinia chrysanthemi*. Eur. J. Biochem. 162: 311-316.

Boyer MH, Cami B, Kotoujansky A, Chambost JP, Frixon C and Cattaneo J, 1987b. Isolation of the gene encoding the major endoglucanase of *Erwinia chrysanthemi*: homology between *cel* genes of two strains of *Erwinia chrysanthemi*. FEMS Microbiol. Lett. 41: 351-356.

Brestic-Goachet N, Gunasekaran P, Cami B and Baratti J, 1989. Transfer and expression of an *Erwinia chrysanthemi* cellulase gene in *Zymomonas mobilis*. J. Gen. Microbiol. 135: 893-902.

Bryant MP, 1986. *Ruminococcus*. In: Bergey's Manual of Systematic Bacteriology, V.2, ed: Sneath PHA, Williams & Wilkins, Baltimore, pp. 661-662.

Carrier MJ, Nugent ME, Tacon WCA and Primrose SB. 1983. High expression of cloned genes in *E. coli* and its consequences. Trends Biotechnol. 1: 109.

Catherine Duong TV, Johnson EA and Demain AL, 1983. Thermophilic, anaerobic and cellulolytic bacteria. In: Topics in enzyme and fermentation biotechnology v.7, ed: Wiseman A, Ellis horwood Ltd., NY, pp. 156-195.

Chen CM, Gritzali M and Stafford DW, 1987. Nucleotide sequence and deduced primary structure of cellobiohydrolase II from *Trichoderma reesei*. Biotechnology 5: 274-278.

Cheng KJ, Stemart CS, Dinsdale D and Costerton JW. 1984. Electron microscopy of bacteria involved in the degradation of plant cell walls. Anim. Feed. Sci. Technol. 10: 93-120.

Chesson A, Stewart CS, Dalgarno K and King TP, 1986. Degradation of isolated grass mesophyll, epidermis and fibre cell walls in the rumen and by cellulolytic rumen bacteria in axenic culture. J. Appl. Bacteriol., 60: 327-336.

Chesson A and Forsberg CW, 1988. Polysaccharide degradation by rumen microoganisms. In: The rumen microcrobial ecosystem. ed: Hobson PN, Elsevier Applied Science, London, pp. 251-284.

Churchward G, Linder P and Caro L. 1983. The nucleotide sequence of repliction and maintenance function encoded by plasmid pSC101. Nucleic Acid Res. 11: 5645.

Chynoweth DP and Isaacson R. 1987. Anaerobic digestion of biomass. Elsevier Applied Science. London

Claverie-Martin F, Diaz-Torres MR and Kushner SR, 1987. Analysis of the regulatory region of the protease III (*ptr*) gene of *Escherichia coli* K-12. Gene 54: 185-195.

Clerke L, Hitzeman R and Carbon J. 1979. Selection of specific clones from colony banks by screening with radioactive antibody. Methods Enzymol. 68: 436.

Collmer A and Wilson DB, 1983. Cloning and expression of a *Thermomonospora* yx endocellulase gene in *E. coli*. Biotechnology 1: 594-601.

Coppolecchia R, Dessi MR, Giacomini A, Lepidi A, Mastromei G, Nuti MP and Polsinelli M, 1987. Cloning in *E. coli* of a *Streptomyces* cellulase gene. Biotechnol. Lett. 9: 495-500.

Cornet P, Millet J, Beguin P and Aubert JP. 1983. Characterization of two *cel* (cellulose degradation) genes of *Clostridium thermocellum* coding for endoglucanase. Biotechnology 1: 589-594.

Coughlan MP, 1985. The properties of fungal and bacterial cellulases with comment on their production and application. In: Biotechnology & genetic engineering reviews. v.3 ed: Russell GN, Intercept, Ponteland, pp.39-109.

Curry C, Gilkes N, O'Neill G, Miller Jr. RC and Skipper N, 1988. Expression and secretion of a *Cellulomonas fimi* exoglucanase in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 54: 476-484.

Dehority BA and Scott HW, 1967. Extent of cellulose and hemicellulose digestion in various forages by pure cultures of rumen bacteria. J. Dairy Sci. 50: 1136-1141.

Dehority BA, 1986. Microbes in the foregut of arctic ruminants. In: Control of Digestion and Metabolism in Ruminants. eds: Milligan LP, Grovum WL and Dobson A, Reston, Prentice-Hall, Englewood Cliffs, New Jersey, pp. 307-325.

Devereux J, Haeberli P and Smithies O, 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acid Res. 12: 387-395.

Dunne CP, 1982. Relationship between extracellular proteases and the cellulase complex of *Trichoderma reesei*. Enzyme Engineering 6: 355-356.

Eriksson KE, 1975. Enzyme mechanisms involved in the degradation of wood components. In: Symposium on Enzymatic Hydrolysis of Cellulose. eds: Bailey M, Enari TM, and Linko M, SITRA, Helsinki. pp. 263-280.

Eriksson KE, 1982. Degradation of cellulose. Experientia 38: 156-159.

Flint HJ, McPherson CA and Bisset J, 1989. Molecular cloning of genes from *Ruminococcus flavefaciens* encoding xylanase and β-(1-3,1-4)glucanase activities. Appl. Environ. Microbiol. 55: 1230-1233.

Forsberg CW, Beveridge TJ and Hellstrom AH. 1981. Celulase and xylanase release from *Bacteroides succinogenes* and its importance in the rumen environment. Appl. Environ. Microbiol. 42: 886-896.

Forsberg CW, Crosby B and Thomas DY, 1986. Potential for manipulation of the rumen fermentation through the use of recombinant DNA techniques. J. Anim. Sci. 63: 310-325.

Fukumori F, Kudo T, Narahashi Y and Horikoshi K, 1986. Molecular cloning and nucleotide sequence of the alkaline cellulase gene from the alkalophilic *Bacillus* sp. strain 1139. J. Gen. Microbiol. 132: 2329-2335.

Fukumori F, Kudo T and Horikoshi K, 1987. Truncation analysis of an alkaline cellulase from an alkalophilic *Bacillus* species. FEMS Microbiol. Lett. 40: 311-314.

Fukumori F, Kudo T, Sashihara N, Nagata Y, Ito K and Horikoshi K, 1989. The third cellulase of alkalophilic *Bacillus* sp. strain N-4: evolutionary relationships within the *cel* gene family. Gene 76: 289-298.

Gardner KH and Blackwell J, 1974. The hydrogen bonding in cellulose. Biochi. et Biophy. acta 343: 232-237.

Gasson MJ. 1986. Genetics of *Streptococcus latis* 712. In: Fifth International Symp. on the Genetics of Industrial Microorganism. Alacevic M, Hranueli D and Toman Z. (eds) Ellis Horwood, NY.

Gayda RC, Stephens PE, Hewick R, Schoemaker JM, Dreyer WJ and Markovitz A, 1985. Regulatory region of the heat shock-inducible CapR (*lon*) gene: DNA and protein sequences. J. Bacteriol. 162: 271-275.

Ghangas GS and Wilson DB, 1987. Expression of a *Thermomonospora fusca* cellulase gene in *Streptomyces lividans* and *Bacillus subtilis*. Appl. Environ. Microbiol. 53: 1470-1475.

Ghangas GS and Wilson DB, 1988. Cloning of the *Thermomonospora fusca* endoglucanase E2 gene in *Streptomyces lividans*: affinity purification and functional domains of the cloned gene product. Appl. Envrion. Microbiol. 54: 2521-2526.

Ghose TK, Bailey HJ, Bisaria VS, Enari TM, 1983. Measurement of cellulase activities. Final recommendations. Commission on Biotechnology, International Union of Pure and Applied Chemistry.

Gijsegem F, Toussaint A and Schoonejans E, 1985. *In vivo* cloning of the pectate lyase and cellulase genes of *Erwinia chrysanthemi*. EMBO 4: 787-792.

Gilbert HJ, Jenkins G, Sullivan DA and Hall J, 1987. Evidence for multiple carboxymethylcellulase genes in *Pseudomonas fluorescens* subsp. *cellulosa*. Mol. Gen. Genet. 210: 551-556.

Gilbertson RL, 1980. Wood-rotting fungi of North America. Mycologia 72: 1-49.

Gilkes NR, Langsford ML, Kilburn DG, Miller Jr. RC and Warren RAJ, 1984a. Mode of action and substrate specificities of cellulases from cloned bacterial genes. J. Biol. Chem. 259: 10455-10459.

Gilkes NR, Kilburn DG, Miller Jr. RC and Warren RAJ, 1984b. A mutant of *Escherichia coli* that leaks cellulase activity encoded by cloned cellulase genes from *Cellulomonas fimi*. Biotechnology 2: 259-263.

Gilkes NR, Kilburn DG, Langsford ML, Miller Jr. RC, Wakarchuk WW, Warren RAJ, Whittle DJ and Wong WKR, 1984c. Isolation and characterization of *Escherichia coli* clones expressing cellulase genes from *Cellulomonas fimi*. J. Gen. Microbiol. 130: 1377-1384.

Gilkes NR, Warren RAJ, Miller Jr. RC and Kilburn DG, 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 NR, Kilburn NG, Miller Jr. RC and Warren AJ, 1989. Structural and functional analysis of a bacterial cellulase by proteolysis. J. Bio. Chem. 264: 17802-17808.

Gong CS and Tsao GT, 1979. Cellulase and biosynthesis regulation. Annual Reports on Fermentation Processes 3: 111-139.

Grahem H, Aman P, Theander O, Kolankaya N and Stewart CS. 1985. Influence of heat sterilization and ammoniation on straw composition and degradation by pure cultures of cellulolytic rumen bacterium. Anim. Feed. Sci. Technol. 12: 195-203.

Greenberg NM, Warren RAJ, Kilburn DG and Miller Jr. RC, 1987a. Regulation, initiation and termination of the *cen*A and *cex* transcripts of *Cellulomonas fimi*. J. Bacteriol. 169: 646-653.

Greenberg NM, Warren RAJ, Kilburn DG and Miller Jr. RC, 1987b. Regulation and initiation of *cen*B transcripts of *Cellulomonas fimi*. J. Bacteriol. 169: 4674-4677.

Greenwood JM, Gilkes NR, Kilburn DG, Miller Jr. RC and Warren RAJ, 1989. Fusion to an endoglucanase allows alkaline phosphatase to bind to cellulose. FEBS Lett. 244: 127-131.

Grepinet O and Beguin P, 1986. Sequence of the cellulase gene of *Clostridium thermocellum* coding for endoglucanase B. Nucleic Acids Res. 14: 1791-1799.

Grepinet O. Chebron MC and Beguin P. 1988. Nucleotide sequence and deletion analysis of the xylanase gene (*xyn*Z) of *Clostridium thermocellum* J. Bacteriol. 170: 4582-4588.

Groleau D and Forsberg CW, 1981. Cellulolytic activity of the rumen bacterium *Bacteroides succinogenes*. Can. J. Microbiol. 27: 517-530.

Gryczan T, Shivakumar AG and Dubnau D. 1980. Characterization of chimeric plasmid clonig vehicles in *Bacillus subtilis*. J. Bacteriol. 141: 246-253.

Guo LH and Wu R, 1982. New rapid methods for DNA sequencing based on *Exo*III digestion followed by repair synthesis. Nucleic Acids Res. 10: 2065-2084.

Guo LH, Stepien PP, Tso JY, Brousseau R, Narang S, Thomas DY and Wu R. 1984. Synthesis of human insulin gene. VIII. construction of expression vectors for fused proinsulin production in *Escherichia coli*. Gene 29: 251.

Guo Z, Arfman N, Ong E, Gilkes NR, Kilburn DG, Warren RAJ and Miller Jr. RC, 1988. Leakage of *Cellulomonas fimi* cellulases from *Escherichia coli*. FEMS Microbiol Lett. 49: 279-283.

Hall J and Gilbert HJ, 1988. The nucleotide sequence of a carboxymethylcellulase gene from *Pseudomonas fluorescens* subsp. *cellulosa*. Mol. Gen. Genet. 213: 112-117.

Halliwell G and Bryant MP. 1963. The cellulolytic activity of pure strains of bacteriol from the rumen of cattle. J. Gen. Microbiol. 32: 441-448.

Hamamota t, 1987. Nucleotide sequence of the xylanase A gene of alkalophilic *Bacillus* sp. strain C-125. Agric. Biol. Chem. 51: 953-955.

Hardy KG. 1985. *Bacillus* cloning methods. In: DNA cloning Vol. II. a practical approach. IRL. Press.

Hawley DK and McClure WR, 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. 11: 2237-2255.

Hazlewood GP and Teather RM, 1988. The genetics of rumen bacteria. In: The rumen microbial ecosystem. ed: Hobson PN, Elsevier Applied Science, London, pp. 323-341.

Hedgpeth J, Goodman HM and Boyer HW. 1972. DNA nucleotide sequences restricted by the RI endonuclease. Proc. Natl. Acad. Sci. USA 69: 3448.

Henderson G, Krygsman P, Liu CJ, Davey CC and Malek LT, 1987. Characterization and structure of genes for protease A and B from *Streptomyces griseus*. J. Bacteriol. 169: 3778-3784.

Henikoff S, 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351-359.

Henner DJ, Ferrari E, Perego MA and Hock JA, 1988. Location of the targets of the *hpr*-97, *sac*U32 (Hy), and *sac*Q36 (Hy) mutations in upstream regions of the subtilisin promoter. J. Bacteriol. 170: 296-300.

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

Hernandez PE, 1982. Transport of D-glucose in *Clostridium thermocellum* ATCC 27405. J. Gen. Appl. Microbiol. 28: 469-477.

Hinchliffe E, 1984. Cloning and expression of a *Bacillus subtilis* endo-1,3-1,4-β-D-glucanase gene in *Escherichia coli* K12. J. Gen. Microbiol. 130: 1285-1291.

Hobson PN. 1988. The rumen microbial ecosystem. Elsevier Applied Science. London.

Hofsten BV and Berg B, 1972. Cellulase formation in *Cellvibrio fulvus*. In: Proceedings IV, IFS, Fermentation Technology Today, pp. 731-734.

Hohn B and Murray K. 1977. Packaging recombinant DNA molecule into bacteriphage particles *in vitro*. Proc. Natl. Acad. Sci. USA. 74: 3259.

Honda H, Iijima S and Kobayashi T, 1988a. Cloning and expression in *Saccharomyces cerevisiae* of an endo-β-glucanase gene from a thermophilic cellulolytic anaerobe. Appl. Microbiol. Biotechnol. 28: 57-58.

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

Howard GT, and White BA, 1988. Molecular cloning and expression of cellulase genes from *Ruminococcus albus* 8 in *Escherichia coli* bacteriophage λ. Appl. Environ. Microbiol. 54: 1752-1755.

Hu YJ and Wilson DB, 1988. Cloning of *Thermomonospora fusca* genes coding for beta 1-4 endoglucanase E_1, E_2, and E_5. Gene 71: 331-337.

Huang CM, Kelly WJ, Asmundson RV and Yu PL, 1989. Molecular cloning and expression of multiple cellulase genes of *Ruminococcus flavefaciens* strain 186 in *Escherichia coli*. Appl. Microbiol. Biotechnol. 31: 265-271.

Hungate 1944. Studies on cellulose fermentation I. The culture and physiology of an anaerobic cellulose digesting bacterium. J. Bacteriol 48: 499-513.

Hungate RE. 1950. The anaerobic mesophilic cellulytic bacteria. Bacteriol. Rev. 14: 1-49.

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

Hungate RE, 1966. The rumen and its microbes. pp. 8-10. Academic, New York

Hungate RE and Stack RJ 1982. Phenylpropanoic acid: growth factor for *Ruminococcus albus*. Appl. Environ. Microbiol. 44: 79-83.

Iborra F, Raynal A and Guerineau M, 1988. The promoter of the β-glucosidase gene from *Kluyveromyces fragilis* contains sequences that act as upstream repressing sequences in *Saccharomyces cerevisiae*. Mol. Gen. Genet. 213:150-154.

Imanaka T, Tsunekawa H and Aiba S. 1980. Phenotypic stability of *trp* operon recombinant plasmids in *Escherichia coli* J. Gen. Microbiol. 118: 253.

Imsenecki AA, 1968. In: Ecology of soil bacteria. Gray TRG, Parkinson D (Eds) pp. 257-269. Univ. Toronto Press.

Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F and Boyer HW. 1977. Expression in *Escherichia coli* of a chemically synthesized gne for the hormone somatostatin. Science 198: 1056.

Johansson G, Stahlberg J, Lindeberg G, Engstrom A and Pettersson G. 1989. Isolated fungal cellulase terminal domains and a synthetic minimum analogue bind to cellulose. FEBS Lett. 243: 389-393.

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

Joliff G, Beguin P and Aubert JP. 1986. Nucleotide sequence of the cellulase gene *cel*D encoding endoglucanase D of *Clostridium thermocellum* Nucleic Acids Res. 14: 8605-8613.

Kadam S and Demain AL. 1988. Molecular cloning of a gene for a thermostable β-glucasidase from *Clostridium thermocellum* into *Escherichia coli*. Enzyme Microb. Technol. 10: 9-13.

Karn J, Brenner S. Barnett L and Cesareni G. 1980. Novel bacteriphage lambda cloning vector. Proc. Natl. Acad. Sci USA 77: 5172.

Kawai S, Honda H, Tanase T, Taya M, Iijima S and Kobayashi T, 1987. Molecular cloning of *Ruminococcus albus* cellulase gene. Agric. Biol. Chem. 51: 59-63.

Kawai S, Okoshi H, Ozaki K, Shikata S, Ara K and Ito S, 1988. Neutrophilic *Bacillus* strain, KSM-522, that produces an alkaline carboxymethyl cellulase. Agric. Biol. Chem. 52: 1425-1431.

Kelly WJ, Asmundson RV and Hopcroft DH 1987. Isolation and characterizatin of a strictly anaerobic, cellulolytic spore former: *Clostridium chartatabidum* sp. nov. Arch. Microbiol. 147: 169-173.

Kennedy JF, Philips GO and Williams PA. 1987. Wood and cellulosics. Inductrial utilization, biotechnology, structure and properties. Ellis Horwood. NY.

Kim JM, Kong IS and Yu JH, 1987. Molecular cloning of an endoglucanase gene from an alkalophilic sp. and its expression in *Escherichia coli*. Appl. Environ. Microbiol. 53: 2656-2659.

Kim H, and Pack MY, 1988. Endo-β-1,4-glucanase encoded by *Bacillus subtilis* gene cloned in *Bacillus megaterium*. Enzyme Microb. Technol. 10: 347-351.

Koide Y, Nakamura A, Uozumi T and Beppu T, 1986. Molecular cloning of a cellulase gene from *Bacillus subtilis* and its expression in *Escherichia coli*. Agric. Biol. Chem. 50: 233-237.

Kolankaya N, Stewart CS, Duncan SH, Cheng KJ and Costerton JW. 1985. The effect of ammonia treatment on the solubilization of straw and the growth of cellulolytic rumen bacteria. J. Appl. Bacteriol. 58: 371-379.

Kotoujansky A, Diolez A, Boccara M, Bertheau Y, Andro T and Coleno A, 1985. Molecular cloning of *Erwinia chrysanthemi* pectinase and cellulase structural genes. EMBO 4: 781-785.

Lamed R and Zeikus JG, 1980. Ethanol production by thermophilic bacteria: relationship between fermentation product yields and catabolic enzyme activities in *Clostridium thermocellum* and *Thermoanaerobium brockii*. J. Bacteriol. 144: 569-578.

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

Lamed R and Bayer EA, 1988. The cellulosome of *Clostridium thermocellum*. Adv. Appl. Microbiol. 33: 1-46.

Langsford M, Gilkes NR, Wakarchuk WW, Kilburn DG, Miller Jr. RC and Warren RAJ, 1984. The cellulase system of *Cellulomonas fimi*. J. Gen. Microbiol. 130: 1367-1376.

Langsford MC, Gilkes NR, Singh B, Moser B, Miller jr. RC, Warren RA and Kilburn DG. 1987. Glycosylation of bacterial cellulases prevent proteolytic cleavage between functional domains. FEBS lett. 225: 163-167.

Larsen JEL, Gerdes K, Light J and Molin S. 1984. Low-copy-number plasmid-coding vectors amplifiable by derepressing of an inserted foreign promoter. Gene 28: 45-54.

Latham MJ, Brooker BE, Pattipher GL and Harris PJ. 1978. Adhension of *Bacteroides succinogenes* in pure culture and in the presence of *Ruminococcus flavefaciens* to cell walls in leaves of perennial ryegrass (L. Perenae). Appl. Environ. Microbiol. 35: 1166-1173.

Leatherwood JM, 1973. Cellulose degradation by *Ruminococcus*. Fed. Proceed. 32: 1814-1818.

Leclerc M, Chemardin P, Arnaud A, Ratomahenina R, Galzy P, Gerbaud C and Raynal A, 1986. Biosynthesis regulation of the β-glucosidase produced by a yeast strain transformed by genetic engineering. Arch. Microbiol. 146: 115-117.

Lee DS and Pack MY, 1987. Use of bacilli for overproduction of exocellular endo-β-1,4-glucanase encoded by cloned gene. Enzyme Microb. Technol. 9: 594-597.

Lee SB, Shin HS and Ryu DDY, 1982. Adsorption of cellulase on cellulose: effect of physicochemical properties of cellulose on adsorption and rate of hydrolysis. Biotech. Bioeng. 24: 2137-2153.

Lejeune A, Colson C and Eveleigh DE, 1986. Cloning of an endoglucanase gene from *Pseudomonas fluorescens* var. *cellulosa* into *Escherichia coli* and *Pseudomonas fluorescens*. J. Indust. Microbiol. 1: 79-86.

Lejeune A, Courtois S and Colson C, 1988a. Characterization of an endoglucanase from *Pseudomonas fluorescens* subsp. *cellulosa* produced in *Escherishia coli* and regulation of the expression of its cloned gene. Appl. Environ. Microbiol. 54: 302-308.

Lejeune A, Dartois V and Colson C, 1988b. Characterization and expression in *Escherichia coli* of an endoglucanase gene of *Pseudomonas fluorescens* subsp. *cellulosa.* Biochi. Biophy. Acta. 950: 204-214.

Lejeune A, Eveleigh DE and Colson C, 1988c. Expression of an endoglucanase gene of *Pseudomonas fluorescens* var. *cellulosa* in *Zymomonas mobilis*. FEMS Microbiol. Lett. 49: 363-366.

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

Lin E and Wilson DB, 1988b. Transcription of the *cel*E gene in *Thermomonospora fusca*. J. Bacteriol. 170: 3838-3842.

Linko M, Ratto M, Vilkari L and Bailey M, 1982. Organisms and enzymes for hydrolysis of cellulose and xylan. In: Proceedings of the International Symposium on Ethanol form Biomass. eds: Duckworth HE and Thompson EA, The Royal Society of Canada, Winnipeg, pp.371-393.

Linko M, 1985. Enzymatic hydrolysis-present status and future developments. In: Bereitstellung und Verwertung von Lignocellulosen. Status-und Expertenseminar Workshop. Kernforschungsanlage Julich GMBH.

Ljungdahl L, Pettersson B, Eriksson KE and Wiegel J, 1983. A yellow affinity substancce invloved in the cellulolytic system of *Clostriduum thermocellum*. Curr. Microbiol. 9: 195-200.

Lo AC, MacKey RM, Seligy VL and Willick GE, 1988. *Bacillus subtilis* β-1,4-endoglucanase products from intact and truncated genes are secreted into the extracellular medium by *Escherichia coli*. Appl. Environ. Microbiol. 54: 2287-2292.

Love DR and Streiff MB 1987. Molecular cloning of a β-glucosidase gene from an extremely thermophilic anaerobic in *E. coli* and *B. subtilis*. Biotechnology 5:384-387.

Love DR, Fisher R and Bergquist PL, 1988. Sequence structure and expression of a cloned β-glucosidase gene from an extreme thermophile. Mol. Gen. Genet. 213: 84-92.

MacKay RM, Lo A, Willick G, Zuker M, Baird S, Dove M, Moranelli F and Seligy V, 1986. Structure of a *Bacillus subtilis* endo-β-1,4-glucanase gene. Nuc. Acids Res. 14: 9159-9170.

Mandels M, 1981. Ann. Reports on Ferm. Processes. V.5

Manfioletti G and Schneider C, 1988. A new and fast method for preparing high quality lambda DNA suitable for sequencing. Nucl. Acid. Res. 16: 2873-2884.

Maniatis T, Fritsch EF and Sambrook J, 1982. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Maxam AM and Gilbert W, 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74: 560-564.

Maxam AM and Gilbert W, 1980. Sequencing end-labled DNA with base-specific chemical cleavages. Methods Enzymol. 65: 499-560.

McGavin M, Forsberg CW, Crosby B, Bell AW, Dignard D and Thomas DY, 1989. Structure of the *cel*-3 gene from *Fibrobacter succinogenes* S85 and characteristics of the encoded gene product, endoglucanase 3. J. Bacteriol. 171: 5587-5595.

McNeil M, Darvill AG, Fry SC and Albersheim P, 1984. Structure and function of the primary cell walls of plants. Ann. Rev. Biochem. 53: 625-664.

Merivuori H, Siegler KM, Sands JA and Montenecourt BS, 1985. Regulation of cellulase biosynthesis and secretion in fungi. Biochem. Soc. Transact. 13: 411-414.

Mertz JE and Davis RW. 1972. Cleavage of DNA by RI restriction endonuclease generates cohensive ends. Proc. Natl. Acad. Sci. USA 69: 3370.

Messing J, Crea R and Seeburg PH, 1981. A system for shotgun DNA sequencing. Nucl. Acids Res. 9: 309-321.

Messing J, 1983. New M13 vectors for cloning. Methods Enzymol. 101: 20-78.

Messner R and Kubicek CP. 1988. Intracellular presusor of endo-β-1,4-glucanase in *Trichoderma reesei*. FEMC Microbiol. Lett. 50: 227-232.

Millet J, Petre D, Beguin P, Raynaud O and Aubert JP, 1985. Cloning of ten distinct DNA fragments of *Clostridium thermocellum* coding for cellulases. FEMS Microbiol. Lett. 29: 145-149.

Monstein HJ and Geijer T. 1986. A rapid and inexpensive method for preparing *E. coli* plasmid DNA. Biochem. International 12: 889-896.

Montenecourt BS, 1983. *Trichoderma reesei* cellulases. Trends in Biotech. 1: 156-161.

Morna Jr. CP, Lang N, LeGrice SFJ, Lee G, Stephans M, Sonenshein AL, Pero J and Losick R, 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. 186: 339-346.

Morosoli R and Durand S, 1988. Molecular cloning of mRNA sequences encoding xylanase from *Cryptococcus albidus*. FEMS Microbiol. Lett. 51: 217-224.

Morris EJ and Van Gylswyk NP, 1980. Comparison of the action of rumen bacteria on cell walls of Eragrostis tef. J. Agric. Sci. 95: 313-323.

Morris EJ, 1984. Degradation of the intact plant cell wall of subtropical and tropical herbage by rumen bacteria. In: Herbivore nutrition in the subtropics and tropics. eds: Gilchrist FMC and Mackie RI, The Science Press, South Africa, pp. 378-395.

Moser B, Gilkes NR, Kilburn DG, Warren RAJ and Miller Jr. RC, 1989. Purification and characterization of endoglucanase C of *Cellulomonas fimi*, cloning of the gene, and analysis of in vitro transcripts of the gene. Appl. Environ. Microbiol. 55: 2480-2487.

Murphy N, McConnell DJ and Cantwwell BA. 1984. The DNA sequence of the gene and genetic control sites for the excreted *B. subtilis* enzyme β-glucanase. Nucleic Acids Res. 12: 5355- 5367.

Murray K. 1977. Applications of bacteriphage lambda in recombinant DNA research. In: Molecular cloning of recombinant DNA. Werner S (ed) Vol. 13. Academic Press. NY

Murray NE, 1983. In: Lambda II. Hendrix RW, Roberts JW and Weraberg RA (Eds) pp. 395-432. Cold Spring Harbor.

Nakai R, Horinouchi S and Beppu T, 1988. Cloning and nucleotide sequence of a cellulase gene, *cas*A, from an alkalophilic *Streptomyces* strain. Gene 65: 229-238.

Nakamura A, Uozumi T, and Beppu T, 1987. Nucleotide sequence of a cellulase gene of *Bacillus subtilis*. Eur. J. Biochem. 164: 317-320.

Nebreda AR, Villa TG, Villanueva JR and Rey F, 1986. Cloning of genes related to exo-β-glucanase production in *Saccharomyces cerevisiae*: characterization of an exo-β-glucanase structural gene. Gene 47: 245-259.

Ng TK and Zeikus JG. 1981. Comparison of extracellular cellulase activities of *Clostridium thermocellum* LQRI and *Trichoderma reesei* QM9414. Appl. Environ. Microbiol. 42: 231-240.

Ng TK and Zeikus JG, 1982. Differential metabolism of cellobiose and glucose by *Clostridium thermocellum* and *Clostridiu thermohydrosulfuricum*. J. Bacteriol. 150: 1391-1399.

Ohmiya K, Shimizu M, Taya M and Shimizu S. 1982. Purification and properties of cellobiosidase from *Ruminococcus albus*. J. Bacteriol. 150: 407-409.

Ohmiya K, Shirai M, Kurachi Y and Shimizu S. 1985. Isolation and properties of β-glucosidase from *Ruminococcus albus*. J. Bacteriol. 161: 432-434.

Ohmiya K, Nagashima K, Kajino T, Goto E, Tsukada A and Shimizu S, 1988. Cloning of the cellulase gene from *Ruminococcus albus* and its expression in *Escherichia coli*. Appl. Environ. Microbiol. 54: 1511-1515.

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

Old RW and Primrose SB. 1985. Principles of gene manipulation: an introduction to genetic engineer. Blackwell Scientific Publications. Oxford, England.

O'Neill GP, Warren RAJ, Kilburn DG and Miller Jr. RC, 1986a. Secretion of *Cellulomonas fimi* exoglucanase by *Escherichia coli*. Gene 44: 331-336.

O'Neill GP, Goh SH, Warren RAJ, Kilburn DG and Miller Jr. RC, 1986b. Structure of the gene encoding the exoglucanase of *Cellulomonas fimi*. Gene 44:325-330.

O'Neill GP, Kilburn DG, Warren RAJ and Miller Jr. RC, 1986c. Overproduction from a cellulase gene with a high guanosine-plus-cytosine content in *Escherichia coli*. Appl. Environ. Microbiol. 52: 737-743.

Ong E, Gilkes NR, Warren RAJ, Miller Jr. RC and Kilburn DG, 1989a. Enzyme immobilization using the cellulose-binding domain of a *cellulomonas fimi* exoglucanase. Biotechnology 7: 604-607.

Ong E, Greenwood M, Gilkes NR, Kilburn DG, Miller Jr. RC and Wreen AJ, 1989b. The cellulose-binding domains of cellulases: tools for biotechnology. Trends in Biotech. 7: 239-243.

Owolbai JB, Béguin P, Kilburn DG, Miller Jr. RC and Warren RAJ, 1988. Expression in *Escherichia coli* of the *Cellulomonas fimi* structural gene for endoglucanase B. Appl. Environ. Microbiol. 54: 518-523.

Paice MG, Desrochers M, Rho D, Jurasek L Roy C, Rollin CF, DeMiguel E and Yaguchi M, 1984. Two forms of endoglucanase from the basidiomycete *Schizophyllum commune* and their relationship to other β-1,4-glucoside hydrolases. Biotechnology 2: 535-539.

Panbangred W, Kondo T, Negoro S, Shinmyo A and Okada H, 1983. Molecular cloning of the genes for xylan degradation of *Bacillus pumilus* and their expression in *Escherichia coli*. Mol. Gen. Genet. 192: 335-341.

Paradis FW, Warren RAJ, Kilburn DG and Miller Jr. RC, 1987. The expression of *Cellulomonas fimi* cellulase genes in *Brevibacterium lactofermentum*. Gene 61: 199-206.

Park SH and Pack MY, 1986. Cloning and expression of a *Bacillus* cellulase gene in *Escherichia coli*. Enzyme Micro. Technol. 8: 725-728.

Patel GB, Khan AW, Agnew BJ and Colvin JR, 1980. Isolation and characterization of an anaerobic, cellulolytic microorganism, *Acetivibrio cellulolyticus* gen. nov., sp. nov. Int. J. System. Bacteriol. 30: 179-185.

Pearson WR. 1982. Automatic construction of restriction site maps. Nucleic Acid Res. 10: 217-227.

Pettipher GL and Latham MJ, 1979. Characteristics of enzymes produced by *Ruminococcus flavefaciens* which degrade plant cell walls. J. Gen. Microbiol. 110: 21-27.

Priest F. 1977. Extracellular enzyme synthesis in the genus *Bacillus*. Bacteriol. Rev. 41: 711-753.

Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ, Jensen MA and Baumeister K, 1987. A system for rapid DNA sequencing with fluorescent chainterminating dideoxynucleotides. Science 238: 336-341.

Radford SE, Laue ED, Perham RN, Martin SR and Appella E, 1989. Conformational flexibility and folding of synthetic peptides represneting an interdomain segment of polypeptide chain in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. J. Biol. Chem. 264: 767-775.

Rees, DA, 1977. Polysaccharide shapes. In: Outline Studies in Biology. Chapman and Hall, London, pp. 41-60.

Rees DA, Morris ER, Thom D and Madden JK, 1982. Shapes and interactions of carbohydrate chains. In: Polysaccharides, V.1, Academic Press, New York, pp. 195-290.

Reese ET, 1977. Degradation of polymeric carbohydrates by microbial enzymes. Recent. Adv. Phytochem. 11: 311-367.

Robson LM and Chambliss GH. 1984. Characterization of the cellulolytic activity of a *Bacillus* isolate. Appl. Environ. Micribiol. 47: 1039-1046.

Robson LM and Chambliss GH, 1986. Cloning of the *Bacillus subtilis* DLG β-1,4-glucanase gene and its expression in *Escherichia coli* and *B. subtilis*. J. Bacteriol. 165: 612-619.

Robson LM and Chambliss GH, 1987. Endo-β-1,4-glucanase gene of *Bacillus subtilis* DLG. J. Bacteriol. 169: 2017-2025.

Romaniec MPM, Clarke NG and Hazlewood GP, 1987. Molecular cloning of *Clostridium thermocellum* DNA and the expression of further movel endo-β-1,4-glucanase genes in *Escherichia coli*. J. Gen. Microbiol. 133: 1297-1307.

Romaniec MPM, Davidson K, White BA and Hazlewood GP, 1989. Cloning of *Ruminococcus albus* endo-β-1,4-glucanase and xylanase genes. Lett. Appl. Microbiol. 9: 101-104.

Russell JB, 1985. Fermentation of cellodextrins by cellulolytic and no-cellulolytic rumen bacteria. Appl. Environ. Microbiol. 49: 572-576.

Russell JB and Wilson DB, 1987. Potential opportunities and problems for genetically altered rumen microorganisms. J. Nutr. 118: 271-279.

Ryu DD and Mandels M, 1980a. Cellulases: biosynthesis and applications. Enzyme Microb. Technol. 2: 91-102.

Ryu DD and Mandels M, 1980b. Competitive adsorption of cellulase components and its significance in a synergistic mechanism. Biotech. Bioeng. 26: 488-496.

Ryu DD, Kim C and Mandels M. 1984. Competitive adsorption of cellulose components and its significance in a synergstic mechnism. Biotech. Bioeng. 26: 488-496.

Sacco M, Millet J and Aubert JP. 1984. Cloning and expression in *Saccharmyces cererisiae* of a cellulase gene from *Clostridium thermocellum*. Ann. Microbiol. (Inst. Pasteur) 135A: 485-488.

Salvovuori I, Makarow M Rauvala H, Knowles J and Kaariainen L. 1987. Low molecular weight high mannose type glcans in a secreted protein of the filamenous fungus *Trichoderma reesei*. Biotechnology 5: 152-156.

Sandhu S and Kennedy JF, 1986. Molecular cloning of the gene for β-D-xylosidase of *Bacillus polymyxa* and its expression in *Escherichia coli*. Enzyme Microb. Technol. 8: 677-680.

Sanger F, Nicklen S and Coulson AR, 1977. DNA sequencing with chain-terminating inhibitors. Proc. Acad. Sci. USA 74: 5463-5467.

Sashihara N, Kudo T and Horikishi K, 1984. Molecular cloning and expression of cellulase genes of alkalophilic sp. strain N-4 in *Escherichia coli*. J. Bacteriol. 158: 503-506.

Schmuck M, Pilz I, Hayn M and Esterbauer H, 1986. Investigation of cellobiohydrolase from *Trichoderma reesei* by small angle x-ray scattering. Biotechnol. Lett. 8: 397-402.

Schwarz W, Bronnenmeier K and Staudenbauer, 1985. Molecular cloning of *Clostridium tehrmocellum* genes involved in β-glucan degradation in bacteriophage lambda. Biotechnol. Lett. 7: 859-864.

Schwarz WH, Grabnitz F and Staudenbaner WL. 1986. Properties of a *Clostridium thermocellum* endoglucanase produced in *Escherichia coli*. Appl. Environ. Microbiol. 51: 1293-1299.

Schwarz WH, Schmming S and Staudenbauer WL. 1987. High-level expression of *Clostridium thermecellum* cellulase genes in *Escherichia coli*. Appl. Microbiol. Biotechnol. 27: 50-56.

Sedmak JJ and Grossberg SE. 1977. A rapid, sensitive and versatile assay for protein using Coomassie Brilliant Blue G250. Anal. Biochem. 79: 544-552.

Sharma P, Gupta JK, Vadehra DV and Dube DK, 1987. Molecular cloning and expression in *Escherichia coli* of a thermophilic *Bacillus* sp. PDV endo-β-1,4-glucanase gene. Enzyme Microb. Technol. 9: 602-606.

Shine J and Dalgarno L. 1975. Determinant of cistron specificity in bacterial ribosomes. Nature 254: 34

Shoemaker S, Watt K, Tsitivsky G and Cox R. 1983. Characterization and properties of cellulases purified from *Trichoderma reesei* strain L27. Biotechnology 1: 687-690.

Sijpesteijn AK, 1951. On *Ruminococcus flavefaciens* a cellulose decomposing bacterium from the rumen of sheep and cattle. J. Gen. Microbiol. 5: 869-879.

Simon LD, Randolph B, Irwin N and Binkowski G. 1983. Stabilization of proteins by a bacteriophage T4 gene cloned in *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 80: 2059.

Sloma A, Ally A, Ally D and Pero J, 1988. Gene encoding a minor extracellular protease in *Bacillus subtilis*. J. Bacteriol. 170: 5557-5563.

Son KH, Jang JH and Kim JH, 1987. Effect of temperature on plasmid stability and expression of cloned cellulase gene in a recombinant *Bacillus megaterium*. Biotechnol. Lett. 9: 821-824.

Sprey B and Lambert C, 1983. Titration curves of cellulases from *Trichoderma reesei*: demonstration of a cellulase-xylanase-β-glucosidase-containing complex. FEMS Microbiol. Lett. 18: 217-222.

Stack RJ and Hungate RE, 1984. Effect of 3-phenylpropanoic acid on capsule and cellulases of *Ruminococcus albus* 8. Appl. Environ. Microbiol. 48: 218-223.

Stack RJ and Cotta MA, 1986. Effect of 3-phenylpropanoic acid on growth of and cellulose utilisation by cellulolytic rumen bacteria. Appl. Environ. Microbiol. 52: 209-210.

Stackebrandt E and Kandler O. 1979. Taxonomy of the genus *Cellulomonas* based on phenotypic characteris and deoxyribonucleic acid-deoxyribonucleic acid homology and proposal of seven neotype strains. Int. J. Syst. Bacteriol. 29: 273-282.

Staden R and McCachlan AD, 1982. Coden preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acid Res. 10: 141-156.

Stahlberg J, Johansson G and Pettersson G, 1988. A binding-site-deficient, catalytically active, core protein of endoglucanase III from the culture filtrate of *Trichoderma reesei*. Eur. J. Biochem. 173: 179-183.

Stark JR and Yin XS, 1987. Fractionation of a commerical cellulase preparation from *Penicillium funiculosum* and its use in the purification of a water soluble α-glucan from milled barley. Enzyme Microb. Technol. 9: 156-160.

Sternberg. D. 1976. Production of cellulase by *Trichoderma*. Biotech. Bioeng. Symp. 6: 35-53.

Stewart CS, Dinsdale D, Cheng KJ and Paniagua C, 1979. The digestion of straw in the rumen. In: Straw decay and its effect on disposal and utilisation. ed: Grossbard E, Wiley, Chichester, pp. 123-130.

Stewart CS and Kuncan SH, 1985. The effect of avoparcin on cellulolytic bacteria of the ovine rumen. J. Gen. Microbiol. 131: 427-435.

Stewart CS, 1986. Rumen function with special reference to fibre digestion. In: Anaerobic Bacteria in Habitats other than man. eds: Barnes EM and Mead GC, Blackwell Scientific Publishers, Oxford, pp. 263-286.

Stewart CS and Bryant MP, 1988. The rumen bacteria. In The rumen microbial ecosystem. ed: Hobson PN, Elsevier Applied Science, London, pp. 21-75.

Stutzenberger F. 1987. Selective adsorption of endoglucanase from *Thermomonospora curvata* on protein-extracted lucerne fibres. Lett. Appl. Microbiol. 5: 1-4.

Swisher EJ, Storvick WO and KIng KW. 1964. Metabolic nonnequivalence of the two glucose moieties of cellobiose in *Cellvibrio gilvus*. J. Bacteriol 88: 817-820.

Tabor S and Richardson CC, 1987. DNA sequence analysis with modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84: 4767-4771.

Talmadge K and Gilbert W, 1982. Cellular location affects protein stability in *Escherichia coli*. Proc. Natl. Acad. Sci. 79: 1830-1833.

Tanaka T and Kawata M, 1988. Cloning and characterization of *Bacillus subtilis iep*, which has positive and netative effects on prodiction of extracellular protease. J. Bacteriol. 170: 3593-3600.

Teather RM and Wood PJ, 1982. Use of Congo red polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl. Environ. Microbiol. 43: 777-780.

Teeri TT, Lehtovaara P, Kauppinen S, Salovuori I and Knowles J, 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. Gene 51: 43-52.

Tomme P, VanTilbeurgh H, Pettersson G, 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM9414, analysis of domain function in two cellobiohydrolases by limited preteolysis. Eur. J. Biochem. 170: 575-581.

Tribe HT, 1960. Ecology of soil fungi. pp. 246-256. Liverpool Univ. Press.

Tucker ML, Durbin ML, Clegg MT and Lewis LN, 1987. Avocado cellulase: nucleotide sequence of a putative full-length cDNA clone and evidence for a small gene family. Plant Mol. Biol. 9: 197-203.

Vieira J, Messing J, 1982. The pUC plasmids, an M13*mp*7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19: 259-268.

Viljoen JA, Fred EB and Peterson WH 1926. The fermentation of cellulose by thermophilic bacteria. J. Agri. Sci. 16: 1-17.

Wakarchuk WW, Kilburn DG, Miller Jr. RC and Warren RAJ, 1984. The preliminary characterization of the β-glucosidases of *Cellulomonas fimi*. J. Gen. Microbiol. 130: 1385-1389.

Wakarchuk WW, Greenberg NM, Kilburn DG, Miller Jr. RC and Warren RA, 1988. Structure and transcription analysis of the gene encoding a cellobiase from *Agrobacterium* sp. strain ATCC 21400. J. Bacteriol. 170: 301-307.

Wald S, Wilke CR and Blanch HW, 1984. Kinetics of the enzymatic hydrolysis of cellulose. Biotech. Bioeng. 26: 221-230.

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

Warren RAJ, Beck CF, Gilkes NR, Kilburn DG, Langsford ML, Miller Jr. RC, O'Neill GP, Scheufens M and Wong WKR, 1986. Sequence conservation and region shuffling in an endoglucanase and an exoglucanase from *Cellulomonas fimi*. Protein 1: 335-341.

Warren RAJ, Gerhard B, Gilkes R, Owolabi JB, Kilburn DG and Miller Jr. RC, 1987. A bifunctional exoglucanase-endoglucanase fusion protein. Gene 61: 421-427.

Weiss B. 1976. Endonuclease II of *Escherichia coli* is exonuclease III. J. Biol. Chem. 251: 1896.

West CA, Elzanowski A, Yeh LS and Barker WC, 1989. Homologous of catalytic domains of *Cellulomonas* glucanases found in fungal and *Bacillus* glycosidases. FEMS Microbiol. 59: 167-172.

White AR and Brown RM Jr., 1981. Enzymatic hydrolysis of cellulose: visual characterization of the process. Proc. Natl. Acad. Sci. USA 78: 1047-1051.

Whittle DJ, Kilburn DG, Warren RAJ and Miller Jr. RC, 1982. Molecular cloning of a *Cellulomonas fimi* cellulase gene in *Escherichia coli*. Gene 17: 139-145.

Wilke CR, Yang RD and Von Stockar U, 1976. Preliminary case analysis for enzymatic hydrolysis of newsprint. Biotech. Bioeng. 6: 155-176.

Williams AG and Withers SE. 1982. The production of plant cell wall polysaccharide-degrading enzymes by hemicellulolytic rumen bacterial isolates grown on a range of carbohydrate substrate. J. appl. Bacteriol. 52: 377-387.

Winterburn PJ, 1974. Polysaccharide structure and function. In: Companion to Biochemistry: Selected topics for further reading. eds: Bull AT, Lagnado JR, Thomas JO and Tipton KE, Longman, London. pp. 307-341.

Wiseman A. 1985. Topics in enzyme and fermentation biotechnology. Ellis Horwood, NY.

Withers SG, Dombroski D, Berven LA, Kilburn DG, Miller Jr. RC, Warren RAJ and Gilkes NR, 1986. Direct ^{1}H N.M.R. determination of the stereochemical course of hydrolyses catalysed by glucanase components of the cellulase complex. Biochem. Biophy. Res. Comm. 139: 487-494.

Wong SL, Price CW, Goldfarb DS and Doi RH. 1984. The subtilisin E gene of *Bacillus subtilis* is transcribed from a σ^{37} promoter *in vitro*. Proc. Natl. Acad. Sci. USA 81: 1184-1188.

Wong WKR, Gerhard B, Guo ZM, Kilburn DG, Warren RAJ and Miller Jr. RC, 1986. Characterization and structure of an endoglucanase gene *cen*A of *Cellulomonas fimi*. Gene 44: 315-324.

Wong WKR, Curry C, Parekh RS, Wayman, Davies RW, Kilburn DG and Skipper N, 1988. Wood hydrolysis by *Cellulomonas fimi* endoglucanase and exoglucanase coexpressed as secreted enzymes in *Saccharomyces cerevisiae*. Biotechnology 6: 713-719.

Wood TM and McCrae SI, 1972. The purification and properties of the C_1 component of *T. koningii* cellulase. Biochem. J. 128: 1183-1192.

Wood TM. 1975. Properties and mode of action of cellulases. Biotech. Bioeng. Symp. 5: 111-137.

Wood TM. 1980. Enzymes and mechanisms involved in fungal degradation of cellulasic materials. In: OECD workshop No.2 on conversion of lignocellulosic substrate to simple carbohydrates. Rijkens BA ed. IBVL. Wageningen.

Wood TM, 1981. Enzyme interactions involved in fungal degradation of cellulosic materials. In: The Ekman-Days International Symposium on Wood Pulping Chemistry, V.3, SPCI, Stockholm. pp. 31-38.

Wood TM, Wilson CA and Stewart CS, 1982. Preparation of the cellulase from the cellulolytic anaerobic rumen bacterium *Ruminococcus albus* and its release from the bacterial cell wall. J. Biochem. 105: 129-137.

Wood TM and Wilson CA, 1984. Some properties of the endo-(1,4)-β-D-glucanase synthesised by the anaerobic cellulolytic rumen bacterium *Ruminococcus albus*. Can. J. Microbiol. 30: 316-321.

Wood TM, 1985a. Observations and speculations on the complex interactions involved in the solubilization of native cellulose. In: Proceedings 16th FEBS, Moscow.

Wood TM, 1985b. Properties of cellulolytic enzyme systems. Biochem. Soc. Transact. 13: 407-410.

Yaguchi M, Roy C, Rollin CF Paice MG and Jurasek L, 1983. A fungal cellulase shows sequence homology with the active site of hen egg-white lysozyme. Biochem. Biophy. Res. Comm. 116: 408-411.

Yang M, Ferrari E, Chen E and Henner DJ, 1986. Identification of the pleiotropic *sac*Q gene of *Bacillus subtilis*. J. Bacteriol. 166: 113-119.

Yang RCA, MacKenzie CR, Bilous D, Seligy VL and Narang SA, 1988. Molecular cloning and expression of a xylanase gene from *Bacillus polymyxa* in *Escherichia coli*. Appl. Environ. Microbiol. 54: 1023-1029.

Yang RCA, MacKenzie CR, Bilous D and Narang SA, 1989a. Identification of two distinct *Bacillus circulans* xylanases by molecular cloning of the genes and expression in *Escherichia coli*. Appl. Environ. Microbiol. 55: 568-572.

Yang RCA, MacKenzie CR, Bilous D and Narang S, 1989b. Hyperexpression of a *Bacillus circulans* xylanase gene in *Escherichia coli* and characterization of the gene product. Appl. Environ. Microbiol. 55: 1192-1195.

Yanischi-Perron C, Vieira J and Messing J, 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13*mp*18 and pUC 19 vectors. Gene 33: 103-109.

Yoshikawa T, Suzuki H and Nisizawa K, 1974. Biogenesis of multiple cellulase components of *Pseudomonas fluoresces* var. *cellulosa* I. effects of culture conditions on the multiplicity of cellulase. J. Biochem. 75: 531-540.

Zappe H, Jone WA, Hone DT and Wood DR. 1988. Structure of an endo-β-1,4-glucanase gene from *Clostridium acetobutylicum* P262 showing homology with endoglucanase genes from *Bacillus* spp. Appl. Environ. Microbiol. 54: 1289-1292.