THE EXPRESSION OF *CELLULOMONAS FIMI* CELLULASE GENES IN *BREVIBACTERIUM LACTOFERMENTUM* AND CHARACTERIZATION OF RECOMBINANT *C. FIMI* ß-GLUCOSIDASE A FROM *E. COLI*.

By

FRANÇOIS WILLIAM PARADIS

B.Sc., Université LAVAL, 1983

M.Sc., Université LAVAL, 1987

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Department of Microbiology)

(Genetics program)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April, 1990

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Microbiology

The University of British Columbia
Vancouver, Canada

Date May 14th, 1990

DE-6 (2/88)

ABSTRACT

In the first part of this thesis, I describe the expression of *C. fimi* cellulase genes in the closely related *Brevibacterium lactofermentum* by generating a shuttle vector able to replicate selectively in the latter and carrying full length cellulase-encoding genes. The expression of those genes apparently originated from some unpredicted regulatory sequences, possibly located within the vector itself. The enzymatic activity was mostly found in the culture medium in *B. lactofermentum* indicating that the organism secreted the enzymes. The putative *C. fimi* promoter sequences did not function in *B. lactofermentum*, making difficult the analysis of their roles in expression of *C. fimi* cellulase genes.

In the second part of this thesis, I describe the characterization of a recombinant *C. fimi* exo-ß-1,4-glucosidase (CbgA) expressed in *E. coli*. The purified enzyme had a M_r of 183 kDa and hydrolysed various ß-glucosides with a preference for cello-oligosaccharides in the order C5>C4>C3>C2. The intact CbgA polypeptide was not required for enzymatic activity since removal of about 700 residues from the amino terminus did not reduce activity. The purified enzyme was used to raise polyclonal antibodies which in turn were used to identify the corresponding enzyme in *C. fimi*. During the fractionation of *C. fimi* ß-glucosidases, several enzymes hydrolyzing various ß-glucosides were isolated together with the native CbgA, which was present in the culture medium as part of a protein aggregate.

Part of the nucleotide sequence of the 7.2 kb insert was determined. Alignments of the N-terminal amino acid sequences of the purified CbgA and truncated polypeptides with the partial nucleotide sequence of the cloned *C. fimi* DNA showed that precise excision was responsible for the appearance of a truncated form of CbgA. Alignment of the amino-terminal sequence of a CbgA:Cex_{CBD} fusion peptide indicated that the pre-mature CbgA starts with a putative leader sequence of 49 amino acids which is followed by a region rich in Pro and Ala residues. Two GTG translational initiation codons followed by sequences resembling

prokaryotic ribosome binding sites and separated by a large open reading frame were identified from data obtained after *in vitro* site-directed mutagenesis of the most upstream initiation codon suggesting that internal re-initiation may occur and that upstream regulatory sequences had not been isolated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A.A. ; amino acid
abg ; *Agrobacterium sp.*gene encoding ß-glucosidase (Abg)
Ap ; ampicillin
ATCC ; American Type Culture Collection
ATP ; adenosine triphosphate
ß-gal ; ß-galactosidase
Bgl ; ß-glucosidase
bp ; base pair (s)
BSA ; bovine serum albumine
CBD ; cellulose binding domain
*cbg*A ; *C. fimi* gene encoding ß-glucosidase A (CbgA)
CbgB ; *Cellulomonas fimi* ß-glucosidase B
CbgC : *Cellulomonas fimi* ß-glucosidase C
CbgD ; *Cellulomonas fimi* ß-glucosidase D
*cen*A ; *C. fimi* gene encoding endoglucanase A (CenA)
*cen*B ; *C. fimi* gene encoding endoglucanase B (CenB)
*cen*C ; *C. fimi* gene encoding endoglucanase C (CenC)
cex ; *C. fimi* gene encoding exoglucanase (Cex)
CIAP ; calf intestine alkaline phosphatase
CMC ; carboxymethyl cellulose
DEAE ; diethylaminoethyl
DMSO ; dimethylsulfoxide
DNA ; deoxyribonucleic acid
DNAse 1 ; deoxyribonuclease 1
DNS ; dinitrosalicylic acid
dNTP ; deoxynucleoside triphosphate (dATP, dCTP, dGTP, dTTP)
ds ; double stranded
DTT ; dithiothreitol
EDTA ; ethylenediaminetetraacetic acid
ELISA ; enzyme-linked-immunoadsorbant assay
Eng ; endo-ß-1,4-glucanase
Exg ; exo-ß-1,4-glucanase

FPLC ; fast protein liquid chromatography
HEPES ; N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
Kb ; 1000 base pair (s)
kDa ; 1000 Daltons
Km ; kanamycin
Lac Z' ; the first 78 amino acid of *E. coli* ß-galactosidase
lac Zpo ; *lac* promoter operator
LB(ls) ; Luria broth (low salt)
LPS ; lipopolysaccharide
MIC ; minimal inhibitory concentration
MOPS ; morpholinepropanesulfonic acid
Mr ; relative molecular mass
mRNA ; messenger RNA
MUG ; 4-methylumbelliferyl-ß-D-glucoside
MUC ; 4-methylumbelliferyl-ß-D-cellobioside
MUGase ; enzyme which hydrolyses MUG
NAD ; nicotinamide adenine dinucleotide
NADP ; nicotinamide adenine dinucleotide phosphate
nt ; nucleotide (s)
O.D. ; optical density
ORF ; open reading frame
PAGE ; polyacrylamide gel electrophoresis
PEG ; polyethylene glycol
PMSF ; phenylmethylsulfonylfluoride
PNK ; polynucleotide kinase
pNPC ; p-nitrophenyl-ß-D-cellobioside
pNPCase; enzyme which hydrolyses pNPC
pNPG ; p-nitrophenyl-ß-D-glucoside
pNPGase; enzyme which hydrolyses pNPG
PSI ; pounds per square inch
RBS ; ribosome binding site
RNA ; ribonucleic acid
RNase A ; ribonuclease A
RPM ; revolution per minute
RS ; modified R2-medium

SDS ; sodium dodecyl sulfate
SET ; Sucrose/Tris/EDTA
s/n ; supernatant
ss ; single stranded
SSC ; standard sodium citrate
TBE ; Tris/Borate/EDTA buffer
TE ; Tris/EDTA buffer
Tc^R ; tetracycline resistance determinant
TSA ; trypticase soy agar
TSB ; trypticase soy broth

ACKNOWLEDGEMENTS

I would like to thank those whom I have trusted, for their guidance and support throughout the course of this work: My supervisor Dr. R.C. Miller, Jr., my committee members Drs. R.A.J. Warren, D.G. Kilburn, S. Withers and C. Douglas, and special thanks to my friend and acting member, Dr. J.T. Beatty (chance favours the prepared mind, L. Pasteur). I wish to thank Dr. N.R. Gilkes for the oligo-saccharides and discussions, Ms. S. Keilland for the amino acid sequences and Mr. T. Atkinson for the synthesis of oligonucleotides. To my fellow graduate students, staff and faculty of the department of Microbiology-many thanks for the unforgettable memories.

Finallement, je tiens à remercier les membres de ma famille pour la patience et le support moral donnés tout au long de ce travail. Cette thèse fût financée en partie par mon père, le Dr. Desmond Paradis. Je dédie cette thèse à celui qui a su éveiller mon esprit scientifique, le Dr. V. Smirnoff, et à celle à qui de loin me protège, ma grand-mère G. Biron.

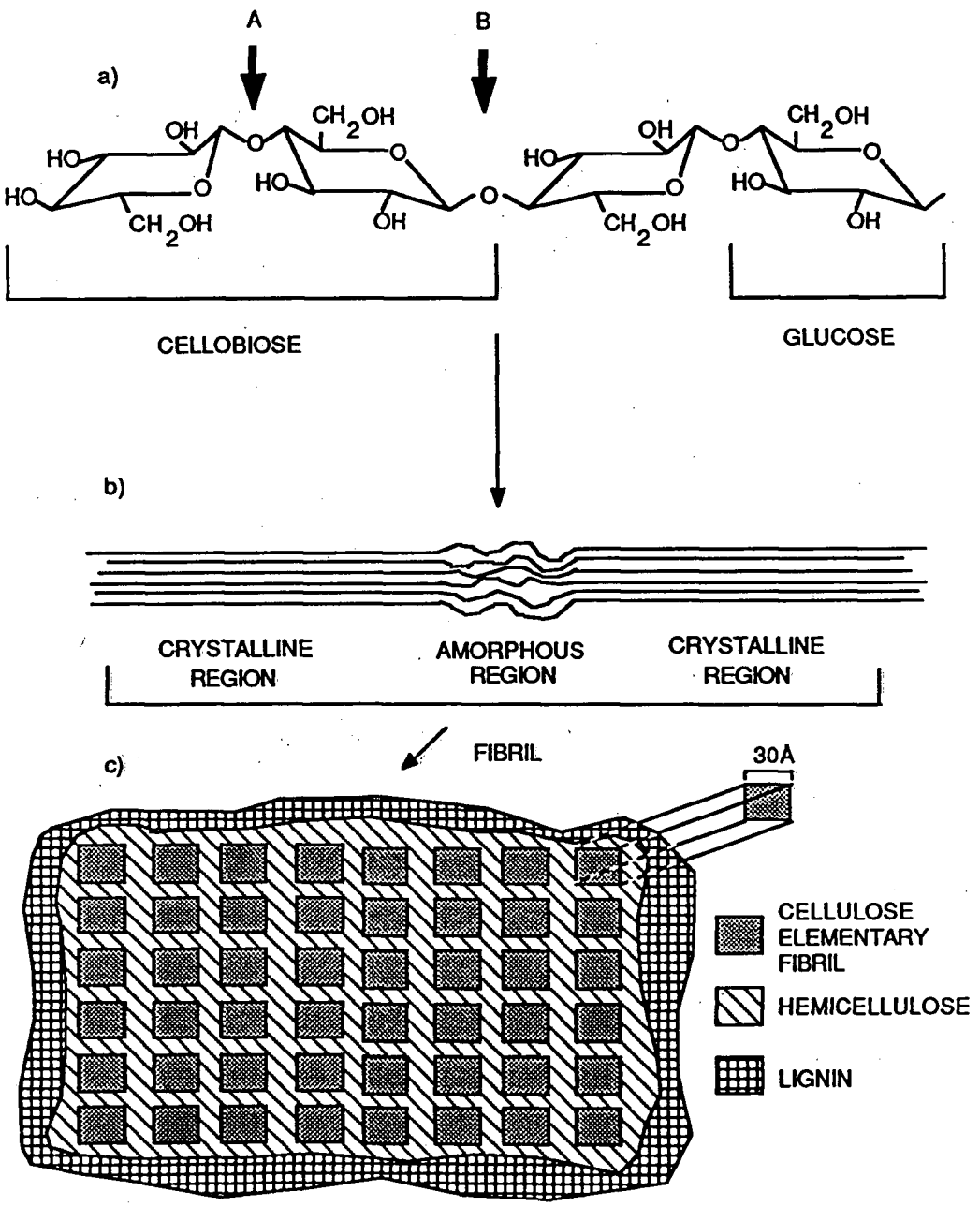
1. INTRODUCTION

1.1. Background

Plant biomass is the most abundant source of carbohydrate available on earth. This nearly unlimited amount of stored solar energy is extensively used by most species, including human, effecting its primary benefits, mainly the production of oxygen and being the first element in the food chain. The scientific community is now concerned about the long term depletion of cellulosic material and methods to improve its utilization by recycling of industrial wastes has become a major concern. Because of its abundance, research has been focussed on the conversion of biomass into more suitable products such as ethanol and acetone, amino acids and nucleotides, even single-cell protein and the production of recombinant material.

In nature, cellulose molecules composed of 100 to 10,000 glucose subunits linked by ß-1,4-glucosidic bonds are associated in a highly resistant matrix with lignin and hemicellulose (Fig. 1). This association forms a major structural barrier to the availability of cellulose, and efficient pretreatment methods which promote the complete hydrolysis of cellulose are practically non-existent. As a result of strong inter- and intramolecular hydrogen bonding, cellulose is insoluble in water and in most organic solvents. Nevertheless, a number of cellulose solvents exist, namely, strong mineral acids, quaternary ammonium bases, non-aqueous organic solvents like DMSO or even metal complex solvents like iron-tartaric acid-alkali. Cellulose can be enzymatically degraded by a set of enzymes called cellulases. Enzymatic degradation is slow because of the structure of the substrate but seems to be a preferred choice because of the absence of corrosion and environmental pollution. A variety of organisms can efficiently hydrolyze crystalline cellulose using a process which requires the synergistic action of several enzymes with various substrate specificities. All cellulolytic enzymes are found to have multiple forms which differ in their relative activities on a broad range of substrate. The complex mechanism of enzymatic hydrolysis of cellulose has been reviewed recently (Coughlan, 1985; Beguin *et al*, 1987).

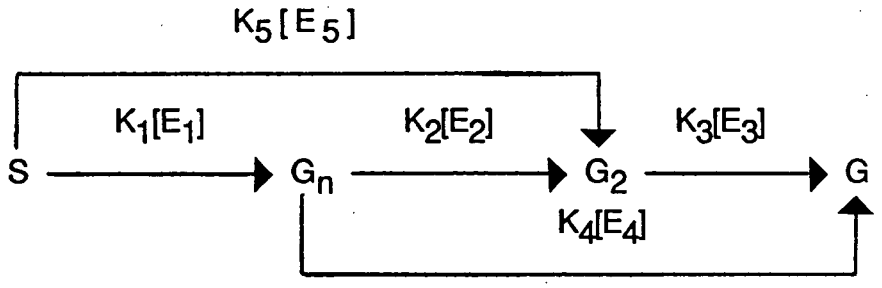
Fig.1. Structure of cellulose. (a) Stereochemical representation of the cellulose molecule. Arrows A and B represent the ß-1,4-linkages within the cellulose fibril. (b) Organisation of cellulose molecules in the elementary fibril. Highly ordered or relative disarray are represented by the crystalline and amorphous regions, respectivelly. (c) Cross-section of a wood fiber. The elementary fibrils are surrounded by hemicellulose and lignin, reducing the accessibility of cellulose to enzymatic degradation (adapted from Fan *et al.,* 1980).



1.2. The enzymatic hydrolysis of crystalline cellulose

Studies on the extracellular enzymes from a number of cellulolytic fungi, particularly *Trichoderma* species, have provided us with much of our basic information on the mechanism of cellulase action (Ericksson and Wood, 1985). However, other mechanisms exist in bacteria where the optimal conditions of pH and temperature may be different.

The overall kinetic scheme can be represented by this rather simplified form:



where S is the original substrate, G_n are cello-oligosaccharides with degree of polymerization greater than 2, G_2 is cellobiose and G is glucose. As can be seen from this scheme, glucose can be formed either from higher oligosaccharides or from cellobiose. The rate of glucose formation essentially depends on the physical state of the substrate. The rate of hydrolysis decreases with solubility of the substrate in the order CM-cellulose (soluble)- filter paper- microcrystaline cellulose (insoluble).

Cellulose hydrolysis requires the synergistic action of various enzymes (Fig.2). In cellulase systems able to degrade crystalline cellulose, three main types of enzymes are found (Shewale, 1982). Endo-1,4-ß-D-glucanases (EC 3.2.1.4, 1,4-ß-D-glucan glucanohydrolase) hydrolyze the 1,4-ß-glucosidic bonds of cellulose molecules randomly generating new ends, and decreasing the degree of polymerization rapidly. The prolonged action of endoglucanases on long molecules of cellulose results in soluble material, mainly cellodextrins and some cellobiose. The rate of attack by endoglucanase which is the first enzyme in the multienzyme chain, decreases progressively with the extent of conversion of cellulose as a result of the exhaustion of more accessible regions of the substrate. The exo-

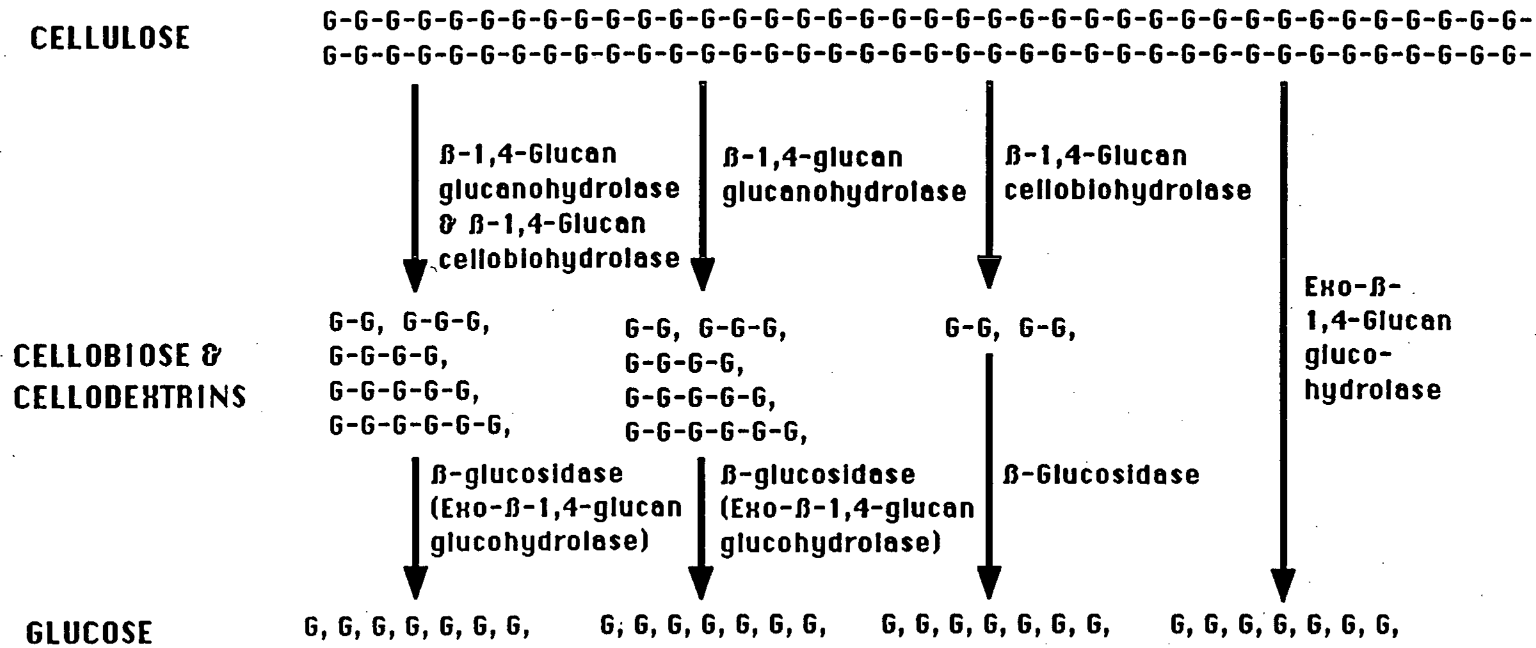


Fig 2. Possible routes of glucose formation in saccharification of cellulose (Shewale, 1982)

cellobiohydrolases (EC 3.2.1.91, 1,4-ß-D-glucan cellobiohydrolase) act by removing cellobiose from the non-reducing ends of cellulose chains. Exoglucanases when present are major components in some cellulase systems but are usually absent from most of them. Finally, ß-glucosidases (EC 3.2.1.74, exo-1,4-ß-glucan glucohydrolase and EC 3.2.1.21, ß-D-glucoside glucohydrolase) are very important components of these systems because they complete the hydrolysis of soluble cellodextrins into glucose. The ß-glucosidases remove the inhibitory effect of cellobiose on endoglucanases and exoglucanases and could represent the rate limiting step in the saccharification of cellulose (Coughlan, 1985). It is assumed that glucose is formed almost exclusively by the hydrolysis of cellobiose by cellobiases. In *Trichoderma longibrachiatum*, *T. reesei*, *T. lignorum* and *Aspergillus feotidus*, the selective inhibition of cellobiase activity by gluconolactone has shown that in the initial stages of hydrolysis of cellulose, exoglucosidase plays the main role in the formation of glucose from soluble and insoluble polymeric substrates (Sinitsyn and Klesov, 1981). Due to the fact that until that time there was no suitable method for determining exoglucosidase activity, exoglucosidases are among the least studied components of the cellulase complex. The properties of ß-glucosidases previously reviewed by Shewale (1982) and Woodward and Wiseman (1982) will be presented since this thesis is largely concerned with the characterization of recombinant *C. fimi* ß-glucosidase A (CbgA).

1.3 The cellulases of *Cellulomonas fimi* and their genes.

C. fimi is a Gram-positive coryneform facultative anaerobic bacterium with DNA of 72 mole % G+C (Stackebrant and Kandler, 1979). By secreting into the culture medium a complex mixture of cellulolytic enzymes (Langsford *et al.*, 1984), *C. fimi* can grow on microcrystalline cellulose as the sole source of carbon and energy. It is one of the most intensively studied cellulolytic bacteria at both the biochemical and genetic levels (Beguin *et al.*, 1987).

The initial biochemical analysis of *C. fimi* cellulolytic enzymes has identified up to 10 different cellulolytic components in culture supernatants, some of which could result from proteolysis or change in glycosylation (Langsford *et al.*, 1984). More recently our studies of cellulose hydrolysis

by *C. fimi* extracellular enzymes have identified at least four different cellulases. The major components of this cellulolytic system are CenA and Cex, endoglucanase and exoglucanase respectively. The two other enzymes, CenB and CenC, are endoglucanases which are minor components of the system. The multiplicity of cellulolytic enzymes may be related to the broad array of substrates generated during the hydrolysis of microcrystalline cellulose. Our ultimate goal has been the characterization of all *C. fimi* components that are secreted and involved in cellulose hydrolysis.

1.3.1. Molecular cloning and characteristics

The molecular cloning of a *C. fimi* cellulase gene was first reported by Whittle *et al.* (Whittle *et al.*, 1982). So far, five cellulase encoding genes have been isolated from this organism and characterized in *E. coli* (Gilkes *et al.*, 1984; Bates, 1987; Moser *et al.*, 1989).

Biochemical analyses of the native and recombinant forms of Cex and CenA has revealed unusual characteristics for these enzymes. On SDS-PAGE, the native endoglucanase A has an M_r of 53.0 kDa, whereas its recombinant counterpart has an M_r of 48.7 kDa. Native exoglucanase has an M_r of 49.3 kDa but the corresponding recombinant enzyme has an M_r of 47.1 to 47.3 kDa (Gilkes *et al.*, 1988). The discrepancies observed in migration between the native and recombinant enzymes reflects glycosylation of the native enzymes (Langsford *et al.*, 1987). *C. fimi* gCex and gCenA are believed to have O-linked sugars which may act in protecting substrate bound forms against proteolytic cleavage.

CenA and Cex contain three distinct regions: an irregular, hydroxyamino acid rich, low charge density region and an ordered, high charge-density region, separated by a short sequence of about 20 amino acids containing only Pro and Thr residues (Fig. 3). Both enzymes hydrolyze CM-cellulose but only the exoglucanase hydrolyses pNPC and MUC (O'Neill, 1986). CenA and Cex in the presence of salt bind strongly to the insoluble substrate Avicel, a microcrystalline cellulose of degree of polymerization between 100 and 250. The cellulose binding domain of CenA (CBD-CenA) is located at the N-terminal part of the enzyme whereas the cellulose binding domain of Cex (CBD-Cex) is located at its C-terminus.

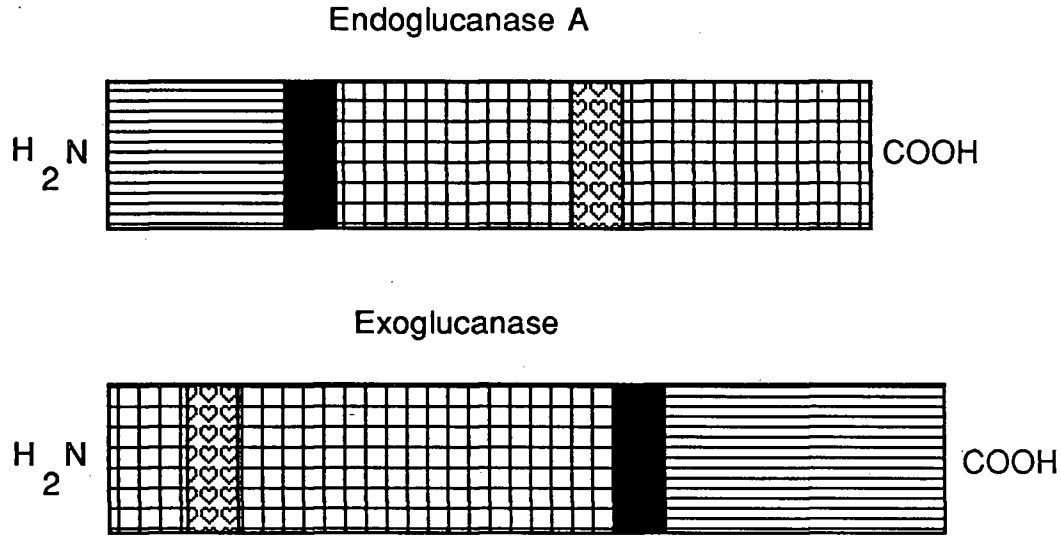


Fig 3. Structures of CenA and Cex.

- denotes the Pro-Thr box
- denotes the active site
- denotes the cellulose binding domain
- denotes the ordered charged region

Recently, Ong *et al.* (1989) have reported the construction of a fusion between the 5'-end of the *abg* gene of an *Agrobacterium sp.* encoding a ß-glucosidase and the 3'-end of the *cex* gene encoding the *C. fimi* CBD-Cex. The hybrid protein was immobilized on Whatman CF1 cellulose column and could retain 48% of its activity. The binding properties were identical to those of Cex in that the bound proteins could be eluted from the cellulose column with water. The non-covalent binding of CBD-Cex to cellulose is strong and stable enough to be used in enzyme immobilization. The CBD of CenA was also used to bind enzymes to cellulose when the CBD was fused at the N-termini (Greenwood *et al.,* 1989). Therefore, fusions of the CBD's at either ends of a molecule can be achieved depending on which *C. fimi* gene is more convenient. The other *C. fimi* enzymes, CenB and CenC, are endoglucanases which hydrolyze CM-cellulose. CenB binds weakly to autoclaved Avicel (Owalabi *et al.,* 1988) whereas CenC does not seem to be substrate associated (Moser, 1988). Apparently, these two enzymes are not glycosylated. A summary of the characteristics of the *C. fimi* cloned genes and the corresponding enzymes is shown in Table I.

1.3.2. The *cex* gene and its product

The *cex* structural gene initiating with an ATG codon is 1452 bp in length and encodes a mature polypeptide of 484 amino acids (O'Neil *et al.,* 1986a). The DNA sequence predicts a 41 amino acid leader peptide the function of which is to export the Exg to the periplasmic space of *E. coli* (O'Neil *et al.,* 1986a). A more detailed analysis of the recombinant Exg has demonstrated the necessity of its leader sequence for export (O'Neil *et al.,* 1986b). By removing all the amino acids of the leader sequence and fusing the mature enzyme to six amino acids of the N-terminus of ß-gal, most of the Exg activity remained in the cytoplasm of *E. coli*(O'Neil *et al.,* 1986b). Both secreted native and recombinant enzymes have been purified to homogeneity (Gilkes *et al.,* 1988). Transcription of the *cex* gene in *E. coli* is not driven by the endogenous *C. fimi* DNA 5'-flanking sequence as shown by the requirement of an external promoter for its expression in the Gram-negative bacterium (Greenberg, 1988).

TABLE I. *C. fimi* cellulases and their genes

Gene	Start codon	ORF (bps)	stop codon	Protein	#A.A. Leader	#A.A. Mature	Specific transcript CMC	Gly	Glu	ref.
cex	ATG	1452	TGA	Cex	41	484	++	+/-	-	1
*cen*A	ATG	1347	TGA	CenA	31	449	++	-	-	2
*cen*B	ATG	3135	TGA	CenB	33	1011	++	+	+	3
*cen*C	GTG	3306	TGA	CenC	32	1180	++	-	-	4

ORF; open reading frame

#A.A.; number of amino acids per polypeptide

++, +, +/-, -; indicates the relative amount ofspecific mRNA detected from cells grown on various carbon sources (CM-cellulose, glycerol, glucose).

1. O'Neill, 1986
2. Wong, 1986
3. Owalabi, 1988; Meinke, personal communication
4. Moser, 1988; Coutinho, personal communication

1.3.3. The *cen*A gene and its product

The *cen*A structural gene initiating with an ATG codon is 1347 bps in length and encodes a mature polypeptide of 449 amino acids. The nucleotide sequence predicts a leader peptide of 31 amino acids, responsible for the export of EngA into the periplasm of *E. coli* (Wong *et al.,* 1986). The location of the native enzyme is extracellular and both native and recombinant forms have been purified to homogeneity (Gilkes *et al.,* 1984a; Wong *et al.,* 1986). The putative *cen*A 5'-flanking ribosome binding site appears to be functional in *E. coli* but the expression of the *cen*A gene in *E. coli* is also dependent on the presence of an external promoter (Wong *et al.,* 1986).

1.3.4. The *cen*B gene and its product

The *cen*B gene encodes a polypeptide of M_r 110 kDa (Owalabi *et al.,* 1988). From the size of this peptide, one could predict that the structural gene would require at least 3 Kb of coding sequence. Some attempts to purify this minor component of the *C. fimi* cellulolytic system have been unsuccessful, but the recombinant enzyme was purified by affinity chromatography on Avicel (Owalabi *et al.,* 1988). The coding sequence starts with an ATG codon. The predicted amino acid sequence includes a 33 amino acid long leader peptide which allows export of the recombinant EngB to the periplasm of *E. coli* (Owalabi *et al.,* 1988). The transcription of *cen*B in *E. coli* also requires the participation of an external promoter. The putative *cen*B ribosome-binding site appears to be functional in *E. coli* (Owalabi *et al.,* 1988).

1.3.5. The *cen*C gene and its product

The DNA sequence of *cen*C encodes two related polypeptides of M_r 130 and 120 kDa. Both recombinant and native polypeptides share the same N-termini and kinetic parameters (Moser *et al.,* 1989). These minor components of the *C. fimi* cellulolytic system, which are apparently not substrate associated, have been purified to homogeneity and require up to 3306 bp of DNA coding sequence (J.B. Coutinho, personal communication). The nucleotide sequence predicts putative 32 amino acid long leader

peptide initiating at a GTG codon.

1.3.6. The *cbg*A gene and its product

The recombinant *C. fimi* ß-glucosidase A (CbgA) is a phosphate-independent aryl-ß-D-glucosidase which preferentially hydrolyses pNPG, but has some activity towards cellobiose (Bates, 1987). The level of expression of *cbg*A in *E. coli* cells carrying the recombinant plasmid pUC13:62 was low, so expression was increased considerably by deleting the 5' end of the insert. The recombinant enzyme may correspond to a secreted enzyme detected in the culture medium of *C. fimi* grown on CMC.

1.3.7. Transcriptional analysis of *C. fimi* cellulase genes

Transcriptional analyses of *cen*A, *cex*, *cen*B, *cen*C and *clg* (or cex-linked gene) by S1 nuclease mapping identified several 5'-mRNA putative start sites (Greenberg *et al*, 1987a,b; Moser *et al.*, 1989). The analysis of putative promoter sequences has revealed a region of 50 bp which display up to 64% homology that could be involved in expression of the *C. fimi* genes (Greenberg, 1988). The closest homologies were found with promoters from *Streptomyces spp.*, a Gram-positive bacterium with high % G+C in its DNA. When RNA was isolated from *E. coli* cells carrying recombinant plasmids with putative *C. fimi* promoter sequences, no transcripts were found which initiated within the inserts (Greenberg *et al.*, 1987b). However, there is no definitive evidence for the presence of the putative *C. fimi* promoter sequences on the cloned DNA fragments since no *in vivo* or *in vitro* transcriptional systems are available for this organism.

1.4. Heterologous expression of *C. fimi* cellulase genes.

1.4.1. Expression in *E. coli*

The first report in 1982 (Whittle *et al.*, 1982) on the heterologous expression of a *C. fimi* cellulase gene was in the host *E. coli* C600 carrying the recombinant plasmid pDW1 (a pBR322 derivative carrying a 20.2 Kb-*Bam*HI insert) expressing Exg activity. Soon after, three clones (namely pEC1, 2 and 3 encoding Cex, CenA and CenB, respectively) expressing various cellulase activities were isolated. Significant fractions of the activities were exported to the periplasm of *E. coli* (Gilkes *et al.*, 1984). Surprisingly, pEC2 encoded a fusion polypeptide in which CenA, lacking

its leader peptide and the first 45 amino acids of the mature protein, was fused to the Tc^R determinant (Wong *et al,* 1986). Cloning of the entire gene led to 50% of the CenA activity being in the periplasm, but there was little activity in the culture supernatant (Wong *et al.,* 1986). An *E. coli* strain leaking up to 40% of total cellulase activity as well as other enzymes into the culture medium was isolated following mutagenesis with nitrosoguanidine (Gilkes *et al.,* 1984b).

The *cex* gene was over expressed by fusing it to a synthetic ribosome-binding site and an initiating ATG which were under the control of the lambda P_L promoter contained on the runaway replication vector pCP3. Up to 20% of total *E. coli* protein was Cex (O'Neil *et al.,* 1986b). The overproduced enzyme formed insoluble aggregates which could be recovered by centrifugation and solubilized in 6M urea or 5M guanidine hydrochloride. However, active enzyme was not recovered.

Subcloning of *cen*B from pBR322 to pUC19 increased its level of expression several fold (Owolabi *et al.,* 1988). A total of seven polypeptides of M_r 110 kDa and lower were produced in *E. coli* carrying pJB301 plasmid, a pUC19:*cen*B construct. The putative CenB leader peptide allowed transport of CenB into the periplasm of *E. coli,* even when the N-terminal 17 amino acids were replaced with part of the *lacZ'* alpha-peptide (Owolabi *et al.,* 1988) suggesting that other regions of the protein may be involved in its secretion.

*Cen*C was detected when *E. coli* cells carrying the recombinant plasmid pTZP-*cen*C were incubated at 30°C but not at 37°C. Apparently, expression of this gene reduces cell viability (Moser, 1988). Electron microscopic analysis demonstrated the aggregation of the recombinant material around the bacterial chromatin (Moser, 1988). Under noninduced condition, *E. coli* JM101 carrying pTZP-*cen*C transported to its periplasm up to 58% of the total CenC enzymatic activity. The pre-enzyme was processed correctly by *E. coli* leader peptidase (Moser *et al.,* 1989). Furthermore, a difference of about 10 kDa in size between the cytoplasmic form and the periplasmic form of the recombinant CenC suggested a proteolytic processing of more than the CenC leader peptide can account for suggesting a processing at the C-terminus of this enzyme (Moser, 1988).

The lack of affinity of CenC for crystalline cellulose could indicate its preference for soluble intermediates generated during cellulose hydrolysis.

1.4.2. Expression in other Gram-negative bacteria

Since the native substrate containing cellulose is often characterized by a high ratio of carbon:nitrogen, the growth of microorganisms that hydrolyze cellulose could become more efficient when supplemented with an additional source of nitrogen (Bisaria and Ghose, 1981). In an attempt to combine cellulose hydrolysis and nitrogen fixation in a single free living organism, the *cex, cen*A and *cen*B genes were subcloned individually into a broad-host-range vector and transferred by conjugation into the photosynthetic bacterium *Rhodobacter capsulatus* (Johnson *et al.,* 1986). A shuttle vector compatible with *Rhodobacter capsulatus* DNA replicative and conjugative mechanisms allowed expression of the cellulase genes under the control of the *R. capsulatus* oxygen-regulated *puf* operon promoter formally known as the *rxc*A promoter. Significant cellulase activity was detectable in cell extracts of *E. coli* and the expression of the cellulase genes was dependent on the presence of the external promoter (Johnson *et al.,* 1986).

With the objective of growing non-cellulolytic bacteria on cellulose, the *cex* and *cen*A genes were placed together under the control of the divergent *tet* promoters of transposon Tn10 and transferred into the broad-host-range plasmid, pJRD215 (Din, 1989). Both genes were expressed in all tested bacteria (*E. coli* C600 MM294, *E. coli* K12 W4860 (Cel+), *R. capsulatus* B10 and *K. pneumoniae* M5a1). Nevertheless, none of them secreted cellulase activity into the culture medium and transformants were unable to grow on cellulose.

1.4.3. Expression in the yeast *Saccharomyces cerevisiae*

The secretion of cellulases is required for their action on cellulose. In contrast to Gram-negative bacteria, Gram-positive bacteria and even the eucaryotic yeasts are better hosts for expression and secretion of *C. fimi* cellulases because of their ability to secrete various proteins into the culture medium. Expression vectors were constructed carrying the *cen*A or the *cex* gene of *C. fimi* and transferred into the yeast *Saccharomyces cerevisiae.* The coding sequences for the mature cellulases were fused to

leader sequences, ribosome-binding sites and promoter sequences of proteins secreted by yeast (Skipper *et al.,* 1985; Curry *et al.,* 1988). *S. cerevisiae* did produce low levels of extracellular glycosylated cellulases but required appropriate yeast leader peptide sequences for secretion of the enzymes. Subsequently, both cellulase genes were inserted in tandem on the expression vector pMV2Adel (Wong *et al.,* 1988). Both enzymes were produced and secreted by *Saccharomyces cerevisiae* using the alpha-Gal signal peptide but the cells were unable to grow on cellulose suggesting that perhaps both enzymes are not sufficient to sustain growth on this substrate.

1.4.4. Expression in Gram-positive bacteria

The Gram-positive bacterium *Bacillus subtilis* secretes a variety of polypeptides, and is well characterized physiologically and genetically. The *cex* gene was subcloned into several plasmid vectors which replicate in *B. subtilis* including an *E. coli/ B. subtilis* shuttle vector. The *cex* gene was not maintained in *B. subtilis,* being excised from all the plasmids tested (D.J. Whittle and V. Tai, personal communication). The instability of the gene may have resulted from the marked difference in the moles per cent G+C contents of *B. subtilis* DNA (41%) and that of *C. fimi* DNA (72%).

1.5. Cloning systems in Coryneform bacteria

Coryneform bacteria are part of the family of Gram-positive, aerobic, non-sporulating irregular rod shaped bacteria with a high genomic % G+C content (54% to 73%) that are widely distributed in nature (Komagata *et al.,* 1969; Bousfield and Callely, 1978). These include *Corynebacterium, Brevibacterium, Arthrobacter, Microbacterium* and *Cellulomonas* genera. Some are pathogens of plants and animals, including man, others are used in industrial amino acid fermentations (Yamada *et al.,* 1972). Despite their physical similarities, DNA hybridization studies as well as 16S and 5S rRNA sequence analyses suggest a clear heterogeneity amongst the species included in these genera (Martin *et al.,* 1987).

Most genetic studies on coryneform strains have focused on the isolation and characterization of amino acid biosynthetic mutants for the production of high levels of selected amino acids. More recently, recombinant DNA techniques through the development of appropriate cloning vectors and transformation systems have become available in several *Corynebacteria* (Santamaria *et al.,* 1984; Katsumata *et al.,* 1984; Santamaria *et al.,* 1985; Yoshihama *et al.,* 1985; Yeh *et al.,* 1986; Singer, 1986; Shaw and Harley, 1988). A large number of endogenous plasmids has now been reported in *Corynebacteria* and used as starting material for vector construction (Martin *et al.,* 1987). Since most of the plasmids are cryptic, this involves the introduction of selectable markers into endogenous plasmids. The selectable markers used so far have been mostly antibiotic resistance genes isolated from different bacteria. Nevertheless, a need for additional markers is evident since only a few are expressed enough to allow direct selection: the *aph* II gene from Tn5, the genes conferring streptomycin and spectinomycin resistance from pCG4, the chloramphenicol and hygromycin resistance genes from *Streptomyces* and perhaps a few more markers that are still poorly characterized (Martin *et al.,* 1987).

Problems related to the construction of cloning vectors in *Corynebacteria* is the lack of knowledge of the indispensable

sequences involved in plasmid establishment and instability of large hybrid plasmids likely due to recombination proficiency of the bacterial host. Furthermore, all published transformation procedures for *Corynebacteria* are based on the treatment of protoplasts with DNA and PEG. The removal of the corynomycolic acid-containing cell wall is difficult to achieve without affecting cell viability, thereby lowering transformation efficiency. Only one in 10,000 protoplasts is transformed under the best experimental conditions (Martin *et al.,* 1987). Other factors, like the secretion of nucleases into the medium and the existence of a restriction-modification system in the host cell leading to low frequencies of transformation have been reported (Katsumata *et al.,* 1984; Smith *et al.,* 1986; Paradis *et al.,* 1987). More recently, transformation of bacterial cells by electroporation techniques has been achieved in *Corynebacteria* and seems promising in terms of better efficiencies (Chassy *et al.,* 1988).

1.6. The first objective of this thesis.

From the perspective of studying heterologous promoter functionality and secretion of cellulases, two *C. fimi* cellulase genes were transferred into the closely related strain *Brevibacterium lactofermentum* (Paradis *et al.,* 1987; this thesis). The use of this industrial strain was believed to allow secretion of the enzymes in the culture medium because of the properties of the leader peptides present on those enzymes. Furthermore, the presence of putative *C. fimi* promoter sequences on the cloned fragments may allow expression of their corresponding genes or display some promoter functionality in *B. lactofermentum* allowing the analysis of those involved in gene expression.

B. lactofermentum is a coryneform bacterium whose genomic DNA contains 64 mol % G+C (Bousfield and Callely, 1978). It and several derivatives are used industrially for the production of L-glutamic acid, L-phenylalanine, L-leucine, L-threonine and L-lysine (Yamada *et al.,* 1972a; Nakamori *et al.,* 1987). It contains the 4.4 Kb cryptic plasmid pBLI (Santamaria *et al.,* 1984) that has been used widely in the

construction of cloning vectors (Santamaria *et al.,* 1984; Miwa *et al.,* 1985; Yeh *et al.,* 1986; Martin *et al.,* 1987), one of which was used to express the *B. subtilis* alpha-amylase gene in *B. lactofermentum* from its own promoter (Smith *et al.,* 1986). As well as being a coryneform bacterium with high % G+C genomic content, *B. lactofermentum* possesses a well established transformation procedure (Santamaria *et al.,* 1984). Expression of cloned *C. fimi* genes in the related organism *Brevibacterium lactofermentum* might allow the role of the individual enzymes to be determined in the absence of other contaminating cellulases. Furthermore, because the gene products might be secreted, production and purification of enzymes could be simplified.

1.7. Characteristics of ß-glucosidases

ß-Glucosidases (ß-D-glucoside glucohydrolase ; EC 3.2.1.21) are enzymes which hydrolyze various compounds with ß-D-glucosidic linkages. They are widely distributed in nature and have been isolated from bacteria, fungi, plants and animals (Shewale, 1982; Woodward and Wiseman, 1982; Dinur *et al.,* 1986). This discussion considers only those involved in the hydrolysis of cellulose to glucose as a component of a cellulase complex.

ß-Glucosidases are not strictly speaking cellulases but they are very important components of cellulase systems. ß-Glucosidases can hydrolyze cellobiose, remove glucose from the non-reducing ends of cello- oligosaccharides, transglycosylate and hydrolyze various artificial ß-D-glucosides (Coughlan, 1985). Their transglycosylation activities may generate inducers of the synthesis of cellulase complexes (Coughlan, 1985). Low levels of ß-glucosidase activity can result in poor saccharification of cellulose because of the inhibition of endoglucanases and cellobiohydrolases by cellobiose (Shewale, 1982). Adequate levels of ß-glucosidase activity are therefore essential to ensure maximal conversion of cellulose to glucose.

Before the hydrolysis of cellulose becomes a commercial process, several problems concerning ß-glucosidases must be addressed. Possible solutions to these problems have already been suggested (Woodward and Wiseman, 1982). The inhibitory effect of glucose could be removed by either chemical modification of the glucose binding site with protection of the substrate binding site or conversion of glucose to fructose by added invertase. The lack of secretion of the enzymes could be overcome by using an appropriate secretion system. The thermo-instability of the enzymes during a reaction could be decreased by addition of Pro residues within the enzymes where the stability of a protein could be increased by changing specific residues by site-directed mutagenesis leading to a decrease in configurational entropy of unfolding (Matthews *et al.,* 1987). The susceptibility of the enzymes to proteolytic cleavages could be changed by the addition of covalently

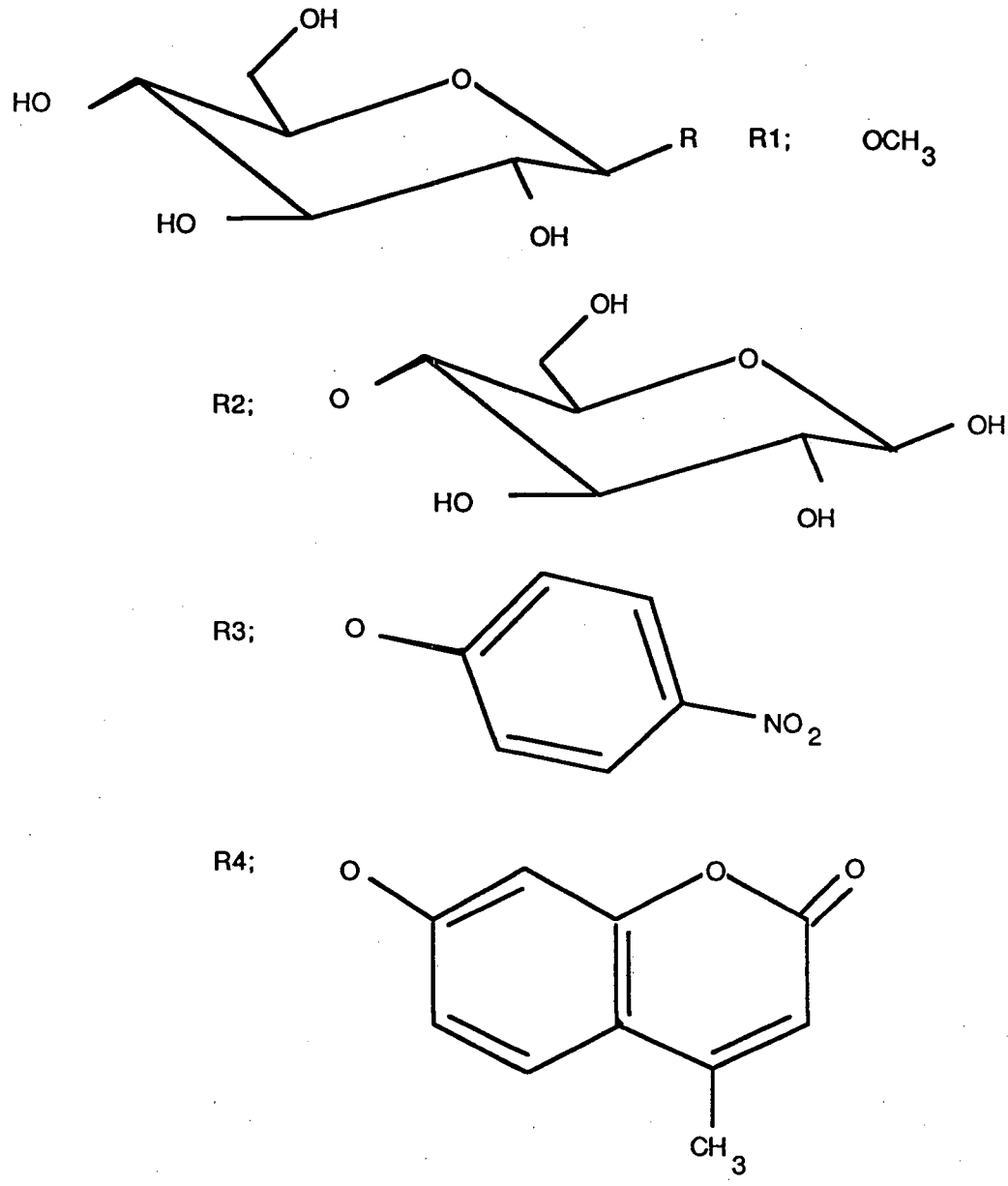
bound carbohydrates and protease inhibitors.

ß-Glucosidases are intracellular, cell-bound or extracellular in different organisms, or in the same organism depending on the type of inducer. ß-Glucosidases are usually secreted into the culture medium by cellulolytic fungi (Moldoveanu and Kluepfel, 1983). Bacterial ß-glucosidases are often intracellular or cell associated (Gong and Tsao, 1979). Nevertheless, extracellular bacterial ß-glucosidases have been reported in *Clostridium stercorarium* (Bronnenmeier and Standenbauer, 1988), *Bacteroides succinogenes* (Groleau and Forsberg, 1981), from an hybrid of *Cellulomonas species* and *Bacillus subtilis* (Gokhale *et al.,* 1984; Gokhale and Deobakar, 1989) and also from the bacterium *Acetivibrio cellulolyticus* (Saddler and Khan, 1981). The M_r for fungal ß-glucosidases, which may contain 0 to 90% (w/w) carbohydrates, ranges from 35000 for an extracellular enzyme of *T. reesei* up to 480000 for the G1 enzyme of *Monilia* sp. The ß-glucosidase of *A. fumigatus* is possibly a dimer with an M_r of 340000 and the *Botryodiplodia theobromae* enzyme with an M_r of about 350000 could be constituted of 8 subunits, each composed of 4 identical non-catalytic polypeptides (Coughlan, 1985). The M_r for bacterial ß-glucosidases may range from 50000 for the enzyme of *C. thermocellum* up to 122000 for the enzyme of *E. herbicola* (Woodward and Wiseman, 1982). The *Agrobacterium* sp. enzyme has an M_r of 100000 and could be a dimer (Day and Withers, 1986).

1.7.1. Substrate specificity and measurement of activity

The ß-glucosidases recognize a wide variety of aglycones provided the glucosyl moiety has a ß-configuration (Fig.4). The rate of hydrolysis depends on the nature of the aglycone moiety. The enzymes are sometimes specific for the hydroxyl group at position C_4. The determination of K_m and Vmax at different pH values during the hydrolysis of cellobiose by ß-glucosidases could indicate the involvement of a carboxylate group and a protonated nitrogen atom of an imidazole group in catalysis. The first step involves splitting of the

FIG. 4. STRUCTURE OF VARIOUS ß-GLUCOSIDES.
R1, 1-0-METHYL-ß-D-GLUCOSIDE; R2, CELLOBIOSE;
R3, 4-NITROPHENYL-ß-D-GLUCOSIDE (PNPG);
R4, 4-METHYLUMBELLIFERYL-ß-D-GLUCOSIDE (MUG);



aglycone moiety with simultaneous formation of a glucosyl-enzyme complex which then reacts with water, yielding glucose.

Reactions catalysed by ß-glucosidases may be assayed by measuring the glucose produced. The glucose is then measured spectrophotometrically using either a coupled glucose oxidase system or a coupled glucose-6-phosphate dehydrogenase and NADP system, or chemically. These assay methods are accurate and sensitive but laborious and time-consuming. An alternative method which allows the reaction to be followed continuously involves following the production of glucose at 340 nm by making use of ATP, hexokinase, glucose-6-phosphate dehydrogenase and NADP. The latter is converted into its reduced form NADPH which absorbs strongly at 340 nm (Hsuanya and Laidler, 1984). More directly, ß-glucosidase activity can be monitored in one step by measuring spectrophotometrically either the release of p-nitrophenol (pNP) from p-NP-ß-D-glucoside at 410 nm (Stoppock *et al.,* 1982) or of 4-methylumbelliferol (4-UM) from 4-MU-ß-D-glucoside at 340 nm (Sprey and Lambert, 1983). Cellobiose cleaving activity should not be determined by using aryl-ß-glucosides or by using assays not optimized for phosphorolytic cleavage, as previously mentioned by Schimz *et al.,* (1983).

More recently, an automated high-pressure liquid chromatography (HPLC) system allows the quasicontinuous analysis of all initial, intermediary and end products that appear during the hydrolysis of oligosaccharides (Schmid and Wandrey, 1989). Additional analysis can be performed by anion-exchange chromatography of the respective carbohydrate borate complexes with an automated sugar analyser (Schmid and Wandrey, 1989).

1.7.2. Classification of ß-glucosidases

In general, ß-glucosidases are separable into two groups: those in the first group have high affinities for aryl- and alkyl-ß-D-glucosides but low affinity for cellobiose and are known as aryl-ß-D-glucosidases; those in the second group are highly active on cellobiose and higher

cellodextrins and are known as true cellobiases. Nevertheless, the distinction between these two groups is not clear cut because many ß-glucosidases hydrolyze both aryl-ß-D-glucosides and cellobiose efficiently.

ß-Glucosidases hydrolyzing preferentially aryl- and alkyl-ß-glucosides have been reported in *Clostridium thermocellum* (Ait *et al.,* 1979), *Flavobacterium* (Sano *et al.,* 1975), *Kluyveromyces fragilis* (Raynal and Guerineau, 1984), *Aspergillus niger* (Pentilla *et al.,* 1984), *Cellulomonas uda* CB4 (Nakamura *et al.,* 1986), *Cellulomonas fimi* (Wakarchuk *et al.,* 1984) and *Escherichia coli* (Schaefler, 1967). Even though the *bgl* operon of *E. coli* encodes all functions necessary for the regulated uptake and utilization of aryl-ß-glucosides, the operon is cryptic and requires mutational activation.

ß-Glucosidases which preferentially hydrolyze cellobiose (or true cellobiases) have been reported in *Clostridium thermocellum* (Alexander, 1968), *Cellulomonas fimi* strain R2 (Sato and Takahashi, 1967), *Cellulomonas* sp., *Cellulomonas flavigena, Cellulomonas cartalyticum, Cellulomonas uda* (Schimz *et al.,* 1983), *Escherichia adecarboxylata* (Armentrout and Brown, 1981), *Agrobacterium faecalis* (Han and Srinivasan, 1969) and *E. coli* (Kricker and Hall, 1984). In *E. coli*, the *cel* operon for cellobiose catabolism is not expressed in wild-type strains but as for the *bgl* operon, it requires activation by mutational event to sustain bacterial growth on this substrate.

Several pathways of cellobiose utilization occur in nature. It can be degraded by hydrolysis, by oxidation to cellobionic acid followed by hydrolysis, by ATP-dependent phosphorylation of the C-6-OH at the non-reducing moiety of cellobiose followed by hydrolysis or by inorganic phosphate dependent phosphorolysis leading to equimolar concentration of glucose-1-phosphate and D-glucose (Schimz *et al.,* 1983). The last two mechanisms occur frequently in bacteria which do not secrete the enzymes into the culture medium.

In cellulolytic fungi, the occurrence of two kinds of ß-glucosidases,

Table II. K_ms for cellodextrin hydrolysis by various exo-ß-1,4-glucosidases from cellulolytic fungi

Enzyme	K_m (mM) C2	C3	C4	C5	C6	Ref.
Aspergillus niger						
Enzyme 1	1.1	0.49	0.13	0.26	0.36	1
Enzyme 2	1.64	0.54	0.087	0.31	0.49	
Enzyme 3	0.91	0.52	0.085	0.63	0.72	
A. phoenicis	0.75	NA	0.36	NA	NA	2
Pyricularia orysae						
GB-1	0.91	0.34	0.26	NA	NA	3
Sclerotium rolfsii						
BG-1	3.6	1.0	0.5	0.49	0.62	4
BG-2	3.07	1.23	0.85	0.40	0.37	
BG-3	5.84	1.98	0.83	0.76	0.37	
BG-4	4.15	0.70	0.50	0.55	0.67	
Trichoderma koningii	0.84	0.41	0.16	0.095	0.22	5
T. viride	1.5	NA	0.30	NA	NA	6
T. reesei	0.54	0.08	0.15	0.16	0.18	7

NA: not available

1. King and Smibert, 1963; 2. Sternberg *et al.,* 1977; 3. Hirayama *et al.,* 1978; 4. Shewale and Sadana, 1981; 5. Halliwell and Vincent, 1977; 6. Berghem and Pettersson, 1974; 7. Schmid and Wandrey, 1989;

one with aryl-ß-glucosidase and one with a broad activity towards different glucosides, is also a common feature. A decrease of K_m with increase in the chain length of the cello-oligosaccharides (at least up to cellopentaose) has been reported for several ß-glucosidases isolated from various fungi (Table II). The exoglucosidase activities of these enzymes were demonstrated by chromatographic identification and quantitative determination of the products of cellodextrin hydrolysis. The enzymes are classified as exo-ß-1,4-glucan glucohydrolase (or exoglucosidases) because of their sequential action on cellodextrins. Finally, it seems that ß-glucosidase in cellulolytic fungi is a multifunctional enzyme in the sense that it is not merely a cellobiase but rather an exo-ß-1,4-glucan glucohydrolase.

1.8. The preliminary characterization of *C. fimi* ß-glucosidases

The first report on *C. fimi* ß-glucosidases described the presence in the cell extracts of at least two enzymes, one an aryl-ß-D-glucosidase and one a ß-D-glucoside glucohydrolase. Both activities are inducible by growth on cellulosic substrates (Wakarchuk *et al.,* 1984). A ß-glucosidase-encoding clone reacting with antiserum to *C. fimi* culture supernatant, was isolated together with the other cellulase genes (Gilkes *et al.,* 1984) and partially characterized (Bates, 1987). The recombinant enzyme, Cbg, did not require phosphate for the hydrolysis of pNPG or cellobiose. Gene expression in *E. coli* requires an *E. coli* promoter (Bates, 1987). Zymograms using MUG as a substrate identify comigrating bands of activity in *C. fimi* culture supernatants and in extracts from *E. coli* cells expressing the gene. This suggested that the recombinant enzyme corresponds to a ß-glucosidase secreted by *C. fimi.* This is unusual because extracellular ß-glucosidase activity was not detected previously in *C. fimi* culture (Wakarchuk *et al.,* 1984). Furthermore, a series of 5'-end deletions of the cloned *C. fimi* DNA fragment gave a mutant called pUC13:62Δ31, with increased activity. Apparently, the increase in expression resulted from a more optimal spacing between the Plac promoter and the *cbg*A RBS (Bates, 1987). More recently, Kim and Pack have reported the molecular cloning in *E.*

coli of two *C. fimi* ß-glucosidase encoding genes (Kim and Pack, 1989). For the expression of ß-glucosidase activity in *E. coli,* both genes require the presence of an *E. coli* promoter. One enzyme hydrolyses both pNPG and cellobiose, the other only pNPG (Kim and Pack, 1989). The corresponding native enzymes were not identified in *C. fimi* ; the recombinant enzymes themselves must be characterized further.

1.9. The second objective of this thesis

The general goal of our group is to characterize every secreted component of the *C. fimi* cellulase complex involved in the hydrolysis of cellulose to glucose. So far, five genes encoding cellulolytic enzymes have been isolated and characterized both at the protein and DNA levels. In *C. fimi* very little is known about its ß-glucosidases. During the characterization of the recombinant CbgA and identification of the corresponding native enzyme, several other enzymes with ß-glucosidase activity were identified. Apparently, up to three different enzymes, based on their catalytic properties, were detected along with the CbgA enzyme which is the only one found secreted in the culture medium. The second part of this thesis describes the characterization of the recombinant enzyme CbgA both at the genetic and biochemical levels.

2. MATERIALS AND METHODS

2.1. Bacterial strains, plasmids, phage and growth conditions

Table III list the bacterial strains, plasmids and phage used. *C. fimi* strain ATCC 484 was grown at 30°C with agitation in low salt-LB medium (Bacto-tryptone, 10 g; Bacto-yeast extract, 5 g; NaCl, 0.5 g; pH 7.0) or in basal medium (Stewart and Leatherwood, 1976) supplemented with 1% (w/v) of an appropriate source of carbon (glucose, glycerol, cellobiose or CMC (Sigma; low viscosity)). *Brevibacterium lactofermentum* strain ATCC 21798 was grown at 30°C with agitation in TSB or MMYC media (Santamaria *et al.,* 1984). *E. coli* strains were grown at 37°C with agitation in LB, TYP or M9CA media (Maniatis *et al.,* 1982). Solid media contained 1.5% (w/v) agar (Difco). If the media contained CMC, the agar was reduced to 1.1% (w/v). Ap and Km sulfate (Sigma Co., St. Louis, MO) were used at a final concentration of 5 to 100 μg/mL, as required.

2.2. Enzymes and reagents

Restriction endonucleases, DNA and RNA modifying enzymes were purchased from various sources (Bethesda Research Lab., Pharmacia, New England Biolabs and Boehringer-Mannheim Inc.) and used according to the supplier. Chemicals for electrophoresis were supplied by Bio-Rad laboratories. Nitrocellulose membranes BA85 were from Schleicher & Schuell Inc. Radionucleotides were from New England Nuclear (Dupont-NEN) and Amersham Inc. Yeast tRNA and redistilled phenol were from Bethesda Research Laboratories. Ribonuclease A, pNPG, pNPC, MUG and MUC were from Sigma Chemical Co.

2.3. Transformation of bacteria by plasmid DNA

E. coli cells were transformed by a modification of the calcium chloride procedure (Mandel and Higa, 1970). *B. lactofermentum* cells were transformed by a protoplast fusion technique according to Santamaria *et al.,* (1984) with some minor modifications. Briefly,

Table III. Bacterial strains, plasmids and phage

Bacterial strains	Relevant characteristics	references
E. coli JM83	ara Δ(lac-pro AB) rpsL Ø80 lacZ ΔM15	1
E. coli JM101	supE thi Δ(lac-proAB)[F' tra Δ36 proAB lacI-Z ΔM15	1
E. coli RZ1032	Hfr KL16 PO/45[lysA(61-62)] dut1 ung1 thi1 relA1 zbd-279::tn10 supE44	2
B. lactofermentum	aecR	ATCC21798
C. fimi	cellulose hydrolysis	ATCC484

Plasmids	Relevant characteristics	
pUC8 to 19	Ap^R lac Z'	3
pUC8:*cen*A	Ap^R lac Z' CenA	4
pUC12:A25	Ap^R lac Z' Cex	5
pUC4K	Ap^R lac Z' Km^R	6
pBR322	Ap^R Tc^R	7
pBL1	cryptic	8
pBK10	Ap^R Km^R	4
pFWP10	Ap^R Km^R Cex	4
pFWP30	Ap^R Km^R CenA	4
pFWP1030	Ap^R Km^R Cex CenA	4
pEC62	Ap^R CbgA	9
pTZ19R	Ap^R lac Z' orif1	10

Phage	Relevant characteristics	
M13K07	Km^R	10

1. Yannish-Perron *et al.*, 1985; 2. Kunkel *et al.*, 1987: 3. Messing, 1983; 4. Paradis *et al.*, 1987; 5. O'Neil, 1986; 6. Oka *et al.*, 1981; 7. Boyer *et al.*, 1977; 8. Santamaria *et al.*, 1984; 9. Bates, 1987; 10. Viera, 1985.

protoplast formation was monitored by microscopy and by increase in osmotic sensitivity after hypotonic dilution. Protoplasts were fused and 50 µL aliquots of the fusion mixture were plated on RS regeneration medium. After 24 h regeneration at 30°C, the agar was overlayed with TSB medium precooled to 37°C and containing 0.4% low-melting agarose and 50 µg Kanamycin sulfate/mL. Putative transformants were transferred to TSA media after five to ten days of incubation at 30°C and tested for cellulolytic activity. Clones expressing *cex* were detected by fluorescence on agar containing MUC (final concentration 500 µM). Those expressing *cen*A were detected on agar containing 1% CMC, using Congo red to visualize the zones of hydrolysis (Teather and Wood, 1982).

2.4. Enzyme assays and protein determination

Cell extracts were prepared by concentrating the cells 10 to 50-fold by centrifugation and resuspending the cells in 50 mM K-phosphate buffer (pH 7.0). The suspensions were either sonicated using a Bronson sonifier with a microprobe set at an intensity of 2 for 3 bursts of 30 sec or ruptured with a French press for larger volumes. The exoglucanase activity of cell extracts was determined with pNPC as a substrate (Gilkes *et al.*, 1984a). The endoglucanase activity of cell extracts was determined by DNS assay (Miller, 1959) using CM-Cellulose as substrate (Gilkes *et al.*, 1984a). ß-Glucosidase activity of cell extracts was determined spectrophotometrically by following the release of p-nitrophenol from pNPG at 410 nm (Stoppok *et al.*, 1982; Berghem and Pettersson, 1973). Glucose-6-phosphate dehydrogenase activity, a cytoplasmic enzyme marker, was measured by following the reduction of NADP at 30°C and 340 nm according to Worthington (1977). ß-Lactamase activity, a periplasmic space marker, was measured by following the hydrolysis of the chromogenic substrate nitrocefine at 37°C and 486 nm according to O'Callaghan *et al.* (1972). NADH oxidase activity, an inner-membrane marker, was measured by the method of Osborn *et al.* (1972). 2-Keto-3-deoxyoctonate content, a sugar

exclusively found in the LPS of outer membranes, was estimated by the modified method of Osborn *et al.* (Darveau and Hancock, 1983). Protein concentration was determined according to Bradford (1976) using BSA as standard. One unit (U) of enzyme is defined as the amount of enzyme required to release 1 µmol of product per min.

2.5. Fractionation of bacterial cells

E. coli cells were fractionated into periplasmic, cytoplasmic and total membrane fractions, and *C. fimi* cultures were fractionated into cytoplasmic, total membrane and culture supernatant fractions according to Gerhardt *et al.,* (1981) with minor modifications.

2.5.1. Fractionation of *E. coli* cells

The location of recombinant CbgA in *E. coli* JM83 carrying pUC13:62Δ31 was established by fractionating the cells and measuring ß-glucosidase activity. Cells were grown in 200 mL of Luria broth at 37°C with agitation. Exponential and stationary phase cultures ($O.D._{600nm}$ of 0.8 and 2.3, respectively) were chilled on ice for 5 min and harvested at 8000 X *g* for 15 min at 4°C in a JA 20 rotor (Beckman). Periplasmic fractions were first prepared by a standard osmotic shock procedure (Nossal and Heppel, 1966). The shocked cells were recovered by centrifugation and re-suspended in 1/20 vol ice-cold 10 mM HEPES buffer, pH 7.4. The cells were ruptured with a French press. Cell debris was removed by centrifugation at 8000 X *g* for 10 min at 4°C and membranes were sedimented by ultracentrifugation at 110 000 X *g* for 2 hrs at 7°C. The supernatants were decanted and kept as cytoplasmic fractions. The membranes were resuspended in 10 mM HEPES buffer (pH 7.4) containing 5 mM-EDTA and further analysed by isopycnic sucrose density gradient centrifugation (Osborn and Munson, 1974).

2.5.2. Fractionation of *C. fimi* cells

C. fimi cells were grown to late log phase ($O.D._{600nm}$ of 1.7) in 1 L

of 1% CM-cellulose containing basal medium, chilled on ice for 5 min prior to collection at 8000 X *g* for 15 min at 4°C in a JA20 rotor (Beckman). Proteins in the culture medium were concentrated by ultrafiltration through a PM10 membrane (Amicon) at 4°C for 12 to 16 hrs. Protein aggregates containing a substantial amount of the total pNPGase activity detected in the culture medium were obtained from the concentrated material by ultracentrifugation. Cytoplasmic proteins and crude envelope fractions were obtained as described above for *E. coli.*

2.6. The purification of ß-glucosidases

2.6.1. The purification of recombinant CbgA

E. coli/pUC13:62Δ31 was grown for 8 hrs at 30°C in 40 L of Luria broth in a LH Scientific 5000 series Fermenter (air pressure at 1 bar, air flow at 20 L/min, agitation at 200 rpm). Cells were harvested at an O.D. $_{600nm}$ of 0.7 using an air-driven Sharples continuous flow centrifuge (Laboratory Presurtite centrifuge), set at 50 000 RPM and 20 PSI, at a flow rate of 700-850 mL/min. Turbidity of the culture supernatant was observed periodically to follow the efficiency of centrifugation. The cell paste was resuspended in 400 mL of ice cold buffer A (50 mM Tris-Cl (pH7.0); 5 mM EDTA; 1 mM PMSF; 10^{-7} M Pepstatin A). Cytoplasmic proteins were prepared by rupturing 100 mL of the cell suspension with a French press. Cell debris was removed by low speed centrifugation and the supernatant was further clarified by ultracentrifugation at 110 000 X *g* for 80 min at 7°C. To the clarified extract was added 1.5% (w/v) streptomycin sulfate and the precipitated nucleic acids removed by centrifugation. Two ammonium sulfate cuts (the first cut 28%, the second cut 45%) were performed on the supernatant fraction, where precipitates were removed by centrifugation. The 45% cut precipitate was dissolved in buffer B and dialysed overnight at 4°C against buffer B (Tris-Cl 10 mM-EDTA 1 mM (pH7.0)). The dialysate was fractionated at 22° by ion-exchange chromatography on DEAE-Sephadex A-50-120. Bound proteins were eluted at 44 mL/hr by stepwise increase in the NaCl salt concentration from 0 to 1 M. Fractions of 2.5 mL each were collected and

the active fractions from two peaks (I and II) containing ß-glucosidase activity were pooled separately (Pool I and Pool II). Proteins in each pool were concentrated by ultrafiltration through a PM10 membrane (Amicon), salt concentration was decreased 10 fold by dilution with buffer B and samples were further concentrated to a final volume of 5 mL by ultrafiltration. The final purification was achieved by anion-exchange chromatography of 1.0 mL of desalted Pool I on Mono Q resin (Pharmacia). The speed of elution was set at 1 mL/min and bound proteins were eluted using a 0 to 1 M NaCl salt gradient.

2.6.2. The purification of *C. fimi* ß-glucosidases.

10 L of 1%-CM-cellulose basal medium were inoculated with a preculture containing *C. fimi* cells and the culture grown at 30°C to an O.D. $_{600\ nm}$ of 1.3. The culture was chilled on ice, then centrifuged at 4400 X g for 10 min at 4°C. The cells were suspended in 100 mL of ice cold buffer A. The suspension was passed 10X through a French press. The resulting extract was fractionated as described above. Active fractions with pNPGase/cellobiase, pNPGase and mainly cellobiase activities were grouped as Pool I, Pool II and Pool III, respectively. The ß-glucosidases present in Pool II and III were further purified by anion-exchange chromatography on Mono Q resin as described above.

2.7. Characterization of purified recombinant CbgA

2.7.1. The effects of pH and temperature on CbgA activity

Reaction mixtures contained either 1.4 µg of purified CbgA mL^{-1} or 140 µg of a crude cell extract mL^{-1}, 100 mM phosphate of the required pH, and 5 mM pNPG in a total volume of 1 mL. After incubation at the appropriate temperature for 60 min, the reactions were stopped by the addition of 1/2 vol 1 M Na_2CO_3 and the absorbance measured at 410 nm.

2.7.2. Determination of the kinetic parameters for CbgA

The kinetic parameters of the purified recombinant enzyme CbgA

were determined from Lineweaver-Burke plots for each ß-glucosidic substrates used in this study (see appendix). All reactions were done in 1 mL final volume in a thermostatted cuvette chamber at 37°C and followed using a spectrophotometer (Varian DMS 100 UV Visible) connected to a Screen-Plotter (SANYO) and laboratory strip chart recorder. Each assay contained 1 µg of the purified recombinant enzyme from *E. coli* JM83/pUC13:62Δ31 strain, 1 X reaction buffer (10 X stock ; 200 mM KPO_4 (pH 7.0)-50 µg BSA/mL-1 mM DTT) and various concentrations of substrates (0.04-8.0 mM pNPG, 0.025-5.0 mM pNPC, 0.5-50 mM cellobiose, 0.25-10 mM cellotriose, 0.25-10 mM cellotetraose, 0.04-5.0 mM cellopentaose).

The hydrolysis of p-nitrophenyl derivatives (pNPG and pNPC) was followed at 410 nm and the molar extinction coefficient for p-nitrophenol at pH 7.0 was estimated from standard solution (Sigma). The hydrolysis of cellodextrins (cellobiose from Sigma; cellotriose, cellotetraose and cellopentaose kindly provided by N.R. Gilkes from U.B.C.) to glucose was followed by making use of a glucose coupled assay (Sigma kit 15-10) containing hexokinase and glucose-6-phosphate dehydrogenase and following the reduction of NADP at 340 nm (Hsuanyu and Laidler, 1984). The hydrolysis of all ß-glucosides was initiated by adding the recombinant CbgA to the prewarmed reaction mixture. Hydrolysis was not detected in the absence of enzyme with all substrates tested. Glucose inhibition (0-11 mM) was measured on pNPG (0.1, 0.3 mM and 1.0 mM) as substrate (Cornish-Bowden, 1979).

2.8. Electrophoretic analysis of proteins

Proteins were analysed by electrophoresis in 7.5 or 10% polyacrylamide gels in the presence of 0.1% sodium dodecyl sulfate (SDS) (Laemmli, 1970) and gels were stained with 0.03% Coomassie brillant blue (R-250) dissolved in 10% acetic acid and 50% methanol. Excess stain was removed by soaking the gel in 10% acetic acid-50% methanol until the background was negligible. When necessary, protein samples were denatured by boiling for 2 min. Calibration standards

used were myosin (205 kDa); ß-galactosidase (116 kDa); phosphorylase B (97.4 kDa); bovine serum albumin (66 kDa); catalase (57 kDa); glutamate dehydrogenase (53 kDa); ovalbumin (45 kDa); alcohol dehydrogenase (43 kDa); glyceraldehyde-3-phosphate dehydrogenase (36 kDa) and carbonic anhydrase (29 kDa).

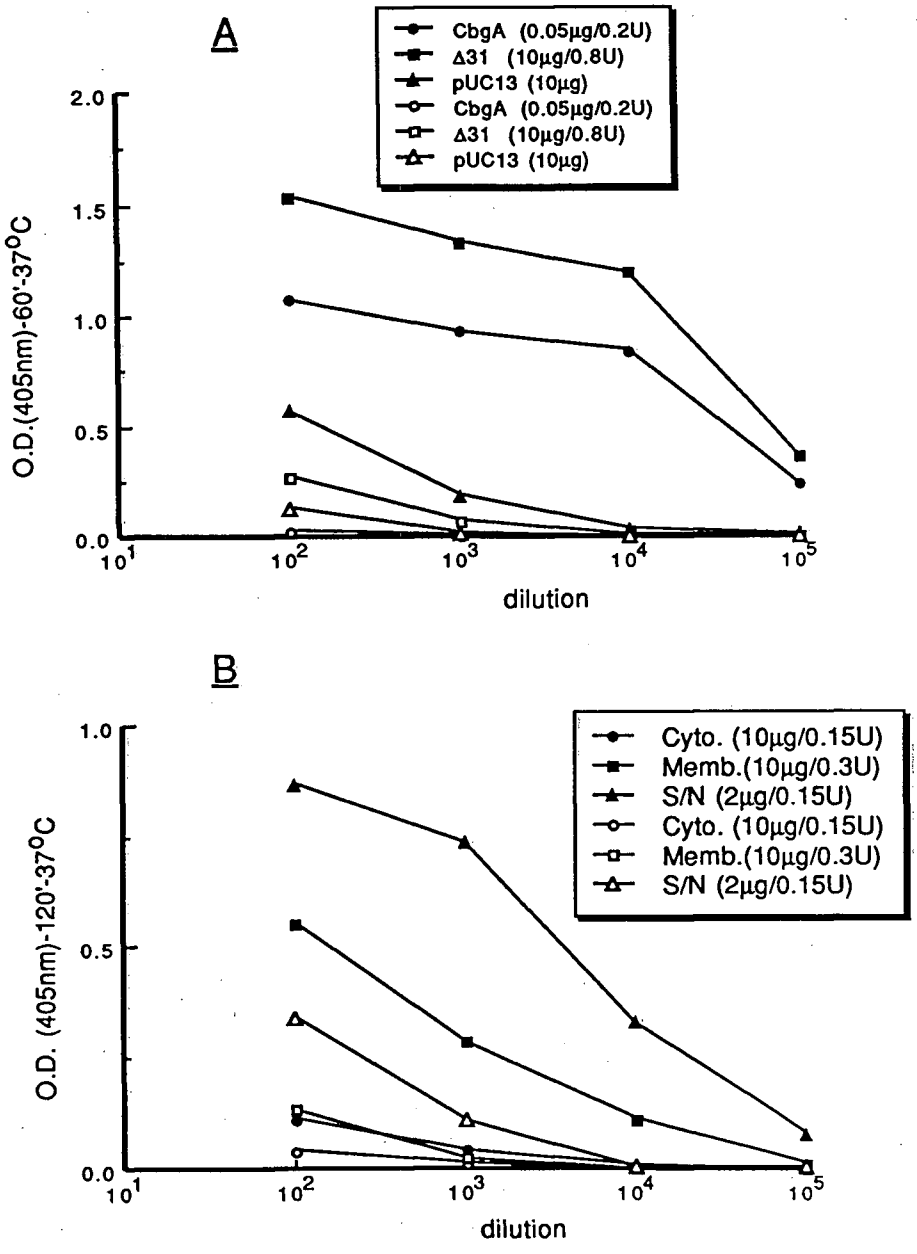
For zymograms, unheated samples were applied to 7% polyacrylamide gels in standard sample loading buffer containing SDS and ß-mercaptoethanol. After electrophoresis, bands of ß-glucosidase activity were visualized by soaking the gel for 30 min at 30°C in a solution containing 1 mM 4-methylumbelliferyl- ß-D-glucoside (MUG)-0.1 M potassium phosphate (pH 6.8)-1% Triton X-100 solution (Lacks and Springhorn, 1980). Fluorescent bands were visualized under shortwave UV light and the gel was photographed using type 57 film (Kodak).

2.9. Production of antiserum and Western blot analysis

A total of 150 µg of purified recombinant CbgA was mixed with Freund's complete and incomplete adjuvant (1:1). At intervals of two weeks, 1/3 of the mixture was injected intramuscularly into a New Zealand white rabbit. One week after the third injection, the rabbit was bled, the serum collected and tested against *E. coli* and *C. fimi* proteins in conjunction with the rabbit preimmune serum using an enzyme-linked-immunoadsorbant assay (ELISA, figure 5A and 5B) as previously described (Voller *et al.*, 1976). The antiserum was made 0.02% in sodium azide and aliquots were kept at 4°C and -20°C.

Polypeptides from polyacrylamide gels were blotted onto nitrocellulose membranes by electrotransfer and detected by immunoblotting as previously described (Towbin *et al.*, 1979) using the alkaline phosphatase/5-bromo- 4-chloro-3-indolyl-phosphate (X-phosphate) detection system (Blake *et al.*, 1984). Prestained calibration standards used were myosin (rabbit skeletal muscle, 210 kDa); phosphorylase B (rabbit muscle, 110 kDa); bovine serum albumin (75 kDa); ovalbumin (hen egg white, 45 kDa); soybean trypsin inhibitor (29 kDa); ß-lactoglobulin (18.4) and lysozyme (hen egg white, 14.3 kDa) (BRL).

Fig. 5. Titration of antiserum by ELISA on (A) *E. coli* and (B) *C. fimi* proteins. Test antigens were coupled in triplicate to microtiter plates using the following amounts of protein per well: (A) 50 ng of purified CbgA, 10 µg of *E. coli*/pUC13:62Δ31 total cell extract and 10 µg of *E. coli*/pUC13 total cell extract; (B) 10 µg of *C. fimi* cytoplasmic fraction, 10 µg of *C. fimi* membrane fraction and 2 µg of *C. fimi* culture supernatant protein aggregate fraction. The pre-immune (open symbols) and immune (black symbols) rabbit sera were tested. The absorbances at 405 nm reflect the mean values of the activity of alkaline phosphatase conjugated to goat IgG antibodies against rabbit IgG bound to the test antigens.



2.10. Determination of the NH_2-terminal amino acid sequences and amino acid compositions of recombinant CbgA polypeptides

The NH_2-terminal amino acid sequences were determined by automated Edman degradation using an Applied Biosystems model 470A gas sequenator utilizing the resident sequencing program. These analyses were kindly performed by Sandy Kielland at the University of Victoria sequencing facility. Amino acid composition was determined at the same facility.

2.11. DNA and RNA methodology

2.11.1. Isolation of plasmid DNA and restriction digest analysis

Plasmid DNA was isolated from *E. coli* strains by a modification of the alkaline-lysis procedure (Birnboim and Doly, 1979). When required, the DNA was purified further by centrifugation to equilibrium in CsCl density gradients containing ethidium bromide (Maniatis *et al.,* 1982). Plasmid DNA was isolated from *B. lactofermentum* by a modification of the procedure of Santamaria *et al.,* (1984). *B. lactofermentum* cells were grown overnight in TSB medium. The cells were collected from a 1.5 mL sample by centrifugation at 12,000 X *g* and resuspended in 200 µL of lysis buffer containing lysozyme at 3 mg/mL final concentration. The mixture was incubated at 37°C for 4 hrs with occasional mixing. Protoplasts were lysed by adding 400 µL of a 1% SDS-0.2 N NaOH solution prewarmed at 65°C and mixing immediately by inversion. After 5 min incubation at room temperature, 300 µL of ice cold 3 M sodium acetate, pH 5.2, were added and mixed in for 10 sec with a vortex mixer. Tubes were kept on ice for 15 min, then the turbid lysates were spun at 12,000 X *g* for 10 min at 4°C. The decanted supernatants were extracted once with 1 vol of phenol and once with 1 vol of chloroform-isoamyl alcohol (24:1). The nucleic acids were precipitated by adding 1 vol of cold isopropanol and the mixtures were stored at -80°C for 30 min. The tubes were centrifuged in a microfuge (Fisher) for 10 min at 4°C. The pellets were rinsed with 70% ethanol, dried and then incubated at

37°C for 30 min in 25 µL of TE buffer containing RNase A at 50µg/mL final concentration.

Single stranded pTZ DNA was isolated by a modification of the procedure of Kristensen *et al.* (1987). A 2.0 mL overnight culture of M13K07 infected JM101 was centrifuged. 1.5 mL of the supernatant were mixed with 200 µL of 25% PEG-3.5 M NaCl and incubated on ice for 30 min. After centrifugation at 12,000 X *g* for 15 min, the phage pellet was treated with 1.0 mL of 4 M $NaClO_4$-10 mM Tris pH 7.5-1 mM EDTA. The DNA released was adsorbed to a glass fiber filter (GF/C from Whatman International Ltd) of 7 mm diameter. The filters were washed 3 X with 500 µL of 70% ethanol. The DNA was eluted with water.

DNA restriction fragments were resolved by agarose gel electrophoresis (0.6 to 1.5% (w/v) agarose). Restriction fragments from lambda DNA digested with *Hind*III or *Hind*III-*Eco*RI served as size markers.

2.11.2. Isolation of RNA and RNase-free work

All chemicals and reagents used during RNA work were purchased only for this purpose and kept separate from regular laboratory supplies to avoid RNase contamination. The glassware used was either baked at 300°C for 3 hrs or bought as disposable labware. When required, solutions were treated with 0.2% (v/v) diethylpyrocarbonate (DEPC) as described elsewhere (Ehrenberg *et al.*, 1976; Maniatis *et al.*, 1982). Plastic pipette tips and microfuge tubes were sterilized by autoclaving only.

RNA was isolated from all bacterial strains by a modification of published procedures (Greenberg *et al.*, 1987a). Briefly, 100 mL cultures were rapidly cooled on ice and transferred to pre-chilled centrifuge tubes (-20°C). Cells were harvested by centrifugation for 5 min at 6000 X *g* and resuspended in 1/10 volume of 50 mM Tris-HCl (pH 6.8)-2 mM EDTA-1.0% SDS. The suspension was transferred to a clean tube and boiled for 2 min in a water bath. The lysate was cooled on ice

for 5 min and 1/2 vol of ice cold NaCl was added and mixed in using a Vortex mixer. The sample was kept on ice for 5 min, then centrifuged for 10 min at 10,000 X *g*. The clear supernatant fluid was decanted into a 30 mL Corex glass tube. The nucleic acids were precipitated by addition of 2.5 volumes of 95% ethanol, stored at -20°C for 12 to 16 hrs and recovered by centrifugation for 20 min at 10,000 X *g*. The pellet was then washed with 70% ethanol at -20°C and redissolved in 2.0 mL of 10 mM Tris-HCl (pH 7.5)-40 mM NaCl-5 mM $MgCl_2$. DNA was removed by treatment with 5 U of RQ1 DNaseI (Promega) for 15 min at 37°C. EDTA was added to 5 mM and the mixture was extracted twice with phenol-chloroform (1:1) and once with chloroform. RNA was recovered by precipitation with 2.5 volumes of 95% ethanol followed by centrifugation for 10 min at 10,000 X *g*. The RNA pellet was washed with 70% ethanol and finally dissolved in RNA storage buffer (20 mM $NaPO_4$ (pH 6.5)-1 mM EDTA). Samples were divided into aliquots and stored at -20°C.

The RNA preparations were analysed by agarose gel electrophoresis after staining with ethidium bromide. The size markers used were ssDNA *Hae*III digested M13 mp11 where the 525-base fragment arised from partial digestion (Greenberg *et al.*, 1987a,b) and a 0.24-9.5 Kb RNA ladder purchased from BRL. RNA concentrations were determined by A_{260} nm using a value of 37 µg per $O.D._{260}$.

2.11.3. Northern blot analysis of mRNA

For Northern blot analysis of specific transcripts, up to 30 µg of bacterial RNA were precipitated with ethanol, redissolved in 10 µL of 20 mM MOPS-1 mM EDTA-5 mM sodium acetate (running buffer [pH 7.0]) with 50% formamide and 2.2 M formaldehyde, heated for 5 min at 68°C, and cooled briefly on ice. Loading dye was added (to give 3%[w/v] Ficol (Pharmacia) and 0.02% (w/v) bromophenol blue and xylene cyanol). The samples were loaded onto a 1.0% agarose-6.6% formaldehyde gel alongside ^{32}P-labelled molecular weight markers. Electrophoresis was at 40 mA with recirculation of running buffer. RNA was blotted onto

BioTrans membranes (Pall, Inc.) in 20X SSC (3 M NaCl-3 M sodium citrate) by electrotransfer for 12 to 16 hrs (Southern, 1975). The membrane was allowed to dry then baked at 80°C for 2 hrs. Prehybridization and hybridization were performed according to the membrane supplier. The ^{32}P-labelled probes (*cex* specific: 5' GTGGCCGGGTGCGGGCGTGGTCCTAGGCAT 3'; and *cen*A specific: 5' CAGCGCTGCGGCGGTTCTGCGGGTGGACAT 3') were synthesized chemically by Tom Atkinson essentially as described (Atkinson and Smith, 1984) and labelled by N. Greenberg at U.B.C. The membranes were first washed at room temperature with 6X SSC buffer followed by gradual increase of temperature and salt stringency. Filters were wrapped in Saran-Wrap^{TM} and exposed to X-ray film (Kodak) with intensifying screens at -70°C for 16 hrs to 3 days.

2.11.4. Preparation of ^{32}P labelled DNA hybridization probes

Plasmid DNA was digested conveniently by restriction enzymes under the appropriate conditions suggested by the manufacturer. The mixture was extracted with phenol-chloroform (1:1) and the DNA was precipitated with 95% ethanol. The 5' phosphate was removed from the DNA ends by treatment with calf intestinal alkaline phosphatase (CIAP) and the ends were labelled with [gamma-^{32}P]-ATP (3,000-7000 Ci/mmol) and T4 polynucleotide kinase (PNK) as described previously (Maniatis *et al.,* 1982). Unincorporated label was removed using a Sephadex G-50 column (Maniatis *et al.,* 1982) and the reaction was monitored by liquid scintillation spectrometry in an ISOCAP-300 (Nuclear Chicago). If required, restriction digests were performed on the labelled DNA to liberate fragments uniquely labeled at one end. The probes were purified by electrophoresis in 8% polyacrylamide gels from which they were electroeluted and further precipitated by addition of 2 vol of 95% ethanol and 0.1 vol of 3M sodium acetate. Pellets were rinsed with 70% ethanol, dried in a spin vac and redissolved in water. The efficiency of the kinasing reaction was monitored by liquid scintillation counting.

2.11.5. Hybrid protection analysis

The 5' ends of mRNA were identified with labeled DNA probes essentially as described previously (Favoloro *et al.,* 1980). Up to 30 µg of RNA and end-labelled DNA probe (0.01 to 0.03 pmol) were precipitated together by ethanol, redissolved in 30 µL of hybridization buffer (0.4 M NaCl-0.04 M sodium phosphate [pH 6.5]-0.4 mM EDTA-80% formamide), heated for 15 min at 85°C and quickly transferred to 60°C water bath for 3 hrs. Samples were diluted with 300 µL ice-cold S1 buffer (30 mM sodium acetate [pH 4.5]-28 mM NaCl-4.5 mM $ZnSO_4$) containing about 1000 U of S1 nuclease, and incubated for 30 min at 37°C. The reactions were terminated by addition of 75 µL stop buffer (2.5 mM ammonium acetate-50 mM EDTA) and 20 µg yeast tRNA as carrier. Nucleic acids were precipitated with 400 µL isopropanol and recovered by centrifugation. Pellets were dissolved in sequencing buffer (90% formamide, 0.02% [w/v] bromophenol blue and xylene cyanol) (Maniatis *et al.,* 1982) and heated at 90°C for 2 min. Nucleic acid fragments were resolved by PAGE alongside a DNA sequencing ladder and the dried gels were exposed to X-ray film (Eastman Kodak) at -70°C with intensifying screens.

2.11.6. DNA sequencing

DNA was sequenced by modifications of the standard chemical (Maxam and Gilbert, 1980) and enzymatic chain termination (Sanger *et al.,* 1977) methods. For chemical sequencing, DNA fragments were 5'end-labelled using PNK, digested with restriction enzymes and purified by PAGE as described above. The chemical modification and cleavage reactions were performed as described (Maniatis *et al.,* 1982). For enzymatic sequencing, several methods were used. For sequencing of double stranded template, DNA was isolated according to Kraft *et al.,* (1988) and the reactions were performed using the Pharmacia Sequencing Kit, T7 DNA polymerase (Pharmacia) and [alpha-^{35}S]dATP

Table IV. Sequencing primers

M13/pUC	17mer-reverse	5'-dCAGGAAACAGCTATGAC-3'
M13/pUC	17mer-universal	5'-dGTAAAACGACGGCCAGT-3'
Micro 1	17mer-94 nt	5'-PO_4-dACAGGCACCGACCAGGC-3'
Micro 2	17mer-380 nt	5'-PO_4-dGCCCAGCGGGTCGCCGG-3'

(500 Ci/mmol) with various primers (Table IV). DNA compressions were resolved by making use of deoxy-7-deaza-guanosine triphosphate (Mizusawa *et al.,* 1986) in place of dGTP and by using Taq polymerase (Pharmacia).

2.11.7. Oligonucleotide directed mutagenesis

The creation of a single base pair frameshift at position 34 of the *cbg*A putative GTG start site was obtained by primer elongation (Atkinson and Smith, 1984). Cells carrying pTZ19R-2 were grown and single stranded DNA was isolated as described previously (Kunkel *et al.,* 1987). In the hybridization mixture, 20 pmoles of single stranded pTZ19R-2 DNA and 200 pmoles of primer FP21 (5' PO_4-GGAGTGA CCG**C**TGCTGCGTGC-OH-3') were mixed in 200 µL of *Hinc*II buffer (BRL, insertion of a **C** destroys the CbgA putative GTG start site). The mixture was heated to 90°C for 5 min and slowly cooled to 25°C. The hybridization mix was then kept at 4°C for 12-16 hrs. Primer extension was initiated by the addition of 15 U of Klenow fragment of *E. coli* DNA polymerase I (BRL) plus 500 µM-dNTPs, 1 mM-DTT, 1 mM ATP and 15 U of T4 DNA ligase (NEB). After 20 min incubation at 37°C, an additional 15 U of Klenow enzyme were added and incubation continued for another 20 min at 37°C. The sample was extracted once with phenol-chloroform (1:1) and the DNA was recovered by ethanol precipitation followed by resuspension in 50 µL TE buffer. An aliquot of 5 µL was used to transform *E. coli* JM101 and some single stranded DNA was isolated from transformants as described (see above). DNA sequencing was performed to identify clones with an extra C at position 34 giving rise to pTZ19R-2FP21. By subcloning the wild-type *Stu*I-*Eco*RI fragment in pTZ19R-2FP21, pTZ19R-7 was finally isolated and screened for the extra C base by sequencing and further assayed for ß-glucosidase activity and Western blot antigenic peptides.

3. RESULTS AND DISCUSSIONS

PART 1. The expression of *Cellulomonas fimi* cellulase genes in *Brevibacterium lactofermentum*.

3.1. Construction of the shuttle vector and subcloning of the cellulase genes

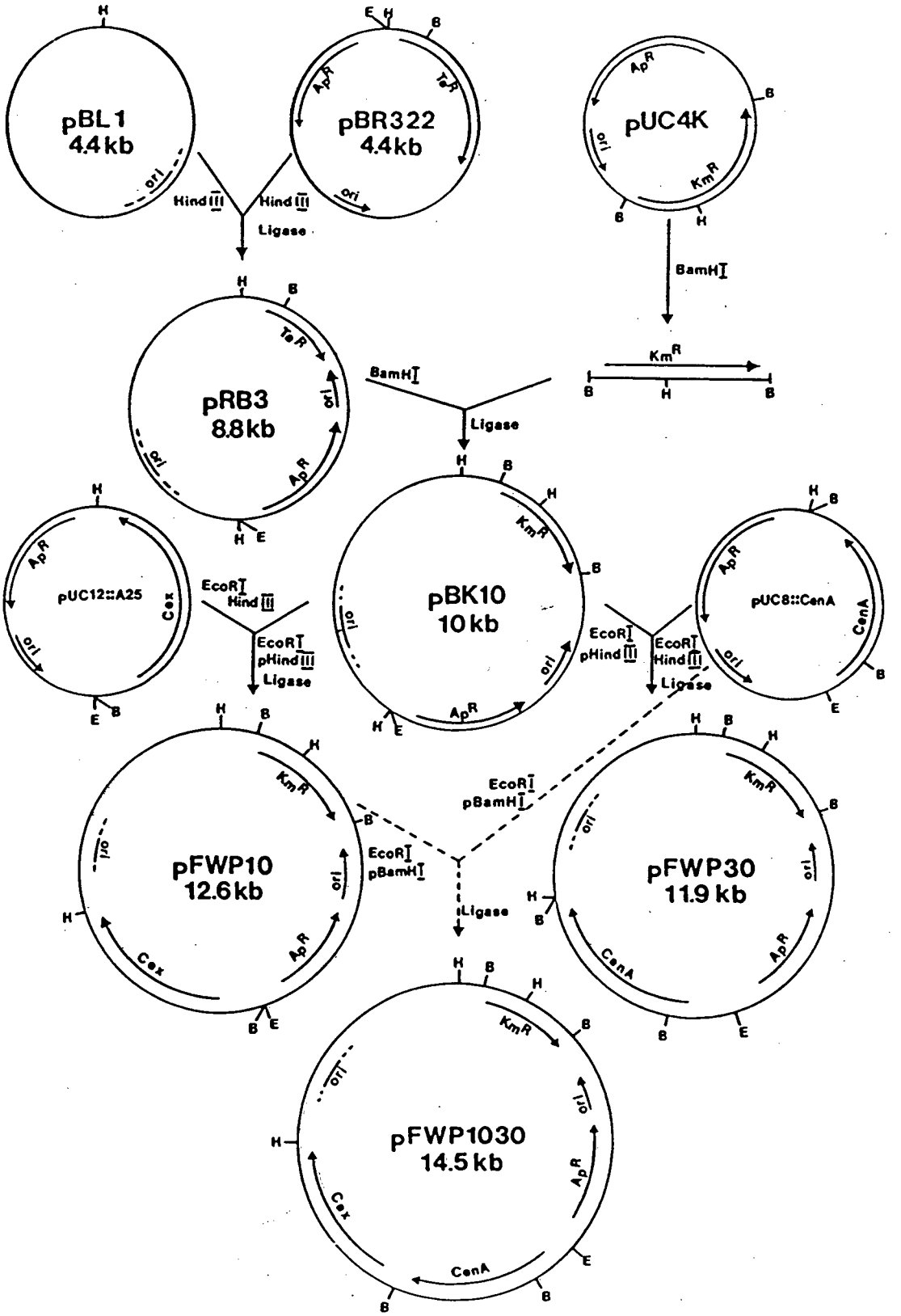
The cryptic plasmid pBL1 was iolated from *B. lactofermentum*; it was joined to pBR322 by their unique *Hind*III sites to give pRB3. The 1.3 Kb kanamycin resistance cartridge of pUC4K (Oka *et al.*, 1981), was inserted into the unique *Bam*HI site of pRB3 to give pBK10 (Fig. 6). The MIC values for kanamycin sulfate in wild type bacterial strains were 0.8 µg/mL for *B. lactofermentum*, 12.5 µg/mL for *E. coli* and 350 µg/mL for *C. fimi*. The values for the strains transformed with pBK10 were 800 µg/mL for *B. lactofermentum* and over 1 mg/mL for *E. coli*.

Three derivatives of pBK10 were constructed (Fig. 6). pFWP10 carries the entire coding sequence of *cex* together with its putative regulatory region (Greenberg *et al.*, 1987a). pFWP30 carries the entire coding sequence of *cen*A together with its putative regulatory region (Greenberg *et al.*, 1987a). pFWP1030 carries both of these genes, separated by more than 1 Kb of *C. fimi* DNA. From previous studies (Stuber and Bujard, 1981), there does not appear to be a putative promoter in the pBR322 sequence upstream of these genes that could be held responsible for the cellulase genes expression.

3.2. Transformation

Transformation efficiency by heterologous plasmid DNA was lower in *B. lactofermentum* (<1 transformant per µg of DNA) than previously reported (Santamaria *et al.*, 1984). This may have been due to the presence of DNA-degrading enzymes which were detected by DNase test agar (Difco) (not shown). The transformation efficiency may also be reduced by restriction because the efficiency of transformation was increased 100-fold when the plasmid DNA was isolated from *B. lactofermentum* rather than from *E. coli* (data not given; see also Santamaria *et al.*, 1985). pBK10 and its derivatives were stable in both

Fig. 6. Construction of the shuttle vector and its derivatives. Restriction enzymes are: B, *Bam*HI; E, *Eco*RI and H, *Hin*dIII. Partial digest are designated by prefix p (e.g., p*Hin*dIII, etc.). Cryptic plasmid pBL1 was isolated from *B. lactofermentum* (ATCC21798). Plasmid pBL1 and pBR322 were digested at their unique *Hin*dIII sites and joined together to form pRB3. pRB3 was digested with *Bam*HI, and the Km^R cartridge of pUC4K was inserted into it to form the shuttle vector pBK10. This was digested with *Eco*RI and partially digested with *Hin*dIII to allow forced cloning of two individual *Eco*RI-*Hin*dIII fragments. The fragment coding for Cex gave rise to pFWP10, that coding for CenA gave rise to pFWP30. Finally, pFWP10 was digested with *Eco*RI and partially digested with *Bam*HI to allow forced cloning of an *Eco*RI-*Bam*HI fragment encoding CenA to give pFWP1030.


pBL1
4.4kb
ori
Hind III
Hind III
Ligase
pBR322
4.4kb
ApR
TcR
ori
pUC4K
ApR
ori
KmR
BamHI
KmR
pRB3
8.8kb
TcR
ori
ori
ApR
BamHI
Ligase
pBK10
10kb
KmR
ori
ApR
pUC12::A25
ApR
Cex
ori
EcoRI
Hind III
EcoRI
pHind III
Ligase
pUC8::CenA
ApR
CenA
ori
EcoRI
pHind III
EcoRI
Hind III
Ligase
pFWP10
12.6kb
KmR
ori
Cex
ApR
EcoRI
pBamHI
EcoRI
pBamHI
Ligase
pFWP30
11.9 kb
KmR
ori
CenA
ApR
pFWP1030
14.5 kb
KmR
ori
Cex
ApR
CenA

E. coli and *B. lactofermentum*. The cryptic plasmid originally present in the wild-type strain of *B. lactofermentum* was incompatible with pBK10 and its derivatives under selective pressure with Km (not shown).

3.3. Gene expression and detection of cellulolytic activities

The cellulase genes were expressed in both *E. coli* and *B. lactofermentum* transformants carrying pFWP10, pFWP30 or pFWP1030 (Fig. 7). Only weak Exg activity was detected in *E. coli* cells carrying pFWP10 or pFWP1030, even after incubation for five days on L agar-MUC plates at 30°C (not shown). On the other hand, strong Exg activity was detected in *B. lactofermentum* carrying pFWP10 (Fig. 7). EngA activity was detected in *E. coli* cells carrying pFWP30 or pFWP1030 (Fig. 7). EngA activity was also detected in *B. lactofermentum* carrying pFWP30 or pFWP1030 (Fig. 7). Surprisingly, Exg activity was not detected by plate assay in *B. lactofermentum* carrying pFWP1030, even though the plasmid could be recovered intact from the cells.

In broth, pFWP30 or pFWP1030 expressed five to seven times more EngA in *B. lactofermentum* (691 units/20 mL and 509 units/20 mL, respectively) than in *E. coli* (97 units/20 mL and 108 units/20 mL, respectively) and pFWP1030 expressed over seven times more Exg in *B. lactofermentum* (5.8 units/20 mL) than in *E. coli* (0.8 units/20 mL) (Table V). There was a striking difference in the location of the enzyme in the two organisms: more than 95% of the activity in *B. lactofermentum* was in the culture supernatant, compared with only 30 to 50% for *E. coli*. However, the level of expression in *E. coli* was very low. Similar results were obtained for cultures grown in minimal medium (Table V). Cellulase activities detected in *E. coli* supernatants probably results from cell lysis since most of the activity is usually found in the periplasm of exponentially growing cells. In *E. coli* carrying pFWP1030, the amount of Exg produced was almost identical to that produced by *E. coli* carrying pFWP10 (Table V). Again, Exg activity in *B. lactofermentum* carrying pFWP1030 was about ten-fold less than that in *B. lactofermentum* carrying pFWP10 (Table V). This difference was not seen with the corresponding *E. coli* strains. The reason for this is unknown.

Fig. 7. Screening of cellulase activities in transformants. Transformants were streaked on L agar containing 50 μg Km/mL; 500 μM MUC; and 1% CMC. The plate was photographed under long-wave UV light to detect Exg activity (plate 1), then stained with Congo red (Teather and Wood, 1982) before photographing under white light to detect EngA activity (plate 2). *E. coli* was transformed with (A) pBK10, (B) pFWP10, (C) pFWP30, (D) pFWP1030; *B. lactofermentum* was transformed with (E) pBK10, (F) pFWP10, (G) pFWP30, (H) pFWP1030.

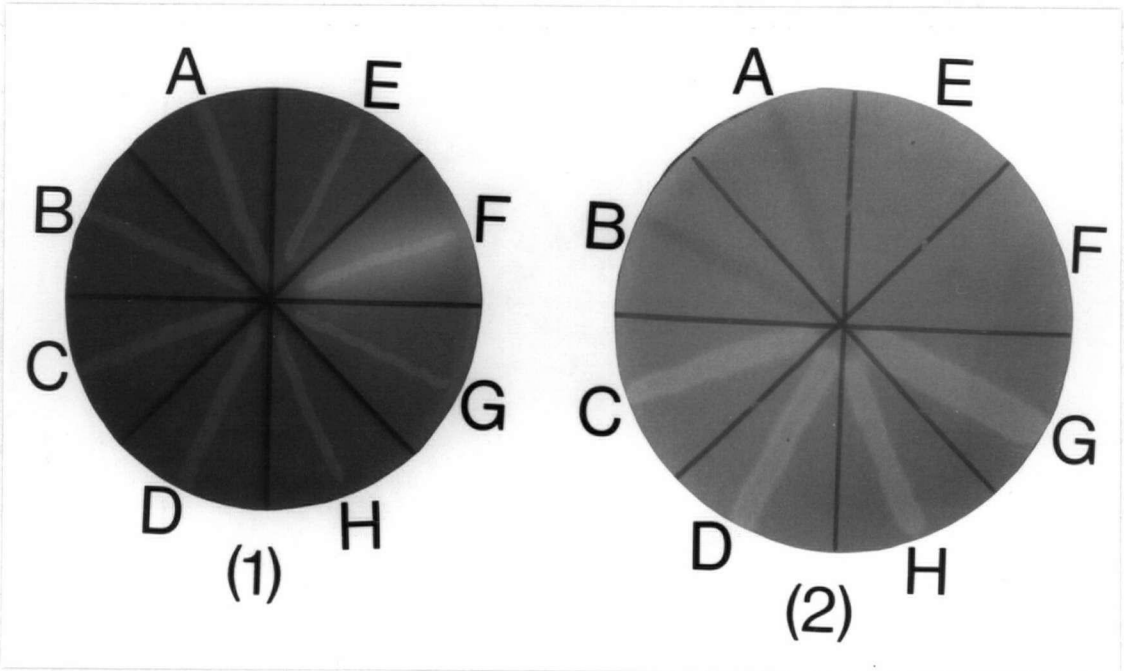


Table. V. Cellulolytic activities of bacterial strains. Each strain was inoculated from a single colony into 20 mL of medium containing Km sulfate at 50 µg/mL. The cultures were incubated at their optimal growth temperature, *B. lactofermentum* at 30°C and *E. coli* at 37°C. The cells were collected by centrifugation at the end of the exponential phase of growth and ruptured by sonication. Cell extracts (c.x.) and supernatants (s/n) were used to measure cellulase activities at 37°C. In the Exg assay, one unit of enzyme released 1 nmol of pNP per min per mL of culture. In the EngA assay, one unit of enzyme released 1 nmol of glucose equivalents per min per mL of culture. Specific activity is units per mg of protein. Symbol "<" is less then 0.1 unit/mg of protein for Exg and less then 1.0 unit/mg of protein for EngA. The total units (Tot. U.) are given for 20 mL of culture.

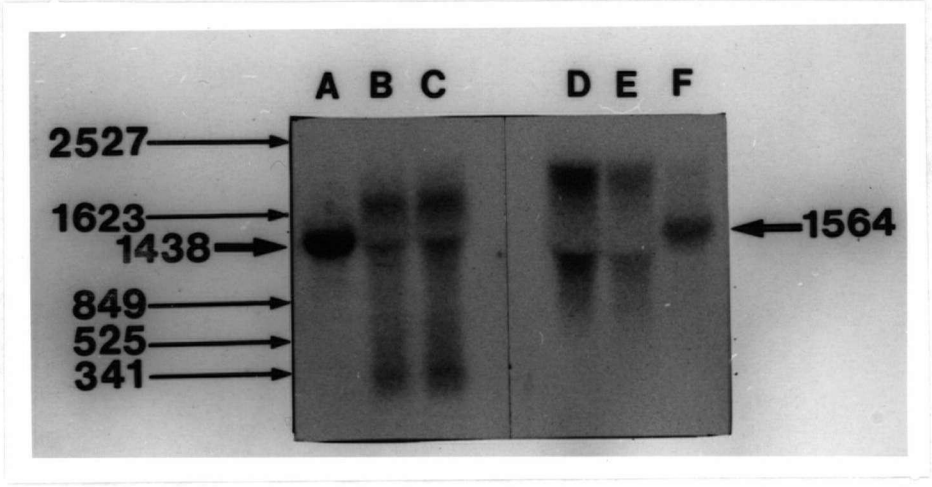
	Brevibacterium lactofermentum				Escherichia coli			
	pNPC		DNS		pNPC		DNS	
	Sp.Act.	Tot.U.	Sp.Act.	Tot.U.	Sp.Act.	Tot.U.	Sp.Act.	Tot.U.
rich media								
plasmids								
c.x.	<		<		<		<	
pBK10		0		0		0		0
s/n	<		<		<		<	
c.x.	0.14		2.7		0.10		2.1	
pFWP10		5.8		71		0.8		14
s/n	3.24		75		0.61		87	
c.x.	<		20		<		20	
pFWP30		0		691		0		97
s/n	<		829		<		33	
c.x.	<		12		0.12		9.4	
pFWP1030		0.5		509		0.8		108
s/n	0.46		1118		0.43		65	
minimal media								
plasmids								
c.x.	<		<		<		<	
pBK10		0		0		0		0
s/n	<		<		<		<	
c.x.	0.34		2.7		0.12		3.4	
pFWP10		3.5		109		0.3		12
s/n	1.7		589		0.48		115	
c.x.	<		20		<		16	
pFWP30		0		1408		0		169
s/n	<		1064		<		154	
c.x.	<		45		<		23	
pFWP1030		0.3		868		0.4		131
s/n	0.2		751		0.54		51	

Restriction analysis of plasmid DNA indicated that the structural gene and the regulatory region were both still present. However, it is possible that some promoter located in the pBR322 plasmid is used by *B. lactofermentum* and is responsible for the expression to the cellulase genes. Since the *bla* gene is not expressed in *B. lactofermentum*(Santamaria *et al.,* 1984). it is possible that the P4 promoter identified by Stuber and Bujard (1981) (located between the origin of replication of pBR322 and the *bla* gene and oriented toward the *bla* gene) could be responsible for the cellulase gene expression.

3.4. Detection of cellulase gene specific transcripts

The positive control RNA species isolated from *C. fimi* and detected by the *cen*A probe was approximately 1400 bases in length (Fig. 8, lane A) as previously observed (Greenberg *et al.,* 1987a). The RNA species isolated from *B. lactofermentum* carrying pFWP30 plasmid and detected by the *cen*A probe are approximately 1700, 1400 and 300 bases in length (Fig. 8, lane B and C). The positive control RNA species isolated from *C. fimi* and detected by the *cex* probe was approximately 1500 bases in length (Fig. 8, lane F) as previously observed (Greenberg *et al.,* 1987a). The RNA species isolated from *B. lactofermentum* carrying pFWP10 plasmid and detected by the *cex* probe are approximately 2300 and 1300 bases in length (Fig. 8, lane D and E). The stringencies of probe hybridizations were based on the specificity of the positive controls. Nevertheless, a negative control like mRNA from *B. lactofermentum* carrying the shuttle vector pBK10 would have been appropriate to justify the specificity of the detected hybrids in that bacterium. Assuming that the RNA polymerase putative stop sites for the cellulase genes are recognized by *B. lactofermentum* transcriptional apparatus and considering the additional 5' end of non-coding *C. fimi* DNA present in pFWP10 and pFWP30 plasmids, the largest *cen*A and *cex* transcripts (1700 and 2300 bases in length, respectively) would originate from a common region in pBR322 just upstream from the *Eco*RI site. In the *cex* lanes, none of the transcript (lane D and E) detected from *B. lactofermentum* carrying pFWP10 plasmid was similar in length with the *cex* positive control *C. fimi* mRNA (lane F). This suggests that

Fig. 8. Detection of specific transcripts by Northern blot analysis. RNA was isolated from *C. fimi* culture grown in basal medium supplemented with 1% CMC and *B. lactofermentum* carrying pFWP10 or pFWP30 plasmids grown in rich medium containing 50 µg Km/mL. Total RNA was denatured with formaldehyde, fractionated on a formaldehyde gel containing 1% (w/v) agarose, and blotted onto a Biotrans membrane (Pall, Inc.). Hybridizations were done with ^{32}P-labelled 30-mers oligos complementary to *cen*A mRNA (lane A, B, C) or *cex* mRNA (Lane D, E, F). The filters were exposed to X-Ray films for 18 hrs in the case of *cen*A and 3 days in the case of *cex*. Lane A and F, 20 µg of *C. fimi* RNA; B and C, 20 and 30 µg of *B. lactofermentum*/pFWP30 RNA, respectively; lane D and E, 30 and 20 µg of *B. lactofermentum*/pFWP10 RNA, respectively. Size markers (in nucleotides) are *Hae*III digested single-stranded M13mp11 ^{32}P-labelled DNA fragments. The expected size of *cen*A mRNA (1438 nt) and *cex* mRNA (1564 nt) are indicated.



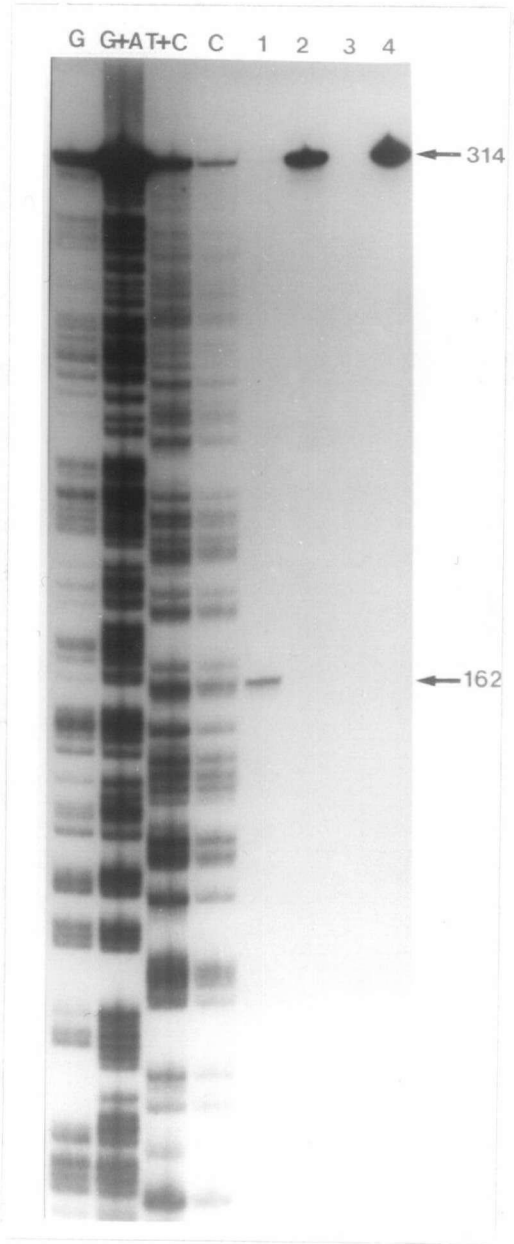
initiation of transcription of the *cex* gene originated upstream from the *C. fimi* regulatory sequences. In contrast, in the *cen*A lanes, one of the transcripts detected from *B. lactofermentum* carrying pFWP30 plasmid is similar in length (about 1400 bases) with the *cen*A positive control *C. fimi* mRNA (Lane A) which suggests that initiation of transcription of the *cen*A gene may occur at the expected site. Nevertheless, Northern blot analysis does not identify transcriptional initiation sites but the approximate size of a transcript which is prone to processing or nuclease degradation. Since only *cen*A gave a comigrating mRNA band, the initiation of transcription of the *cen*A mRNA in *B. lactofermentum* carrying plasmid pFWP30 was further investigated by fine resolution S1 nuclease mapping.

3.5. Mapping the *cen*A mRNA 5' ends

Transcripts produced *in vivo* by *C. fimi* cells grown on CMC containing basal medium and by *B. lactofermentum* carrying pFWP30 plasmid were analysed by high resolution S1 nuclease mapping essentially as described (Greenberg *et al.,* 1987a). A 315 bp *Sma*I-*Sal*I DNA restriction fragment isolated from the plasmid pNG101 (Greenberg *et al.,* 1987a) was used as the hybridisation probe. The probe uniquely labelled at the 5' *Sal*I site, was made single stranded by denaturation and hybridised with total RNA isolated from *C. fimi* and *B. lactofermentum* carrying pFWP30 plasmid. The mixtures favouring RNA-DNA hybrids were degraded with S1 nuclease. The length of the protected probe species were determined by gel electrophoresis under denaturing conditions. A sequencing ladder of the *Sma*I-*Sal*I probe created by chemical cleavage served as control to measure the length of the residual probe and identify the 5' end of the CenA mRNA.

The autoradiograph (Fig. 9) revealed a major mRNA 5' end in the *C. fimi* control lane (Lane 1) with a protected fragment of 162 bases in length. When these sequencing data and the weaker exposure of the ladder were compared with the previously published data, it was estimated that the mapped 5' end corresponded to a C base located 52 bases upstream of the ATG as previously observed as a -1 start site (Greenberg *et al.,* 1987a). The conditions of hybridization must have

Fig. 9. S1 nuclease protection analysis of *cen*A transcripts. After hybridisation with RNA and treatment with S1 nuclease, the remaining *Sma*I-*Sal*I *cen*A-specific labeled DNA probe was analysed in a 8% polyacrylamide-7 M urea sequencing gel alongside probe sequenced by the base-specific chemical cleavage method (Maxam and Gilbert, 1980). The chemical sequencing ladders of the probe is shown. Protection of the probe by: lane 1, RNA from *C. fimi* ; lane 2, RNA from *B. lactofermentum*/pFWP30; lane 3, by yeast tRNA. Lane 4 represent the probe alone. Sizes are in nucleotides.



been optimal since a unique band was detected (Lane 1) where the usual self DNA-DNA probe hybrid (Lane 4) was not. Interestingly, the full length of the probe was protected by *B. lactofermentum* mRNA (lane 2) carrying pFWP30 compared with the untreated probe alone (Lane 4) indicating that the 5' end of the *cen*A mRNA in that strain is located upstream of the *Sma*I site. Furthermore, this observation suggests that the *cen*A putative regulatory sequence is not used or recognized by the *B. lactofermentum* transcriptional apparatus and that an unreported promoter located somewhere upstream is used by *B. lactofermentum* for expression of the cloned genes. This is a likely explanation of the lack of exoglucanase activity in *B. lactofermentum* carrying the pFWP1030 construct. The labelled probe was not protected in mapping studies with control yeast tRNA (lane 3). With RNA isolated from *B. lactofermentum* as well as *E. coli* cells carrying recombinant plasmids with putative *C. fimi* promoter sequences, transcripts were never found which initiated within the inserts (Greenberg *et al.*, 1987b; this thesis).

Several reasons could be suggested for these observations: (1) *E. coli* and *B. lactofermentum* RNA polymerases may not have been able to recognize the *C. fimi* promoters due to the absence of an appropriate sigma factor, (2) the RNA polymerases may have recognized the promoters but were incapable of initiating or elongating the transcripts, or (3) the resulting "hybrid" transcripts may have been intrinsically unstable. It is proposed that the expression of the *cex* and *cen*A genes in *B. lactofermentum* originated from a similar promoter sequence located within the vector itself and that conclusive evidence for that promoter would require more stringent studies beyond the scope of this study.

In conclusion, even though the cellulase gene putative promoter sequences did not function as expected in the closely related bacterium *B. lactofermentum*, the cellulase genes were expressed and their products were secreted into the culture supernatant. The *B. lactofermentum* system could allow studies of the implicated sequences and factors required for secretion of the cellulase enzymes in this bacterium.

PART 2. The characterization of recombinant *C. fimi* ß-glucosidase A from *E. coli.*

3.6. The expression of recombinant CbgA in *E. coli*

3.6.1. Cloning, subcloning and specific activities

The original clone *cbg*A was first isolated from a family of immunopositive *E. coli* transformants generated from a pBR322-*C. fimi* genomic library (Gilkes *et al.,* 1984). The recombinant plasmid pEC62 carries a 7.2 Kb-*Bam*HI insert (Fig. 10) and expresses in *E. coli* an enzyme with ß-glucosidase activity (CbgA). Expression of the cloned gene, *cbg*A, in *E. coli* requires a heterologous promoter (Bates, 1987; this thesis). To test for the presence of a CbgA translational initiation codon within the 7.2 Kb insert, a translational frameshift was created at the 5'-end *Bam*HI site of pEC62 (Fig. 11). A loss of the enzymatic activity would have resulted if the recombinant ß-glucosidase originated from a translational fusion between CbgA and the Tc^R determinant of pBR322, whereas, the detection of ß-glucosidase activity would have confirmed the presence of a CbgA translational start site within the insert, and perhaps the presence of the entire CbgA structural gene as well. As an initial step, the problem of having an additional *Bam*HI site at the 3'-end of the insert was circumvented by deleting the 3' end-1.5 Kb *Bgl*II-*Bam*HI fragment followed by religation of the compatible ends giving rise to pEC62.1. No loss in specific activities was observed from either the pEC62.1 or pEC62.2 genetic constructs when expression was compared to that of the original plasmid pEC62 (Fig. 10). These data confirmed the presence of a CbgA translational start site within the 7.2 Kb *Bam*HI insert and suggests that 5.6 Kb of insert DNA is sufficient to produce an active peptide.

To increase the expression of *cbg*A in *E. coli* , the 7.2Kb *Bam*HI insert was transferred from pEC62 to the high copy number plasmid pUC13 giving rise to pUC13:62 (Bates, 1987). No increase of the specific activity was observed from this construct (Fig. 10). Deletions were generated at the 5' end of the 7.2Kb insert with exonuclease *Bal*31 and fragments of various sizes were isolated and subcloned into pUC13. *E. coli* transformants were screened by the MUGase plate assay for

Fig. 10. Diagram of pEC62 and derivatives. The circular plasmids are shown in a linear fashion. The open bars represent pBR322 DNA; the dashed bars represent *C. fimi* DNA; the solid bars represent pUC13 DNA; The single lines represent deleted DNA regions. The arrow indicates the functional orientation for the Tc^R or Lac promoters as well as the orientation of the *cbg*A gene in the insert. The scale of the drawing is given at the top. The length of each plasmid is indicated. The deletion in pUC13:62Δ31 extends to the 5' proximal *Xho*I site. The specific ß-glucosidase activity for each plasmid expressed in *E. coli* is given in units (1 U releasing 1 nmol of pNP per min at 37°C) per mg of protein of total cell extracts. Restriction sites : B, *Bam*HI; Bg, *Bgl* II; K, *Kpn*I; X, *Xho*I.

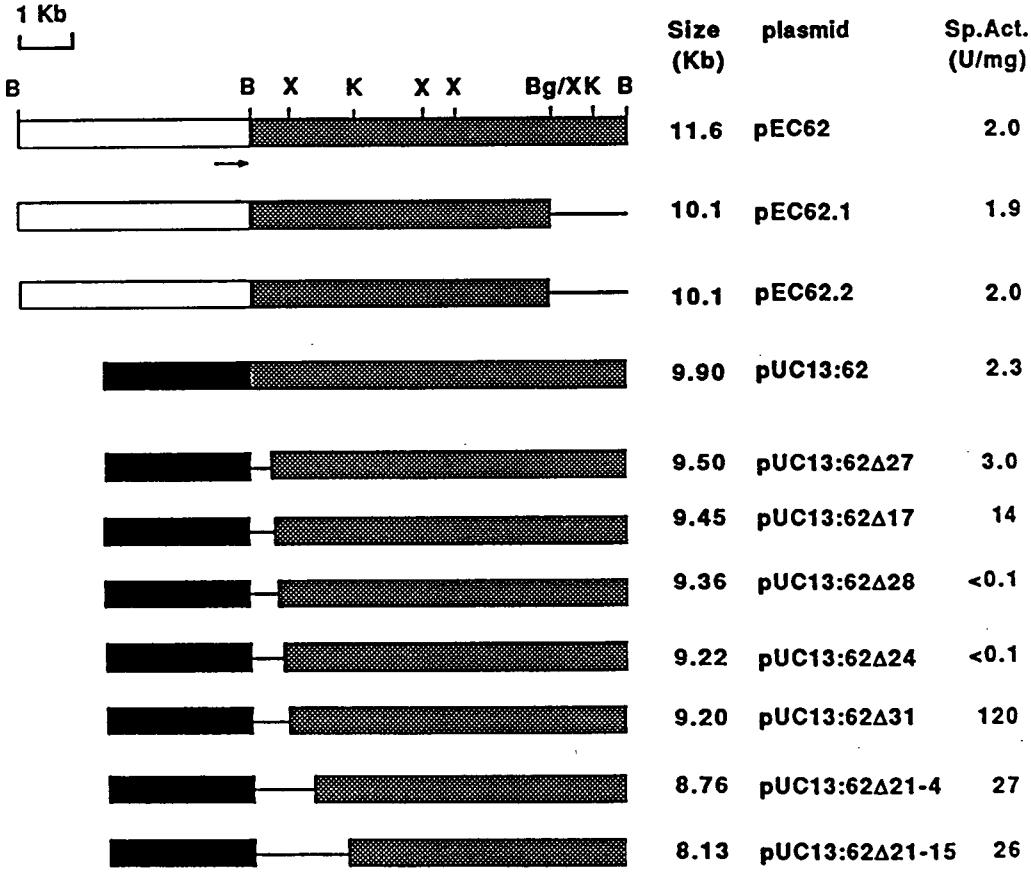
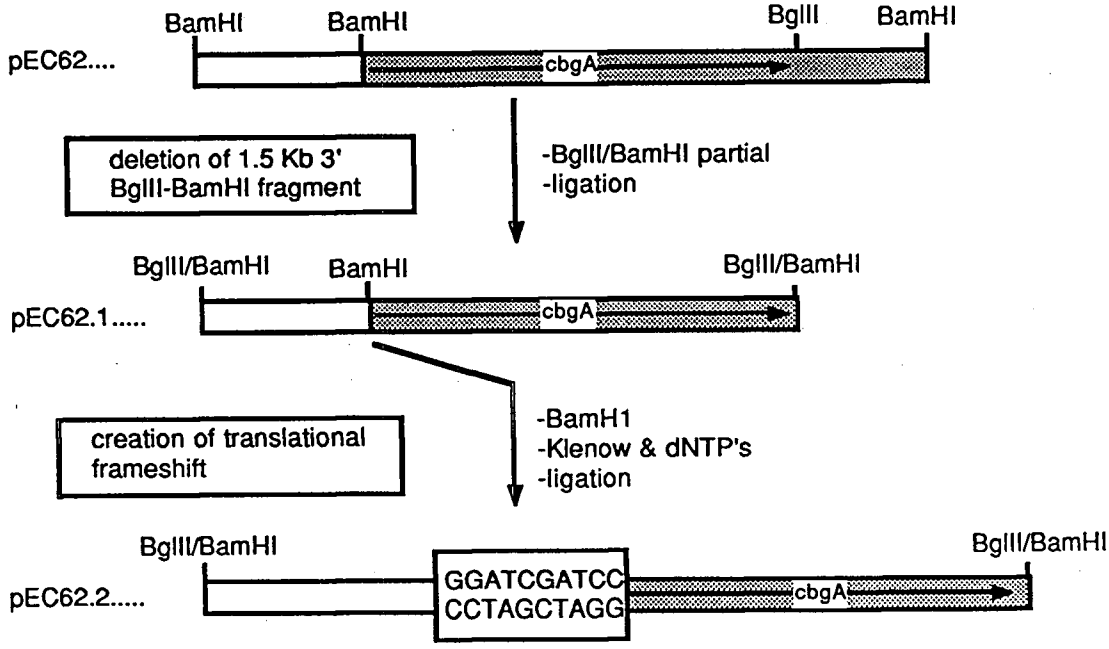


Fig. 11. Scheme for the generation of a translational frameshift in pEC62. Plasmids are shown in a linear fashion. The original recombinant plasmid pEC62 carrying a 7.2 Kb *Bam*HI fragment of *C. fimi* DNA (dash boxes) in pBR322 (open boxes) was digested to completion with *Bgl*II and partially digested with *Bam*HI. Ligation of compatible cohesive ends of the remaining 10.1 Kb fragment eliminated the 3' proximal *Bam*HI site creating pEC62.1 containing a unique 5' proximal *Bam*HI site. Finally, pEC62.1 was digested to completion with *Bam*HI and the ends were filled with dNTP's by Klenow enzyme followed by religation of the newly generated blunt ends creating pEC62.2. This construct contains the translational frameshift at the *Bam*HI Tc^R promoter proximal region. Relevant restriction sites are indicated. The location of the *cbg*A gene and orientation of transcription is given. Putative and newly generated DNA sequence is shown in lower diagram of pEC62.2.



increase in ß-glucosidase activity (Bates, 1987). The size of the deletions was estimated by restriction digest analysis. The plasmid giving the highest expression in the *E. coli* (pUC13:62Δ31) has a deletion of about 700 bp from the 5' end of the insert and gives a 60 fold increase in specific activity compared with the original clone pEC62 (Fig. 10). The shortest deletion (pUC13:62Δ28) which results in the loss of ß-glucosidase activity is about 540 bp. This suggests that the CbgA translational start site is located within this 540 bp region. It is appropriate to mention at this point that the Δ24, Δ27 and Δ28 deletions are out of frame fusions with the ß-Gal alpha peptide, whereas, the other deletions are in frame fusions (see section 3.10) and they will be discussed later on. The expression of ß-glucosidase activity detectable in *E. coli* cells carrying pUC13:62Δ21-15 or pEC62.1 plasmids suggests that less than 3.8 Kb of DNA is sufficient to produce an active peptide. The high expressing clone pUC13:62Δ31 was used to determine the location of recombinant enzyme in *E. coli*.

3.6.2. The location of recombinant CbgA in *E. coli*

The location of CbgA in *E. coli* cells carrying pUC13:62Δ31 was determined by fractionation of the cellular compartments into cytoplasm, periplasm and total membrane fractions. Cells grown to midlog and stationary phases were fractionated and each fraction was tested for the presence of enzymes and sugar markers that are specific for and representative of *E. coli* compartments (Table VI). In either logarithmic or stationary phase cells, almost all of the ß-glucosidase activity was cell associated, mostly found in the cytoplasmic fraction and partially in the membrane fractions of *E. coli* cells (Table VI). Little ß-glucosidase activity was detectable in the periplasmic fractions of *E. coli* cells carrying either pUC13:62 (not shown) or its Δ31 derivative (Table V). No activity was detectable in culture supernatants of either cultures (data not shown). This is not surprising since *E. coli* rarely secretes proteins into the culture medium.

Since ß-glucosidases are sometimes membrane associated (Hwang and Suzuki, 1976; Umile and Kubicek, 1986), the total membrane fraction of the stationary phase culture was analysed further by isopycnic

Table VI. Localisation of ß-glucosidase activity in *E. coli*JM83/pUC13:62Δ31

Fractions	Protein (mg)	ß-glucosidase (total mU)	G-6-PDH (total mU)	ß-lactamase (total mU)	KDO (µg)	NADHox (total mU)
(LOG PHASE)						
SET wash	0.3	9	n.d.	12	n.d.	<1
cytoplasm	2.5	415	513	2	<1	1220
membranes	1.3	30	<1	<1	63	1459
periplasm	0.3	7	<1	28	<1	<1
Total cells	5.0	496	586	37	100	5200
(STATIONARY PHASE)						
SET wash	0.9	<1	n.d.	15	n.d.	<1
cytoplasm	22.4	3219	698	147	<1	255
membranes	10.0	1374	14	16	87	1894
periplasm	0.9	1	<1	53	18	<1
Total cells	61.6	4468	635	235	224	2080

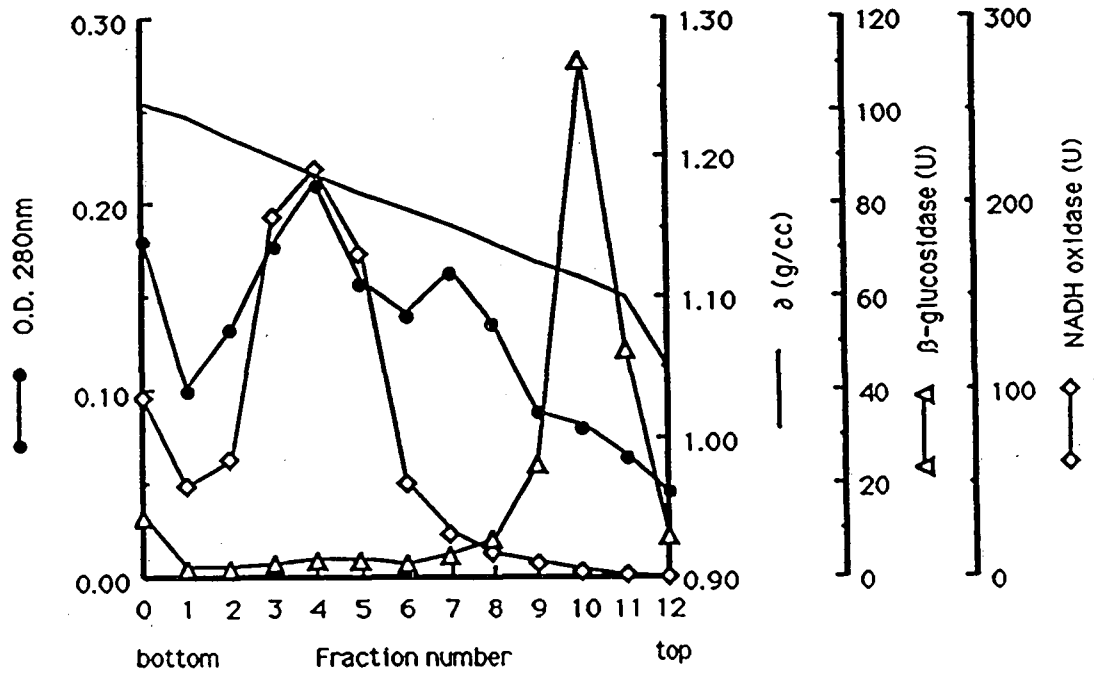
Protein fractions in this table were obtained as described in section 2.5.1. Enzymatic assays, sugar content and protein determination in this table are as described in section 2.4.

SET : 40% sucrose- 33 mM Tris- 0.1 mM EDTA (pH 7.0)

< : less then

One unit of enzyme in this table releases one µmol of product per min under specified conditions.

Fig. 12. Profile of *E. coli* JM83/pUC13:62Δ31 membrane fraction separated on a sucrose density gradient. Approximately 950 µg of proteins from the membrane fraction of stationary phase cells was loaded onto a 30% to 55% sucrose step gradient containing 5 mM-EDTA and 10 mM-HEPES buffer (pH 7.4). Fractions of 350 µL were collected dropwise, absorbance at 280 nm, ß-glucosidase and NADH oxidase activities were measured. The bottom and top of the gradient are shown. Sample 0 corresponds to the pellet at the bottom of the tube.



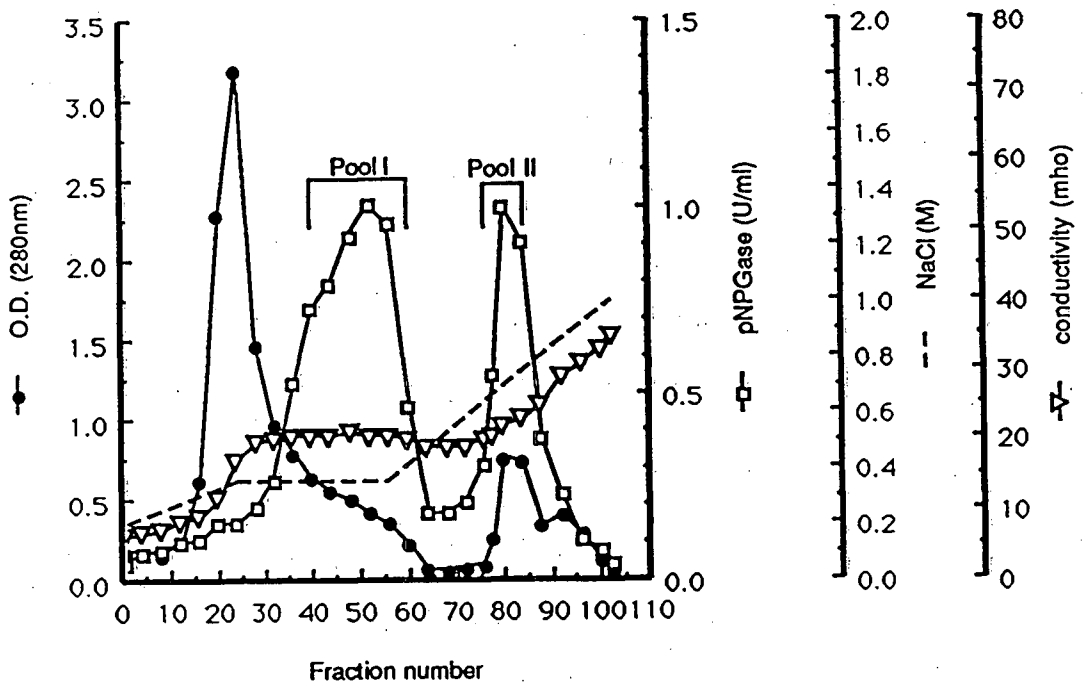
sucrose density gradient centrifugation (Fig. 12). The location of ß-glucosidase activity was compared with that of NADH oxidase which is membrane associated in *E. coli* (Osborn *et al.,* 1972). The small amount of sediment at the bottom of the tube (sample 0) was kept for analysis. One has to consider the fact that the use of Mg^{2+} during the osmotic shock treatment or at any point in the isolation procedure results in extremely poor separation of inner and outer membranes (Osborn and Munson, 1974). Also, membranes from cells grown on very rich media appear to be more difficult to separate because of a possible increase in the number of zones of adhesion between the outer and cytoplasmic membranes in rapidly growing cells (Gerhardt *et al.,* 1981). Therefore, virtually all the membrane material (Fig. 12) was recovered at the intermediate position "M" (Osborn and Munson, 1974). About 94% of the recovered NADH oxidase activity was found at buoyant density of about 1.19 g/cc indicating the location of the bacterial inner membranes. About 67% of the recovered ß-glucosidase activity was located on the top of the gradient. These data suggest that the recombinant enzyme CbgA expressed from pUC13:62Δ31 plasmid is not bound to the bacterial inner membrane like the *E. coli* NADH oxidase enzyme but rather seems to co-sediment with the membranes during ultracentrifugation. This suggests that CbgA may form dense particles perhaps by multimeric association with itself or other proteins present in the sample.

3.7. Characterization of recombinant CbgA from *E. coli* /pUC13:62Δ31

3.7.1. Purification of CbgA from *E. coli.*

The high expressing clone pUC13:62Δ31 was used to produce the recombinant ß-glucosidase in *E. coli* JM83 for purification. The elution profile of *E. coli* cytoplasmic fraction from DEAE-sephadex A-50-120 column (Fig. 13) shows two major peaks of ß-glucosidase activity eluting at 0.35 M NaCl and 0.55 M NaCl respectively. The fractions under the first peak of activity which contained about 50% of the total ß-glucosidase activity (fraction 40 to 60) were pooled (Pool I) and processed further. The fractions under the second peak of activity which contained about 25% of the total ß-glucosidase activity (fraction 76 to 84) were also pooled (Pool II) and proteins were concentrated by ultrafiltration.

Fig. 13. Fractionation of *E. coli*JM83/pUC13:62Δ31 cytoplasmic proteins on a DEAE-sephadex A-50-120 column. The cytoplasmic fraction of *E. coli* carrying pUC13:62Δ31 was applied to a DEAE-sephadex A-50-120 (2.5 cm x 11 cm) column. After washing the column at a flow rate of 44 mL/hr with 50 mM Tris-5 mM EDTA (pH 7.0), the bound proteins were eluted by stepwise increases in the NaCl concentration as shown. Fractions of 2.5 mL were collected and the absorbance at 280 nm, conductivity and pNPGase activity were determined. Fractions 40 to 60 (Pool I) and 76 to 84 (Pool II) were combined separately.



The proteins in Pool I were separated further by ion exchange on Mono Q column (Fig. 14 and 15). Several fractions obtained during purification of the recombinant enzyme were analysed by SDS-PAGE. The purification scheme gave a fraction (Fig. 14, fraction 24) containing a major polypeptide, as visualized by Coomassie blue staining of SDS-PAGE (Fig. 16-lane 7). The purified recombinant enzyme CbgA from Mono Q resin has a molecular weight of approximately 183 000 (p183) and a specific activity of 12957 mU/mg of protein with pNPG as substrate (Table VII). The recombinant enzyme was purified about 100-fold to homogeneity, as visualized from SDS-PAGE. A yield of 2.1% was obtained from the purification scheme and was estimated from the total recovery after Mono Q column (Fig. 14). The presence of more than one peak of activity eluting from DEAE-sephadex A-50-120 is not because of the presence on the 7.2 Kb insert of two separate ß-glucosidase encoding genes. Western blotting analysis of fully denatured proteins from Pool II has indicated the presence of a major polypeptide of 183 kDa as well as much larger components that could originate from aggregation of CbgA with itself or with other peptides (data not shown).

The storage at 4°C of Pool I from DEAE-Sephadex A-50-120 gave rise to additional products (Fig. 16-Lane 10 and 11) which coeluted with the p183 peptide during a second purification on Mono Q column (Fig. 15) . Based on their apparent molecular weights and associativity, these additional peptides (p137 and p60) could originate from proteolysis of the p183 peptide (see section 3.10). Furthermore, the specific activities of fractions containing p183-p137 or p183-p137-p60 peptides suggests that p60 does not have ß-glucosidase activity. Further evidence that the p137 polypeptide is a cleavage product of the p183 polypeptide comes from the observation that p137 reacted with the anti-CbgA antibodies on Western blots (see section 3.9). DNA sequence and N-terminal amino acid sequence analysis of p137 brings about conclusive evidence of this relatedness. However, I have no evidence to support the origin of the p60 polypeptide (see section 3.10). Some cleavage products are also observed in *C. fimi*, and in both cases, they seem to interact in an aggregating fashion. The function of the cleavage products and the nature of the aggregation of CbgA subunits remains unknown.

Fig. 14. Fractionation of Pool I on Mono Q. One fifth of the Pool I fraction from the DEAE-sephadex column was applied to a Mono Q anion-exchange column. After washing the column at a flow rate of 1 mL/min with 50 mM Tris-5 mM EDTA (pH 7.0), bound proteins were eluted by stepwise increases in the NaCl concentration as shown. Eluting fractions were monitored by absorbance at 280 nm and analysed for pNPGase activity.

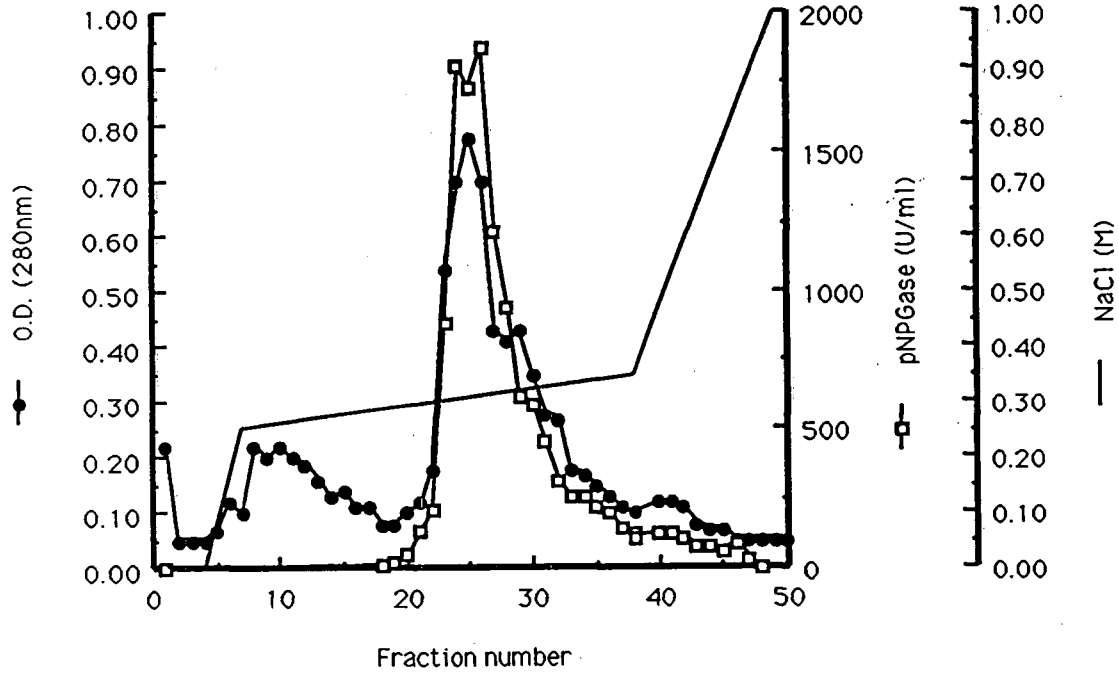


Fig. 15. Fractionation of stored Pool I on Mono Q. One fifth of the Pool I fraction from the DEAE-sephadex column that was stored at 4°C for several months was applied to a Mono Q anion-exchange column. After washing the column at a flow rate of 1 mL/min with 50 mM Tris-5 mM EDTA (pH 7.0), bound proteins were eluted by stepwise increases in the NaCl concentration as shown. Eluting fractions were monitored by absorbance at 280 nm and analysed for pNPGase activity. Fractions 22 to 27 and 28 to 31 were pooled separately.

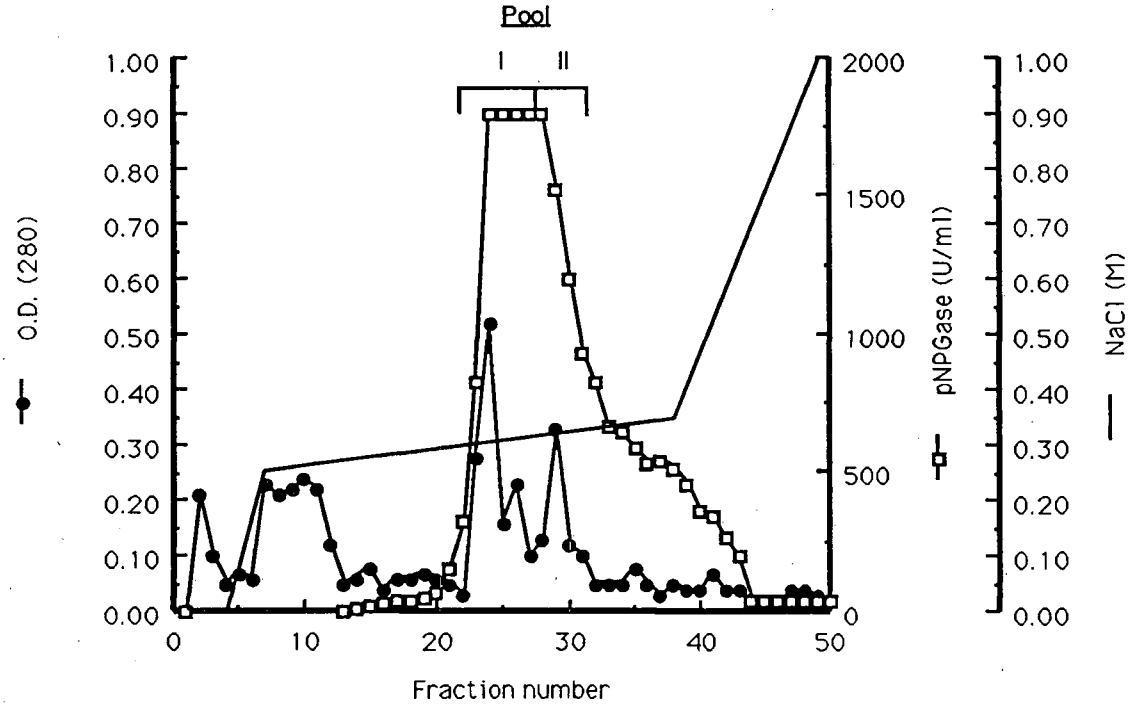


Fig. 16. 7% SDS-PAGE analysis of protein fractions from the ion-exchange purification of recombinant CbgA. Lane 1, 26 µg (10 U) of French press total crude extract; lane 2, 35 µg (10 U) of the cytosol fraction; lane 3, 58 µg (10 U) of the membrane fraction; lane 4, 28 µg (10 U) of the streptomycin sulfate fraction; lane 5, 15 µg (10 U) of the ammonium sulfate fraction; lane 6, 3 µg (10 U) of Pool I fraction; lane 7, 1.4 µg (7U) of fraction 24 from first Mono Q column; lane 8, 1.4 µg (6U) of fraction 25 from first Mono Q column; lane 9, 1.4 µg (5U) of fraction 26 from first Mono Q column; lane 10, 6.9 µg (27 U) of fraction 28 to 31 from second Mono Q column; lane 11, 6.3 µg (56 U) of fraction 22 to 27 from second Mono Q column. Proteins from lane 10 and 11 were isolated from the stored Pool I fraction and ran on a separate gel where Rf's were adjusted accordingly. M_r standards (kDa) are shown on the left. Polypeptides relevant to this thesis are identified on the right.

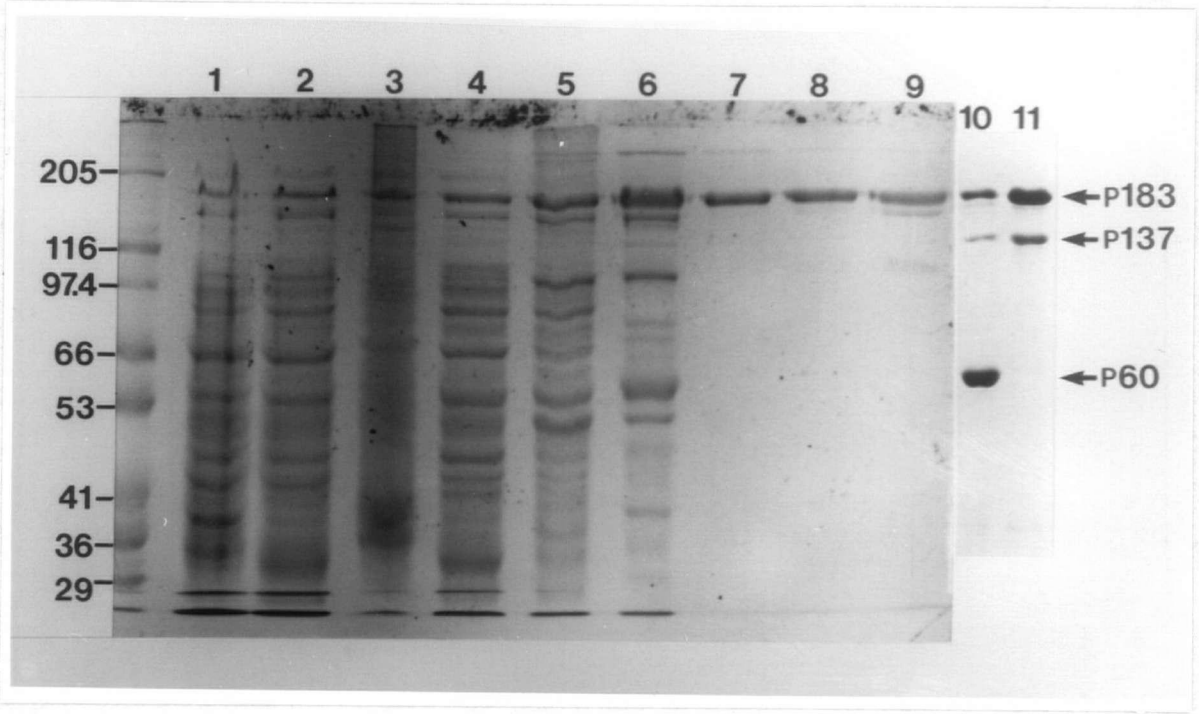


Table VII. Purification of recombinant CbgA from *E. coli* JM83/ pUC13:62Δ31

Fraction	Tot. protein (mg)	Tot. units	Specific activity (U/mg of prot.)	Purification factor	Yield (%)
Crude extract	6300	$9.5x10^5$	151	-	-
Cytoplasm	5900	$7.1x10^5$	120	1	100
Amm. SO4 cut	352	$2.3x10^5$	653	5.4	42
DEAE-Sephadex	44	$4.9x10^4$	1114	9.3	35
Mono Q-fraction 24	0.14	$1.8x10^3$	12957	108	2.1

The values in this table are from 10 L of cell culture with various fractions obtained during the purification of recombinant CbgA.
The units in this table are defined as nmol of pNP released per min at 37°C.

3.7.2. The effects of pH and temperature on CbgA activity

The ß-glucosidase activity of the crude extract from *E. coli* carrying pUC13:62Δ31 and the purified CbgA was tested using 5 mM pNPG as substrate at 37°C at various pH values (Fig.17-A). Both the crude extract and purified enzyme showed maximal activity at about pH 5.5. The crude extract shows higher activity than the purified enzyme below and above pH 5.5 perhaps because of a protein-protein interaction protecting CbgA from acid/alkaly inactivation or activation at the active site.

The ß-glucosidase activity of the crude CbgA enzyme extract from *E. coli* carrying pUC13:62Δ31 and the purified enzyme was tested after 60 min incubation at various temperatures in the presence of pNPG (5 mM) (Fig.17-B). Both samples showed identical behavior with highest activity at a temperature of about 42°C. Nevertheless, at that temperature, a significant denaturation of the enzyme occurs (Fig. 18). The thermal inactivation of the purified enzyme CbgA was monitored by measuring the remaining activity against pNPG at 37°C. The enzyme was incubated at various temperatures in phosphate buffer (pH 7.0) without substrate. The residual activity presented under the form of an Arrhenius plot shows the effect of temperature on the purified enzyme.

3.7.3. The kinetic parameters of recombinant CbgA

The kinetic parameters of the purified recombinant CbgA were determined from Lineweaver-Burke plots for each of the substrates tested (Table VIII). An example of such plots is given in the Appendix (1-A to D). The recombinant enzyme CbgA can hydrolyze a wide variety of aglycones. It acts preferentially on longer cellodextrins (C5>C4>C3>C2) with K_m values comparable to those previously reported for ß-glucosidases from fungi (see Table II). The rate of hydrolysis depends on the nature of the aglycone moiety. The relationship between the V_{max}/K_m values and the number of glucose molecules per cello-oligosaccharide shows an increase in substrate specificity with the increase in length of the substrate. Thus, in the hydrolysis of cellulose, CbgA could be involved more in the hydrolysis of cellodextrins than of cellobiose, assuming equal concentrations of both substrates. This strongly suggests that CbgA is an exo-ß-1,4-glucosidase and must be

Fig. 17-A. Effect of pH on the hydrolysis of pNPG by recombinant CbgA

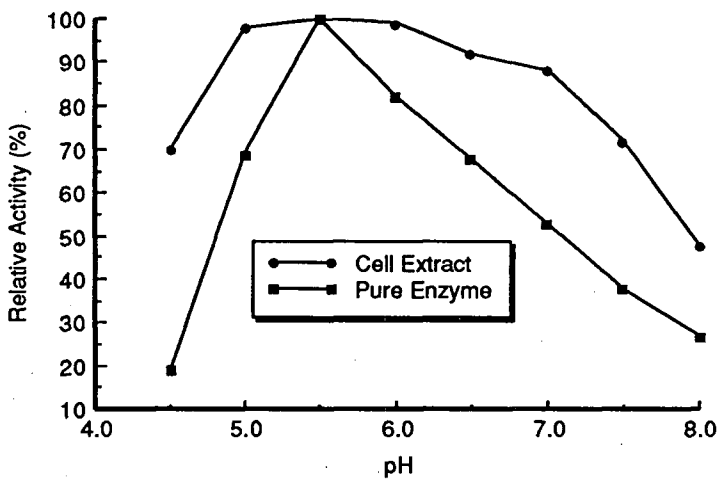


Fig. 17-B. Effect of temperature on the hydrolysis of pNPG by recombinant CbgA

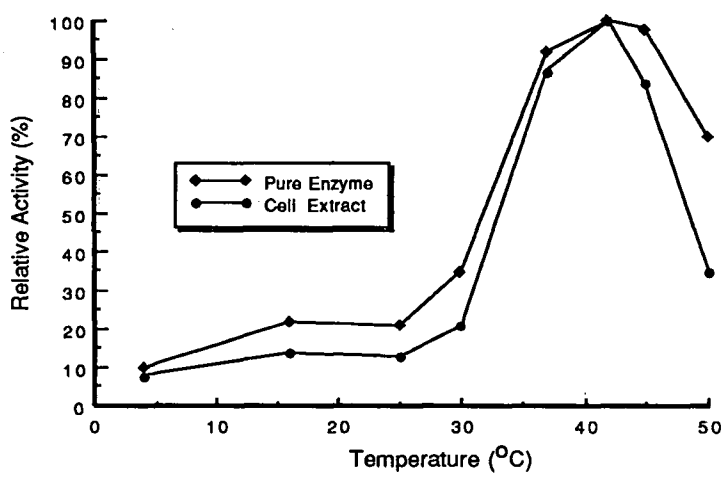


Fig. 18. Thermostability study of purified CbgA

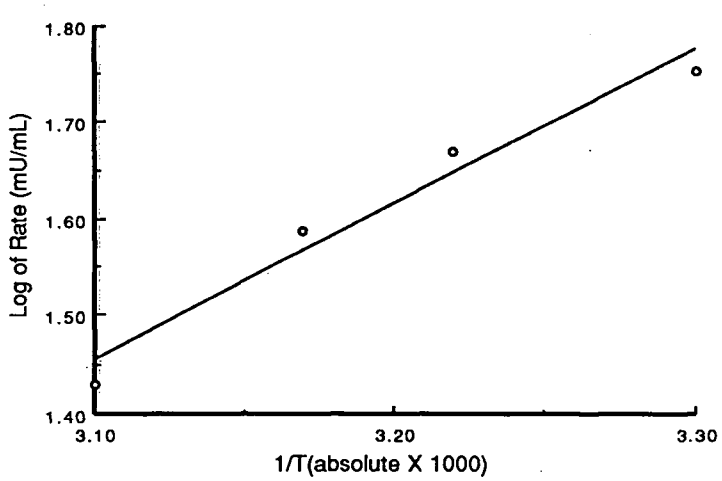


Table VIII. Kinetic parameters of recombinant CbgA on various ß-glucosides

Substrate	K_m (mM)	V_{max} (µmol prod/min/mg)	V_{max}/K_m (µmol prod/min/mg) / mM substrate
pNPG	0.13	4.4	33.8
Cellobiose	2.50	31.2	13
Cellotriose	0.27	8.8	33
Cellotetraose	0.33	20.0	61
Cellopentaose	0.08	9.2	115
Glucose (Ki)	2.00	-	-

The kinetic parameters from this table were determined from Lineweaver-Burk plots as described in section 2.7.2 (also Appendix 1). Proteins were determine according to Bradford (1976) using BSA as a standard.

transported outside the cell to be able to act on the cellodextrins released from hydrolysis of cellulose by *C. fimi* endoglucanases. If this is the case, CbgA should possess a leader peptide like sequence for secretion as for the Exg and EngA, B, and C from the same organism. Furthermore, one may speculate that other ß-glucosidases may be present in *C. fimi* to hydrolyze cellobiose and other ß-glucosides from plants.

The Ki value of 2 mM glucose was measured using pNPG as substrate and is in the same range as those for bacterial and fungal ß-glucosidases ranging from 0.66 to 135 mM glucose. Product inhibition for ß-glucosidases has been reported to be either competitive or non-competitive depending on the source of the enzyme (Coughlan, 1985). In this case, the inhibition of CbgA activity on pNPG by glucose is of the competitive type (see appendix 1-E) meaning that the product glucose competes with the substrate during the formation of enzyme-substrate complexes.

An interesting behavior was observed for the hydrolysis of pNPC (see appendix 1-F) which was also observed and reported elsewhere (Day and Withers, 1986). At high substrate concentration, very little nitrophenol is initially detected at 410 nm whereas at low concentration, a faster release of nitrophenol occurs. This suggests that the enzyme initially hydrolyses the bond between the two sugar residues to produce glucose and pNPG with no change in the absorbance at 410 nm. Subsequently, pNPG will compete with the remaining pNPC and soon enough, adsorbance at 410 nm will increase notably. The recombinant ß-glucosidase would therefore be better classified as an exo-glucosidase enzyme. Interestingly, the recombinant enzyme is also specific for the hydroxyl group at position C2 since it does not hydrolyze the substrate p-nitrophenyl-ß-mannoside (pNP-man) which differs from pNPG by having the C2 hydroxyl group in the opposite configuration. The recombinant enzyme shows weak activity on pNP-Acetylglucosamine which has a larger group and right configuration at this position (data not shown). This result could suggest a role for the C2-OH group of the sugar during enzyme activity or may simply reflect poor binding of the pNP-man with its axial hydroxyl. No change in absorbance was measured on pNP-

Galactoside, pNP-Lactoside or pNP-Xyloside (not shown).

Interestingly, only at high cellopentaose substrate concentration does the Lineweaver-Burk plot show an upward deflection of the curve (see Appendix 1-D). The upward deflection in Lineweaver-Burk plots for cellopentaose hydrolysis at high substrate concentration could be due to the excess of substrate causing enzyme inhibition. In this case, the substrate may bind at a different position away from the catalytic site with no reaction occuring preventing proper binding of another substrate molecule. Another explanation could be because of the transglycosylation reaction that could take place under those conditions as previously observed (Gusakov *et al.,* 1984) where cellopentaose could become the acceptor molecule generating C6 and C4. Interestingly, the synthesis of inducer molecules are believed to require transglycosylation activity by enzymes that are associated with the membranes generating molecules which are capable of inducing for example, the lactose operon (Jobe and Bourgeois, 1972) and cellulase complexes (Nisizawa *et al.,* 1971; Vaheri *et al.,* 1979).

3.7.4. The amino acid composition of CbgA and various enzymes

The amino acid composition of the purified enzyme CbgA and related enzymes is given in Table IX. No major differences are evident.

Table IX. Amino acid composition of various enzymes

	Molar ratio[a]					
A.A.	CbgA	Abg[1]	Cex[2]	CenA[2]	CenB[3]	CenC[4]
asx	0.73	1.00	0.83	0.86	0.71	0.68
glx	0.76	0.74	0.62	0.67	0.53	0.94
ser	0.71	0.35	0.50	0.59	0.66	0.60
gly	1.36	0.89	0.70	1.10	0.84	1.00
his	0.11	0.31	0.08	0.06	0.11	0.13
arg	0.33	0.43	0.27	0.35	0.19	0.22
thr	0.53	0.35	0.72	0.90	1.03	0.59
ala	1.00	1.00	1.00	1.00	1.00	1.00
pro	0.16	0.54	0.47	0.67	0.60	0.65
tyr	0.12	0.39	0.17	0.24	0.39	0.33
val	0.60	0.59	0.57	0.51	0.65	0.67
met	0.04	0.28	0.08	0.04	0.02	0.05
cys	0.11	0.11	0.10	0.12	0.06	-
ile	0.48	0.26	0.13	0.20	0.11	0.12
leu	0.13	0.67	0.37	0.49	0.52	0.68
phe	0.11	0.39	0.35	0.18	0.23	0.22
lys	0.17	0.35	0.28	0.27	0.35	0.15
trp	-	0.24	0.20	0.33	0.19	-
#A.A.	1634	458	443	418	1065	1161
M_r (kDa)	183	51	49	52	110	120

1, Wakarchuk, 1987; 2, Langsford, 1988; 3, Owolabi, 1988;
4, Moser, 1988.

a, The amino acid composition is given on a Mol % basis relative to alanine residues equal to 1.00.

(-) , Not available

#A.A., Number of residues per enzyme, approximate for CbgA

3.8. The identification of the native *C. fimi* ß-glucosidase A.

To be able to compare the size, cellular location and enzymatic activity of the recombinant enzyme with the corresponding native enzyme, *C. fimi* ß-glucosidases were partially purified and characterized. Proteins from a *C. fimi* CM-Cellulose grown culture were fractionated into cytoplasmic, membrane associated and secreted protein fractions.

3.8.1. The effect of carbon source on ß-glucosidase activity

During the study of cellulose hydrolysis by *C. fimi*, several enzymes with ß-glucosidase activity have been detected. The enzyme activity is always observed independently of the carbon source suggesting that in *C. fimi*, ß-glucosidase activity is constitutive. Nevertheless, the total pNPGase activity detected in *C. fimi* cultures varies with the carbon source used (Fig. 19). An increase of 2.4 fold in total ß-glucosidase activity is observed when cells are grown in rich medium and a 3.7 fold increase is observed in minimal medium containing CM-cellulose. ß-glucosidase activity was found mostly in the cytosol fraction of *C. fimi* (from 65% up to 100% depending on the carbon source) but some activity was detected in the membrane fractions and in the culture supernatants, which suggested a multiple location of ß-glucosidases in this organism (data not shown).

3.8.2. The fractionation of *C. fimi* ß-glucosidases

About 80% of the pNPGase activity in a culture grown on basal medium containing 1% CMC was cell associated. About 66% of the cell-associated activity was in the cytoplasm and about 13% was membrane-associated. About 16% of the activity in cultures was in the culture supernatant and 21% of the unrecovered activity was attributed to experimental loss, most of it still present in the discarded pellet of cell debris which was not processed.

The ß-glucosidases in the cytoplasm of CMC-grown cells were first fractionated by anion-exchange chromatography on a DEAE-Sephadex A-50-120 column (Fig. 20). ß-Glucosidase activities were separated into 3 pools. Pool I (fractions 27 to 32) contained both pNPGase and

Fig. I9. Total ß-glucosidase (pNPGase) activity in *C. fimi* cultures grown on various carbon sources. *C. fimi* cells were grown at 30°C in 100 mL of media containing 1% of the various carbon sources to late log phase. Each Culture was tested for total ß-glucosidase activity using pNPG as substrate. Values are given in total miliunits (mU) of enzyme per culture per $O.D._{600nm}$ where one unit releases one µmol of pNP per min at 37°C.

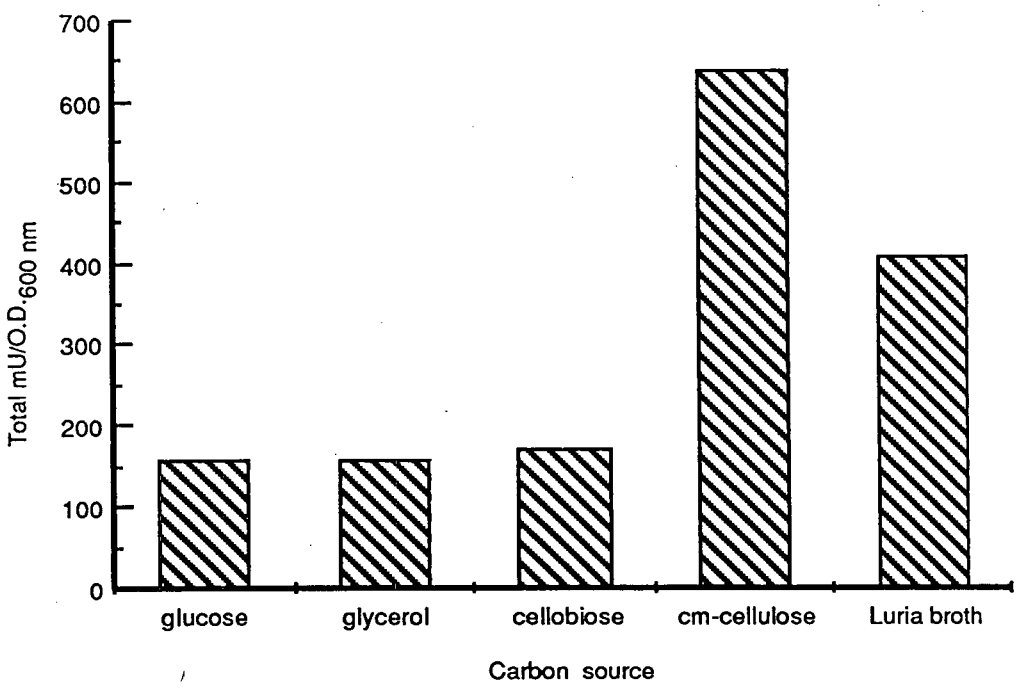
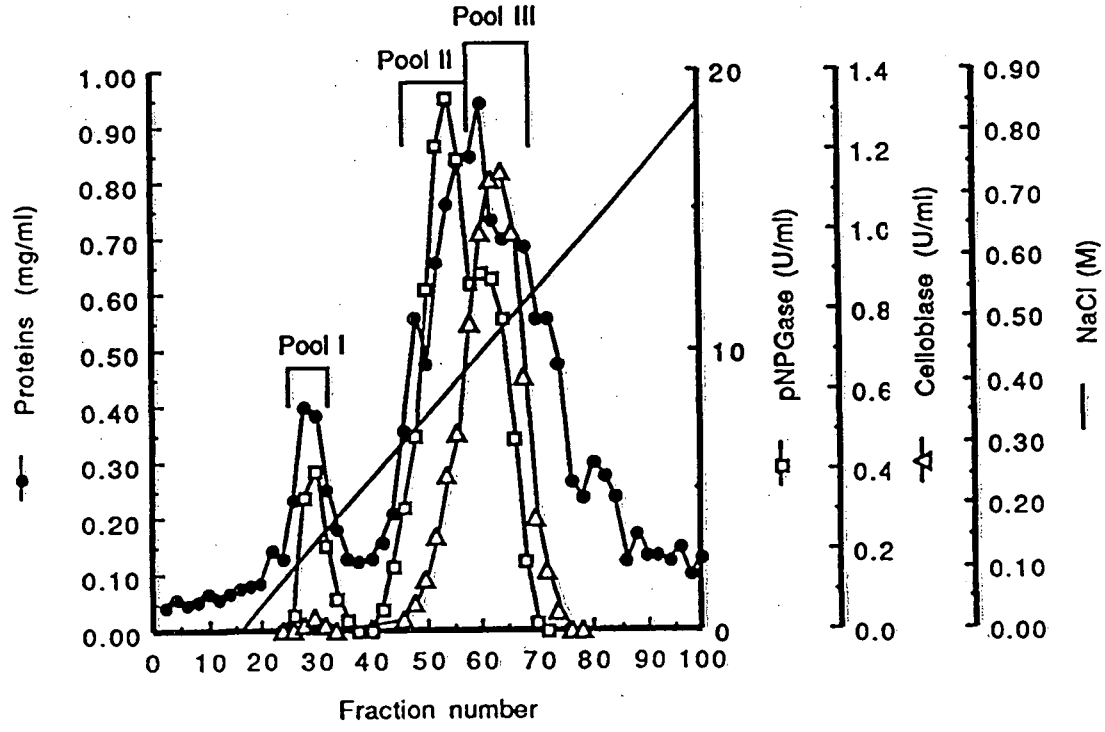


Fig. 20. Fractionation of *C. fimi* cytoplasmic proteins on a DEAE-Sephadex A-50-120 column. *C. fimi* cytoplasmic protein fraction obtained from cells grown in basal medium containing 1%-CM-cellulose was applied to a DEAE-Sephadex A-50-120 column (2.5 cm x 11 cm). After washing with 50 mM Tris-5 mM EDTA (pH 7.0), bound proteins were eluted with a linear salt gradient (0 - 0.85 M NaCl). Fractions were tested for protein concentration, pNPGase and cellobiase activities (legend on axis). Fractions 27 to 32 (Pool I), 44 to 57 (Pool II) and 58 to 69 (Pool III) were pooled separately. The proteins in Pool II and III were fractionated further on a Mono Q column.



cellobiase activities; Pool II (fractions 44 to 57) contained mostly pNPGase activity; and Pool III (fractions 58 to 69) contained mostly cellobiase activity with a small shoulder of pNPGase activity.

The proteins concentrated from Pools II and III were purified further on a Mono Q column (Fig. 21 and 22) because it was beleived that one of those fractions may contain the native CbgA enzyme and that Pool I with its low amount of activity was not to be considered important. A single peak of pNPGase activity was obtained from Pool II (Fig 21-fraction 14 and 15) and fractions 14 and 15 were combined. Three major peaks of pNPGase activity were obtained from Pool III (Fig. 22- fraction 19 and 20, 32 and 33 and 39 to 41) and these fractions were combined, respectively. There was also a major peak with activity on cellobiose but not on pNPG (Fig. 22-fraction 34 to 37) and these active fractions were also combined. The combined active fractions 14-15, 32-33, 34-37 and 39-41 were then analysed further by gel electrophoresis for their content.

3.8.3. The detection of ß-glucosidase activity after gel electrophoresis

The proteins from the various active fractions were resuspended in SDS-PAGE loading buffer containing ß-mercaptoethanol and were kept at 25°C prior to loading on gels. The proteins in those samples are beleived to have migrated in the polyacrylamide gel according to their globular size since all samples were not fully denatured by heat treatment. The rationale for not fully denaturing the proteins prior to the electrophoresis comes from an early observation that showed the absence of migration of CbgA during non-denaturing gel electrophoresis where SDS is omitted (not shown). ß-Glucosidase activity from those cells had previously been detected at the junction of the stacking and separating gels but was not discussed further because active bands were detected in the separating gels (Wakarchuk, personal communication). The usual explanation for absence of appropriate migration of a protein under non-denaturing conditions is often attributed to its isoelectric point which could be located above 7.0 preventing a net negative charge to the peptide (Jovin *et al.*, 1964; Sigma, technical bulletin no. NKR-137). In the case of CbgA, aggregation or multimerization of the enzyme could have precluded its migration in the gel.

Fig. 2I. Fractionation of Pool II on Mono Q. Pool II from the DEAE-sephadex column was applied to a Mono Q anion-exchange column. After washing the column with 50 mM Tris-5 mM EDTA (pH 7.0), bound proteins were eluted by a linear increase in the NaCl concentration as shown. Fractions were analysed for protein and pNPGase activity. Fractions 14 and 15 were pooled.

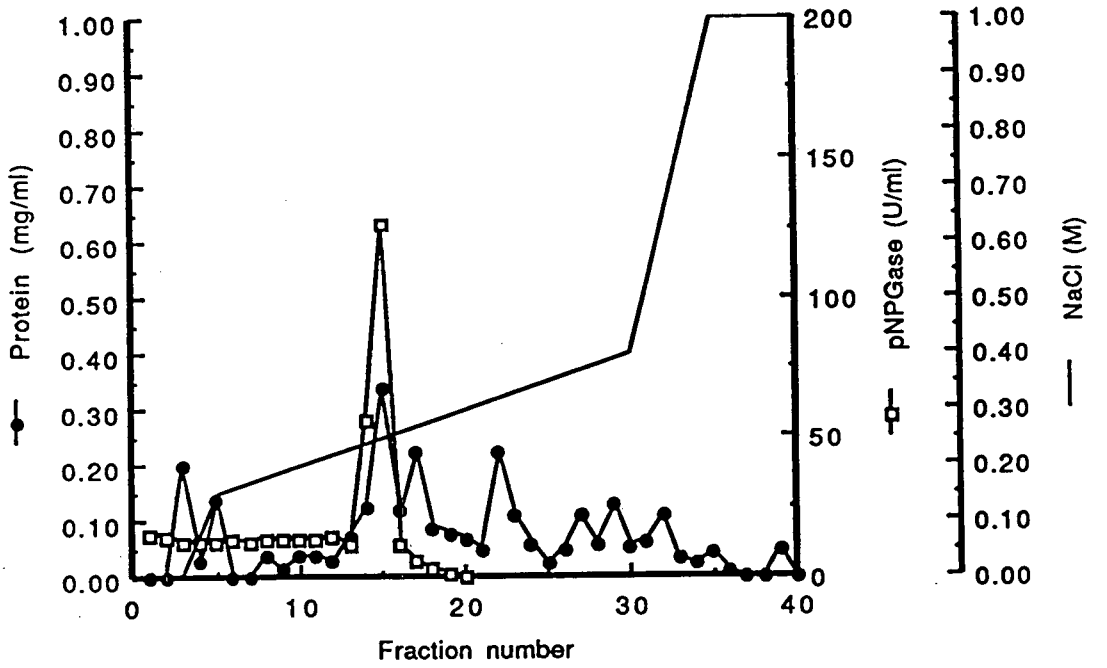
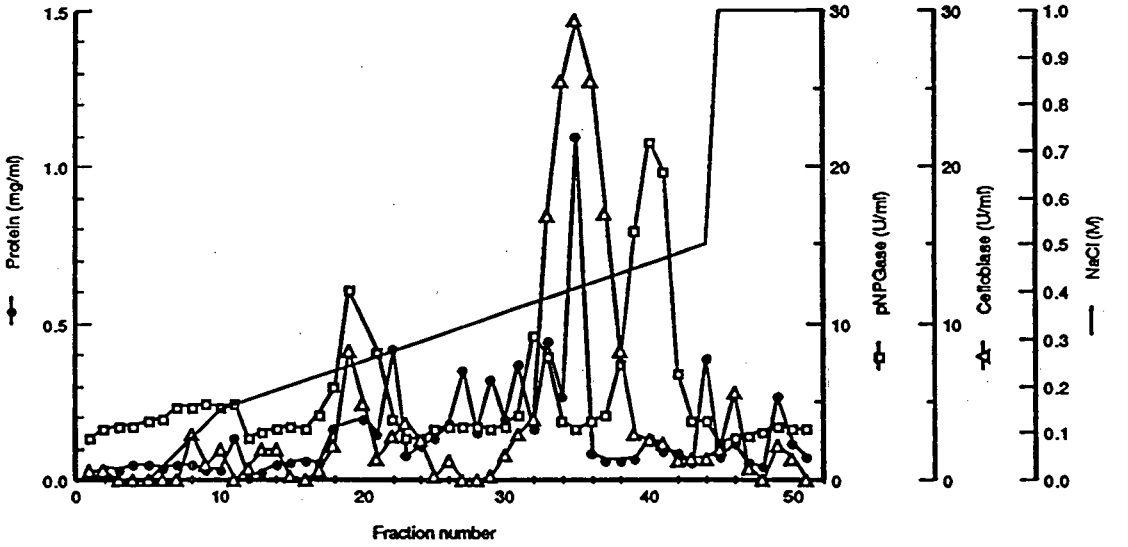


Fig. 22. Fractionation of Pool III on Mono Q. Pool III fraction from the DEAE-sephadex column was applied to a Mono Q anion-exchange column. After washing the column with 50 mM Tris-5 mM EDTA (pH 7.0), bound proteins were eluted by linear increase in the NaCl concentration as shown. Fractions were analysed for protein, and for cellobiase and pNPGase activities. Fractions 19 and 20, 32 and 33, 34 to 37 and 39 to 41 were pooled seperately.



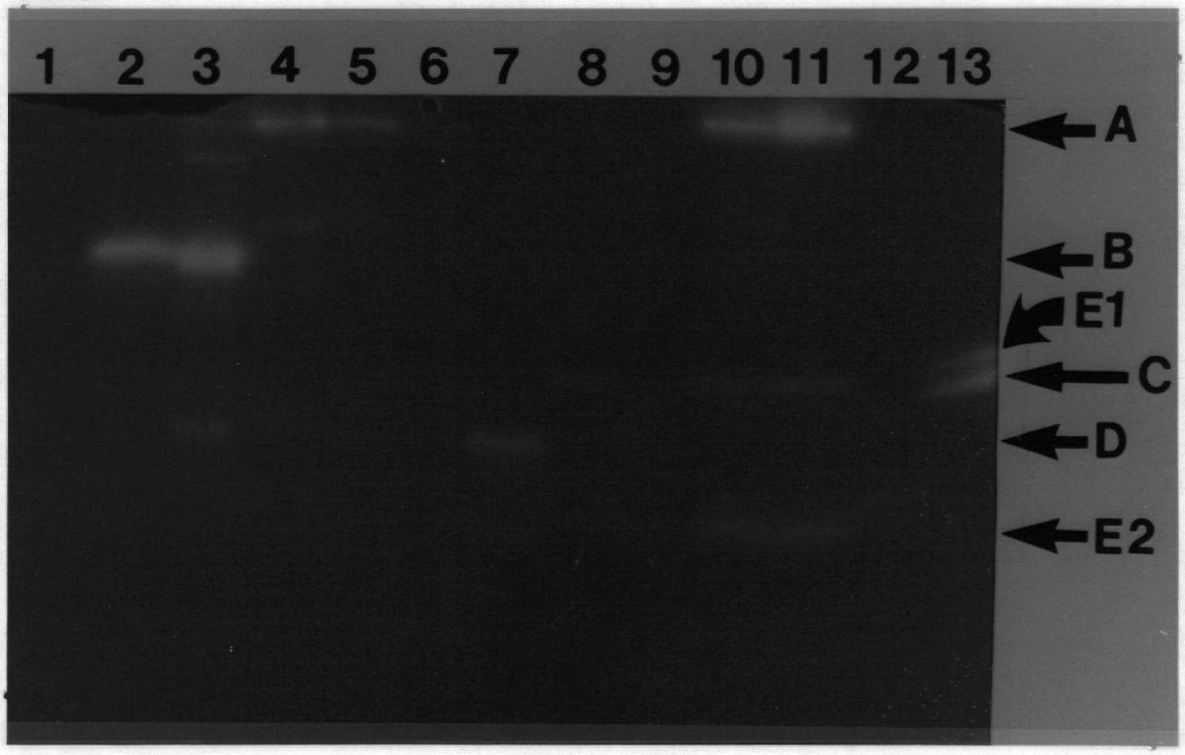
Furthermore, heat denaturation of CbgA and other proteins (see Lacks and Springhorn, 1980) sometimes completely inactivates the enzyme. Activity can not be restored after removal of SDS (data not shown). The negatively charged SDS present in the gel and sample buffer which may have bound partially to the protein allowed CbgA to enter the stacking and separating gels and MUGase activity to be observed after removal of the SDS.

Various protein fractions from *E. coli* and *C. fimi* cultures were analysed by zymogram (Fig. 23). Total cell extract from *E. coli*/pUC13 lacked MUGase activity (lane 1), that from *E. coli*/pUC13:62Δ31 gave a major active band at position B (lane 2) which comigrated with the major active component in the purified enzyme (lane 3). The purified enzyme contained several active components with a fast migrating band at position D. All these active components were found to react with the CbgA-directed antiserum (Fig. 24-lane 3). SDS-PAGE of the fully denatured purified enzyme, however, gave a unique band stained with Coomassie blue or reacting with the antibodies (Fig 16, lane 7; Fig 25, lane 1).

The *C. fimi* enzyme(s) behave similarly (Fig. 23, lane 4 to 13). The *C. fimi* culture supernatant contains an apparent aggregate (lane 4) which also contains multiple components. A major active band (A) was located at the junction of the stacking and separating gels. The *C. fimi* membranes (lane 5) also gave an active band (A) located at the top of the separating gel. Due to the low concentration of ß-glucosidase activity present in the *C. fimi* cytoplasm (lane 6), no MUGase activity was detectable. Nevertheless, *C. fimi* cytoplasm which was further fractionated into more concentrated and active fractions (see above), is shown in lanes 7 to 13 (Fig. 23).

The Pool I fraction (lane 7) gave an active band at position D which comigrated with the fastest migrating band of the purified recombinant enzyme. The band at position D from Pool I also reacted with the antiserum (Fig. 24-lane 7). The native enzyme present in Pool I could correspond to the cloned enzyme and was named CbgA. I wish to emphasize that at this point, it is an early suggestion to relate CbgA to the enzyme in Pool I as more evidence will be shown later in the thesis.

Fig. 23. Zymogram of partially denatured protein samples using MUG as a substrate. By mixing various protein samples with loading buffer, non-fully denatured protein samples from *E. coli* (lanes 1 to 3) and *C. fimi* (lanes 4 to 13) were analysed by 10% SDS-PAGE. Active bands were visualised by soaking the gel in 1 mM MUG -1%(v/v) Triton X-100 -100 mM phosphate buffer (pH 6.6). Lane 1, 40 μg of *E. coli* JM83/pUC13 total cell extract; lane 2, 40 μg of *E. coli* JM83/pUC13:62Δ31 total cell extract; lane 3, 188 ng of purified recombinant CbgA. Lane 4, 19 μg of *C. fimi* culture supernatant protein aggregate fraction; lane 5, 30 μg of *C. fimi* membrane fraction; lane 6, 40 μg of *C. fimi* cytosol fraction. Lane 7, 10 μg of *C. fimi* Pool I fraction; lane 8, 37 μg of *C. fimi* Pool II fraction; lane 9, 10 μg of *C. fimi* fraction 14 and 15 from Mono Q of Pool II; lane 10, 40 μg of *C. fimi* Pool III fraction; lane 11, 21 μg of *C. fimi* fraction 32 and 33 from Mono Q of Pool III; lane 12, 11 μg of *C. fimi* fraction 34 to 37 from Mono Q of Pool III; lane 13, 7.7 μg of *C. fimi* fraction 39 to 41 from Mono Q of Pool III. After 30 min incubation at 30°C, active bands were visualized by U.V. illumination and picture was taken. Letters A, B, C, D, E1 and E2 on the left indicate the locations of active bands.



The Pool II fraction (lane 8) gave very weak MUGase activity with a band located at position E1. The Pool II (lane 8), fraction 14-15 (lane 9) and fraction 19-20 (not shown) seem to lack MUGase activity since appreciable amount of pNPGase activity was present in those lanes. This suggests that MUGase activity from Pool II may originate from contaminating activity from Pool III (see Fig. 20) and that fractions 14-15 and 19-20 could contain identical enzymes. The Pool III fraction (lane 10) contained multiple components with 3 major active bands detected at position A, E1, and E2. Similar components are detected in the lane containing fraction 32-33 (lane 11). The fraction 39-41 (lane 13) showed active bands at position E1 and uniquely at position C.

The enzyme from Pool II (Fig. 21) called CbgB1 is responsible for the major peak of pNPGase activity from *C. fimi* cytoplasm (Fig. 20) and is anticipated to have contaminating amount of activity in Pool III. The contaminating activity could correspond to fraction 19 to 20 from Pool III (Fig. 22) and this enzyme was called CbgB2. This suggests that CbgB1 and CbgB2 enzymes originating from different Pools (Pool II and III, respectively), may contain identical enzymes and could be named CbgB. The major active bands detected at the junction of the gels (A) could originate from aggregation of proteins. The other enzymes from fractions 32 to 33 and 39 to 41 of Pool III (Fig. 22) could also be identical enzymes that are isomers and were called CbgD1 and CbgD2, respectively. Nevertheless, this could not be confirmed by the data obtained so far. The fraction 34-37 with strong cellobiase activity did not hydrolyze MUG efficiently (lane 12) and was called CbgC. The weak MUGase activity from that sample which was detected above position E2 is also detected in Pool III (lane 10) and in sample containing CbgD1 (lane 11). This suggests that CbgC is not responsible for the appearance of this MUGase active band and, consequently, CbgC may not have strong affinity for MUG or for pNPG.

To be able to compare protein migration patterns and determine the sizes of the active peptides, various samples were analysed by SDS-PAGE after heat denaturation then stained with Coomassie blue in parallel with the corresponding non heat-denatured samples analysed by zymogram and Coomassie blue (data not shown). The CbgB1 and

CbgB2 samples stained with Coomassie blue contained several peptides and had identical protein migration patterns. These samples contained a major peptide with an Mr of 57 kDa. Nevertheless, the CbgB enzymes could not be identified with certainty because these enzymes were only partially purified and could not hydrolyze MUG when analysed by zymogram. The sample containing CbgC which lacked MUGase activity was stained with Coomassie blue and showed a unique band with an Mr of 88 kDa. From identical migration patterns between denatured and non-denatured samples containing either CbgD1 or CbgD2 enzymes, Mr were assigned to the active peptides. The CbgD1 enzyme has an Mr of 35 kDa. The CbgD2 enzyme showed two active bands comigrating with denatured peptides with Mr of 45 and 47 kDa. The Pool I fractions containing CbgA enzyme was not analyzed under those conditions.

The Cbg enzymes were quantitatively analysed for their ability to hydrolyse pNPG and cellobiose (Table X). The CbgA enzyme could hydrolyse pNPG. The CbgB enzymes could hydrolyze pNPG and cellobiose. The CbgC enzyme could only hydrolyze cellobiose. The CbgD enzymes could hydrolyze pNPG but only CbgD1 had some activity on cellobiose.

3.8.4. Summary of *C. fimi* ß-glucosidases

A summary of the characteristics obtained so far on *C. fimi* ß-glucosidases indicates the various differences between those enzymes and the relationship between the recombinant CbgA and its corresponding form in *C. fimi*. The *C. fimi* cytoplasmic fraction shows heterogeneity in its ß-glucosidases and so far, 4 different enzymes have been partially purified by ion exchange chromatography. Based on their enzymatic activities, a summary of the properties of these enzymes is given in table X.

The CbgB enzymes have the highest specific activity on pNPG whereas the CbgC enzyme had the highest specific activity on cellobiose. Consequently, most of the pNPGase activity of *C. fimi* seems to be due to CbgB, whereas, most of the cellobiase activity seems to be due to CbgC. The CbgD enzymes were able to hydrolyse

Table X. Characteristics of *C. fimi* ß-glucosidases

Fractions	Enzyme	Specific activity (U/mg prot.)		MUGase	M_r (kDa)
		pNPGase[1]	Cellobiase[2]		
Pool I	CbgA	47	0.3	+	183(a)
Pool II		250	n.d.	-	
14 and 15	CbgB1	450	5.5	-	57(b)
Pool III		50	n.d.	+	
19 and 20	CbgB2	146	2.1	-	57(b)
32 and 33	CbgD1	52	5.3	+	35
34 to 37	CbgC	<1	218	+/-	88
39 to 41	CbgD2	122	<0.1	+	45-47

The fractions in this table are from *C. fimi* culture grown on CMC basal medium and were obtained as described in section 2.6.2.
Pool I, II and III fractions are from DEAE-sephadex column (Fig. 20).
Fraction 14 and 15 is from Mono Q column of Pool II (Fig. 21).
Fractions 19 and 20, 32 and 33, 34 to 37 and 39 to 41 are from Mono Q column of Pool III (Fig. 22).

1. One unit of enzyme releases 1 µmol of pNP per minute at 37°C.
2. One unit of enzyme releases 1 µmol of glucose per minute at 37°C.

+, -, presence or absence of MUGase activity from zymogram
(a), M_r of larger peptide from Western blot analysis
(b), M_r after SDS-PAGE of major peptide in that sample
n.d. not determined

pNPG but not as efficiently as the CbgB enzymes. The activity of CbgD1 on cellobiose could originate from contaminating CbgC activity since CbgD1 originates from fraction 32 to 33 and CbgC from fraction 34 to 37 (Fig. 22), fraction 33 containing appreciable amount of cellobiase activity. Eventhough a definite statement should not be made until further purification is achieved, the Cbg enzymes appear to be distinctive from each other in their relative substrate specificity.

The presence in bacteria of more than one ß-glucosidase gene has been reported antecedently (Gokhale and Deobagkar, 1989). Multiple ß-glucosidase components are common in microorganisms and may possibly result from (i) complex formation between enzymes and polysaccharides; (ii) dissociation of subunits during experimentation; (iii) proteolytic cleavage of the peptide; (iv) multiple enzymes with different functions required for growth of the organism (Shewale, 1982). A preliminary study has shown that *C. fimi* contains at least two distinct ß-glucosidases (Wakarchuk *et al.,* 1984). More recently, Kim and Pack (1989) have reported the molecular cloning in *E. coli* of two ß-glucosidase genes from *C. fimi.* One of them encodes an aryl-ß-glucosidase that hydrolyzes only pNPG but not cellobiose and may well correspond to the CbgD enzymes. The other gene encodes a "true" cellobiase that hydrolyzes both cellobiose and pNPG, and may well correspond to the CbgB enzymes. In *C. fimi*, CbgB is possibly active in cleaving phenolic glucosides of plant origin that may be liberated by plants following the attack by this cellulolytic bacterium. In addition, a third enzyme, called CbgC, has been isolated and was shown to have high activity on cellobiose. It is noteworthy to mention that the *C. fimi* genes were isolated by screening a genomic bank with pNPG as a substrate making unlikely the molecular cloning of the true cellobiase, CbgC, which does not hydrolyze this substrate. It appears that CbgC is the *C. fimi* ß-glucoside glucohydrolase or "true" cellobiase that is involved in the hydrolysis of cellobiose released from the hydrolysis of cellulose by *C. fimi* cellulases. Finally, more studies are required to identify the corresponding native CbgA that could be present in the cytoplasmic fraction Pool I and/or in the secreted fraction as part of the protein aggregate.

3.9. Antiserum and Western blot analysis

Various protein samples were obtained from *E. coli* and *C. fimi* where the relationship between the recombinant enzyme CbgA and the native enzyme remained to be established on another basis than simply enzymatic activity. A highly specific antiserum to CbgA was necessary for the detection of the recombinant forms of the enzyme within a mixture of bacterial proteins. Also, the detection of the native forms of CbgA would give some interesting information regarding its cellular location. Finally, a stronger relation could be established between the previously characterized recombinant CbgA and its corresponding forms in *C. fimi.*

3.9.1. Generation of rabbit polyclonal antibodies.

For the detection of small quantities of antigenic peptides through the powerful technique of Western blot analysis, polyclonal antibodies to purified recombinant enzyme CbgA were raised in New Zealand white rabbit (section 2.8). An ELISA on protein fractions from *E. coli*/pUC13:62 gave a titre of 10^{-4} for the antiserum with virtually no background for the *E. coli*/pUC13 cell extracts (Fig. 5A). The pre-immune rabbit serum also shows zero cross reactivity with the protein extracts from *E. coli*/pUC13:62 or with the purified enzyme CbgA. This indicates that the antiserum is highly specific for CbgA from *E. coli* and that it can be used at the appropriate dilution of 10^{-4} to detect antigenic peptides from SDS-PAGE. An ELISA on protein fractions from *C. fimi* culture supernatant (Fig. 5B) also gave a titre of 10^{-4} for the antiserum when compared to the rabbit pre-immune serum. Interestingly, both the *C. fimi* culture supernatant fraction, containing the protein aggregate, and also the membrane fraction from ultracentrifugation of crude protein extract gave a positive ELISA response. This result suggests that CbgA in *C. fimi* could be secreted in the culture medium. It isn't clear if the native CbgA from the *C. fimi* membrane fraction is physically attached to the membranes or, as for the recombinant enzyme detected in the *E. coli* membrane fraction, is present as aggregates which sediment at high speed centrifugation. The rabbit antiserum was used to analyse various protein fractions by Western blot.

3.9.2. Western blot analysis of the activity gel.

The proteins present in the MUGase activity gel were blotted onto NC membrane and further analysed by Western blot (Fig. 24). The negative control containing *E. coli*/pUC13 cell extract (lane 1) shows no reactivity with the antiserum whereas all the enzymatically active components of *E. coli*/pUC13:62Δ31 cell extract (lane 2) or of the purified recombinant enzyme (lane 3) reacted with the antibodies. It is appropriate to mention that once boiled at 100°C for 2 min, the purified enzyme shows only a single reactive band (Fig. 25-lane 1). This suggests that the ladder of active bands in fig. 23 lane 3 could represent a set of partially denatured CbgA or aggregating fragments forming multimeric forms.

This multiple banding phenomenon was also observed with the proteins in the *C. fimi* culture supernatant aggregate (Fig. 23-lane 4) which could also contain various denatured or aggregating forms of CbgA. All the MUGase active components in that fraction reacted with the antibodies. It is also appropriate to mention that once boiled at 100°C for 2 min, the *C. fimi* protein aggregate shows only 3 reactive bands (Fig. 25-lane 8). The nature of the aggregation is unknown but ß-glucosidase enzymes have been reported to form multimers in other organisms (Coughlan, 1985) and could be responsible for these observations . The active band in Pool I reacts with the antibodies and the kinetic characteristics of the Pool I enzyme are very similar to those previously reported for the recombinant CbgA enzyme. This indicates that the purified recombinant enzyme CbgA corresponds to the native enzyme present in Pool I from *C. fimi* cytoplasm. In the other lanes (Fig. 24-lane 8 to 13) none of the MUGase active components reacted with the antibodies indicating a clear difference between CbgA and the other ß-glucosidases characterized so far.

3.9.3. Western blot analysis of denatured protein samples.

Western blot analysis using antiserum raised against the recombinant enzyme CbgA was used to detect specific peptides in fully denatured fractions separated by SDS-PAGE (Fig. 25). The purified recombinant CbgA (lane 1) shows a single band of about 183 kDa. The negative

Fig. 24. Western blot analysis of the zymogram. Proteins separated by electrophoresis were blotted onto NC filter and analysed by Western blot using anti-serum raised against purified recombinant CbgA. Lane 1, 40 µg of *E. coli*JM83/pUC13 total cell extract; lane 2, 40 µg of *E. coli* JM83/pUC13:62Δ31 total cell extract; lane 3, 188 ng of purified recombinant CbgA. Lane 4, 19 µg of *C. fimi* culture supernatant protein aggregate fraction; lane 5, 30 µg of *C. fimi* membrane fraction; lane 6, 40 µg of *C. fimi* cytosol fraction. Lane 7, 10 µg of *C. fimi* Pool I fraction; lane 8, 37 µg of *C. fimi* Pool II fraction; lane 9, 10 µg of *C. fimi* fraction 19 and 20 from Mono Q of Pool II; lane 10, 40 µg of *C. fimi* Pool III fraction; lane 11, 21 µg of *C. fimi* fraction 32 and 33 from Mono Q of Pool III; lane 12, 11 µg of *C. fimi* fraction 34 to 37 from Mono Q of Pool III; lane 13, 7.7 µg of *C. fimi* fraction 39 to 41 from Mono Q of Pool III.

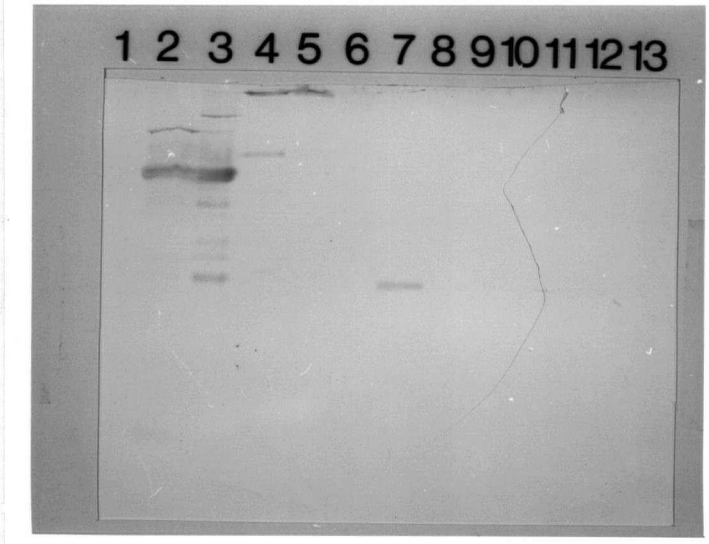
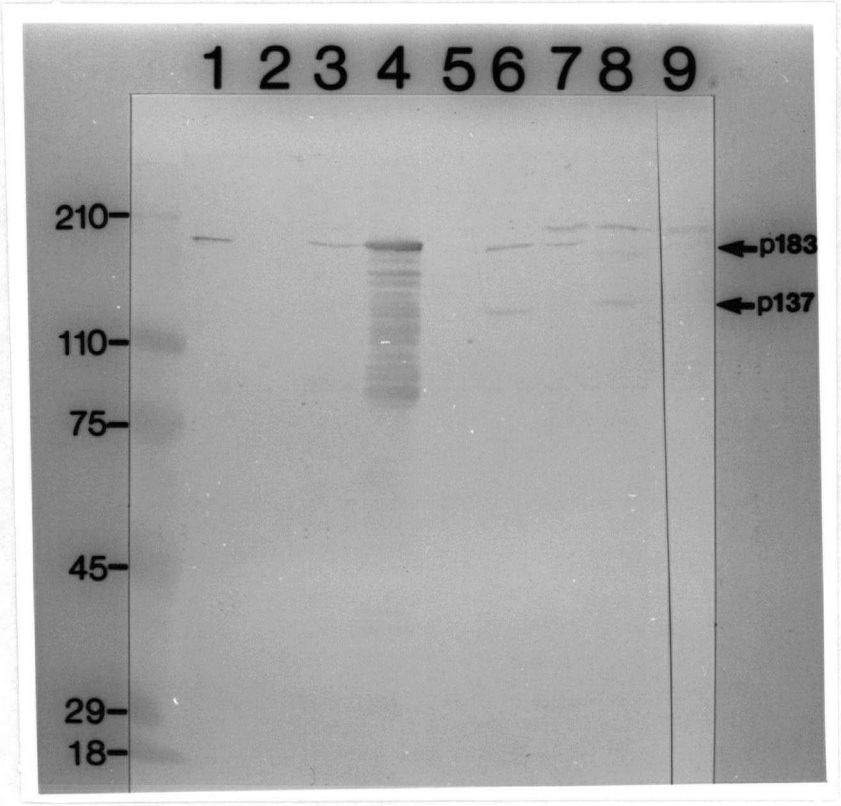


Fig. 25. Western blot analysis of the recombinant and native enzyme CbgA. Various protein samples were fractionated by 7% SDS-PAGE after denaturation at 100°C for 2 min. Samples were blotted onto NC filter prior to detection of proteins with antiserum raised against the purified recombinant CbgA. Lanes 1 to 4 contain *E. coli* proteins and lanes 5 to 9 contain *C. fimi* proteins. Lane 1, 50 ng of purified recombinant CbgA; lane 2, 30 µg of *E. coli* JM83/pUC13 total cell extract; lane 3, 30 µg of *E. coli*JM83/pUC13:62 total cell extract; lane 4, 30 µg of *E. coli* JM83/pUC13:62Δ31 total cell extract. Lane 5, 30 µg of *C. fimi* cytosol fraction; lane 6, 5 µg of *C. fimi* Pool I fraction; lane 7, 30 µg of *C. fimi* membrane fraction; lane 8, 9.5 µg of *C. fimi* culture supernatant protein aggregate fraction; lane 9, 50 µg of *C. fimi* cytosol (100X) from LBIs grown culture. Numbers on the left indicate Mr markers (kDa).

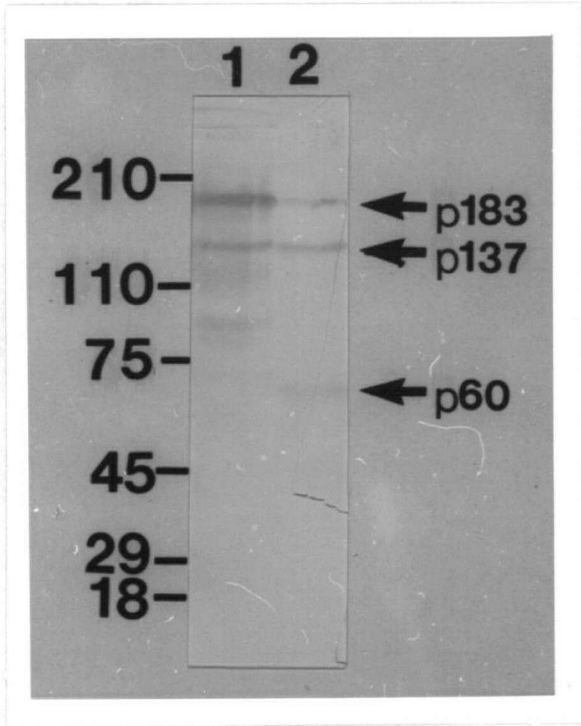


control *E. coli*/pUC13 (lane 2) shows no reactivity with the antibodies whereas proteins from *E. coli*/pUC13:62 (lane 3) contains a p183 band as well as very weak bands of about 197 and 210 kDa in size. When CbgA is over expressed in *E. coli* cells carrying pUC13:62Δ31 plasmid proteolysis of the recombinant enzyme is evident (lane 4). The major peptide in that lane comigrates with the purified enzyme in lane 1. One would expect a larger peptide from *E. coli* cells carrying pUC13:62 plasmid which contains the entire 7.2 Kb fragment (lane 3). A number of products, including a processed p183 product, are present. This suggests that the size of the major peptide expressed from pUC13:62Δ31 roughly correlates with the size of the processed product produced by *E. coli* carrying pUC13:62 plasmid. The intensity of the immunogenic protein bands is comparable to the specific activities of the samples. A more accurate study of the proteins made from the various deletions will be discussed later.

Specific bands could be detected in fractions of *C. fimi* cells grown on CMC (Fig. 25). In the cytoplasmic fraction of CMC-grown cells (lane 5) and LBIs-CMC grown cells (lane 9), two bands of low intensity with molecular weights of 197 and 183 kDa were only visible on the original filter. Partially purified *C. fimi* cytosol fraction (lane 6) contains p183 and p137 peptides and a weakly reacting p60 peptide (see below). The *C. fimi* fraction enriched for membranes by ultracentrifugation of cleared broken cell lysate shows the presence of two polypeptides with M_r of 197 and 183 kDa (lane 7). Finally, the *C. fimi* protein aggregate isolated from its culture supernatant shows the presence of several polypeptides (lane 8) with M_r of 197, 170 and 137 kDa in sizes.

The p60 polypeptide visible as a weak band on the original Western blot (and not visible on the photograph Fig. 25) is clearly visible in Fig. 26 (lane 2). Lane 1 (Fig. 26) which corresponds to a FPLC purified fraction of *E. coli* proteins (also Fig. 16-lane 10) contains a p183 and a p137 polypeptide but does not contain the p60 polypeptide. This p60 polypeptide is evident in the *C. fimi* pool I fraction (lane 2). For comparisons, this figure shows the p183 and p137 polypeptides from different origins and suggests that specific proteolysis of CbgA occurs in a similar way in both organisms. Nevertheless, only in the lane

Fig. 26. Western blot analysis of *E. coli* and *C. fimi* proteins. Proteins from the fractions containing the purified p183, p137 and p60 polypeptides from *E. coli* (Fig. 15-fraction 28 to 31) and *C. fimi* Pool I (Fig. 20-fraction 27 to 32) were fully denatured, fractionated by 7% SDS-PAGE and blotted onto a NC filter. These fractions were then analysed by Western blotting using CbgA anti-serum. Lane 1, 13 µg (114 U) of protein from *E. coli* p183, p137 and p60 peptides. Lane 2, 5 µg (1 U) of protein from *C. fimi* Pool I fraction. The numbers on the left indicate Mr markers (kDa).



containing *C. fimi* proteins is the p60 polypeptide detectable with the antibodies. The absence of a p60 reacting polypeptide in lane 1 suggests that in *E. coli* (Fig. 16-lane 10), the FPLC purified p60 polypeptide may not correspond to the p60 polypeptide detected from *C. fimi*. The native enzyme in *C. fimi* has been identified, and like the recombinant enzyme CbgA, it could be cleaved into smaller polypeptides presumably by proteolysis. The presence of the native enzyme in the culture supernatant of CMC-grown *C. fimi* cells suggests again that CbgA is probably secreted by this organism into the culture medium to facilitate hydrolysis of cello-oligosaccharides to glucose.

3.9.4. Western blot analysis of proteins from various clones.

Various ß-glucosidase expressing clones obtained from deletions of 5' end DNA regions were analysed by Western blot (Fig. 27). Plasmid DNA was digested with *Pst*I restriction enzyme and restriction fragments were fractionated by agarose gel electrophoresis. The decrease in size of the major peptides detected by Western blot analysis (Fig. 27) correlates roughly with the decrease in size of the DNA in Δ17, Δ31, Δ21-4 and Δ21-15 clones (Fig. 28). With the exception of clone Δ27, all peptides show a gradual decrease in size with decrease in length of the DNA coding sequences. A peptide was detected from *C. fimi* cytoplasmic proteins purified by affinity chromatography on an antibody column (lane 11) and could comigrate with a peptide from Δ27 clone (Fig. 27-lane 3).

Proteolysis is likely responsible for the appearance of lower-Mr species. A peptide of approximately 137 kDa in size is detected in most samples suggesting that proteolysis of CbgA in *E. coli* could occur at a specific location downstream from the Δ21-15 fusion point. This peptide may correspond to the previously isolated p137 polypeptide originating from processing of CbgA by *E. coli*. Therefore, up to 2 Kb of DNA can be deleted and still produce an active polypeptide suggesting that the CbgA catalytic domain is located in the carboxy terminal two third of the protein requiring less than 3.4 Kb of DNA coding sequence.

Fig. 27. Western blot analysis of protein extracts from ß-glucosidase deletion clones expressed in *E. coli* . Various denatured protein samples from *E. coli* cells carrying plasmids encoding ß-glucosidase were fractionated by 7% SDS-PAGE and analysed by Western blotting using CbgA anti-serum. Lanes 1 to 7, 30 µg of proteins from *E. coli* total cells extracts carrying : 1, pUC13; 2, pUC13:62; 3, pUC13:62Δ27; 4, pUC13:62Δ17; 5, pUC13:62Δ31; 6, pUC13:62Δ21-4; 7, pUC13:62Δ21-15 plasmids. Lane 8, 0.7 µg (2.6 U) of purified CbgA; lane 9, 2.5 µg (23 U) of p183-p137 polypeptides; lane 10, 5.5 µg (21 U) of p183-p137-p60 polypeptides. Lane 11 contains 0.7 µg of protein from an antibody affinity purified fraction of the cytosol of LBIs-CMC grown *C. fimi* cells. Prestained M_r markers (kDa) are shown and the location of p183 and p137 is indicated.

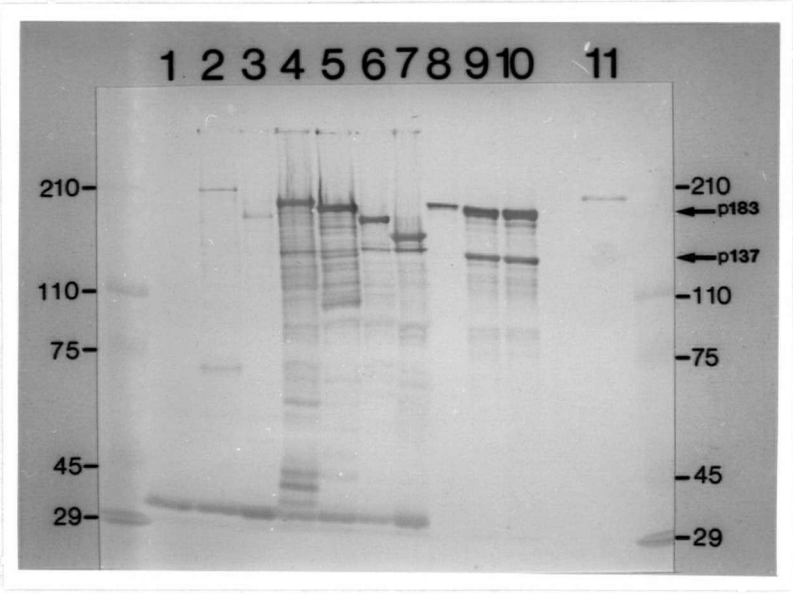
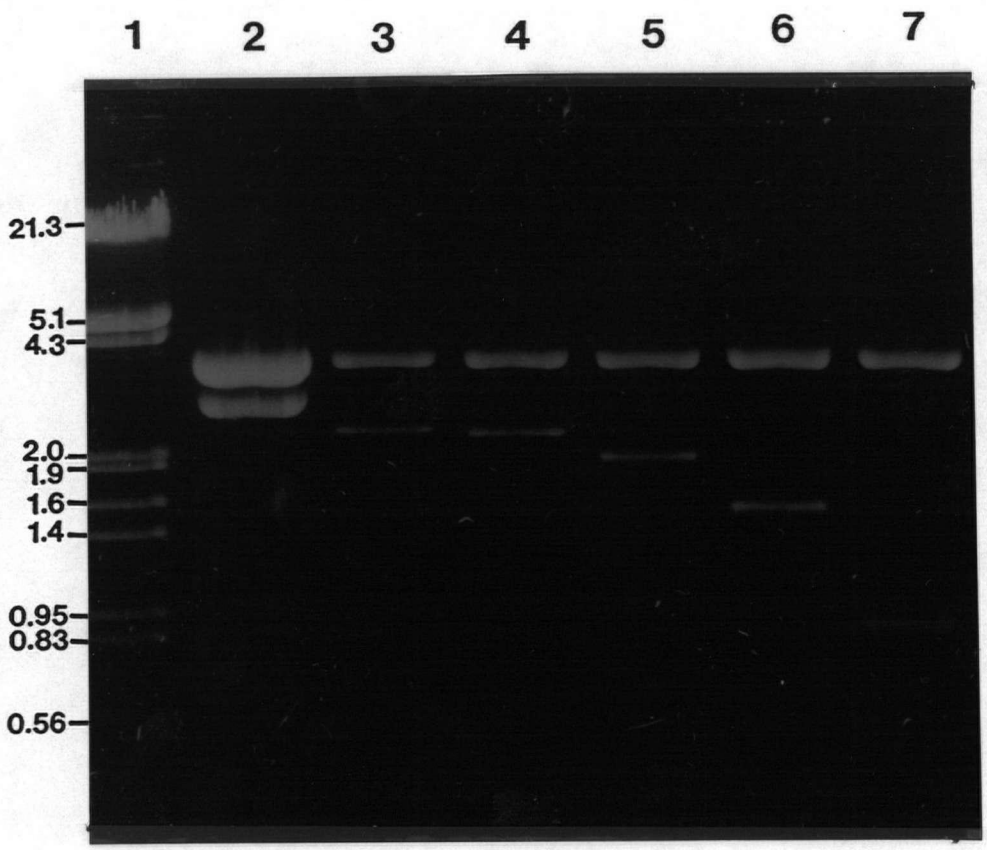


Fig. 28. Restriction digest analysis of various deletion clones. Plasmid DNA isolated from various clones were digested with *Pst*I and fractionated by 1% agarose gel electrophoresis in 1 X TBE buffer. Lane 1, *Hin*dIII-*Eco*RI lambda DNA markers; lane 2, pUC13:62; lane 3, pUC13:62Δ27; lane 4, pUC13:62Δ17; lane 5, pUC13:62Δ31; lane 6, pUC13:62Δ21-4; lane 7, pUC13:62Δ21-15 plasmids. Sizes of M_r standards are shown (Kb).



3.10. DNA and amino acid sequence determination

DNA sequencing was initiated at the region surrounding the Δ31 fusion point and going in both direction. Part of the nucleotide sequence of the 7.2 Kb insert was determined, the region covered starts from the *Bam*HI site at the 5' end of the insert up to approximatly 2200 bp downstream (Fig. 29) but not all regions were sequenced on both strands. Nevertheless, some important conclusions could be drawn.

3.10.1. DNA sequence analysis of the 5' end of *cbg*A.

A unique ATG located at position 636 was found to be in frame with the CbgA coding sequence of the Δ31 clone. Nevertheless, the upstream region did not indicate the presence of a typical prokaryotic RBS which would allow initiation of translation of CbgA at this ATG. To verify whether the region located upstream of this ATG does promote expression of *cbg*A, S1 nuclease analysis of total RNA from *C. fimi* CMC-grown cells was performed. Initially, a 214 bp 5' labelled probe which covers this region was generated by labelling the 5' end of the *Xho*I site at position 705 followed by digestion with *Stu*I at position 493. The uniquely labelled fragment released was isolated and used to detect the presence of a specific mRNA from *C. fimi* CMC-grown cells by Northern blot analysis (Appendix 2). The autoradiogram was put in Appendix because of the poor quality and low intensity of the Northern blot. This may have resulted from a low transfer efficiency or degradation of large mRNA's. Nevertheless, a specific mRNA band of approximatly 5400 nucleotides in length was detected as well as a weaker band just under the 2.37 nt marker. This RNA preparation was allowed to hydridize to the heat denatured probe which was then followed by S1 nuclease analysis of the protected DNA-RNA hybrids (Appendix 3). At the lowest concentration of S1 nuclease used (lane 1), several 5' ends were detected and were mapped at the positions 593, 637 and 675 nt relative to the DNA sequence (Fig. 29). These detected fragments were 112, 68 and 30 bases in length, respectively. As more sequencing data will become available and in the context of those identified 5' ends, the importance of those ends remains to be established since the Δ28 fusion located upstream of those mapped ends is out of frame with the CbgA

Fig. 29. Nucleotide sequence of the 5' end of *cbg*A. The 2189 nucleotides shown start at the *Bam*HI site at the 5' end and extend past the *Kpn*I site in the 7.2 Kb insert. RBS indicates the putative ribosome-binding sites preceding the GTG initiation codons. The proposed signal sequence is indicated in the boxed area. The N-terminal amino acid sequences of CbgA:Cex$_{CBD}$ and polypeptides p183 and p137 are underlined. Relevant restriction sites are indicated. The fusion points of various deletion clones are shown (▼). Numbers refered as #bp/#residue below.

```
1 BamHI  1                    RBS
gga tcc ggt cgt ggt gtc acg aag gag tga cc
33   /   1                                  63   /   11
gtg ctg cgt gcc cgt ccg acc ctg ctc cgc cgc acc cgc acc acc gcc ggc agc ccg ccg
met leu arg ala arg pro thr leu leu arg arg thr arg thr thr ala gly ser pro pro
93   /   21                                 123   /   31
gac agg cac cga cca ggc gtc gcc gcg ctg acg gcg ctc gcg ctc acg gtc ccc ctc gcg
asp arg his arg pro gly val ala ala leu thr ala leu ala leu thr val pro leu ala
153  /   41                                 183   /   51
ctc gcc gcg gcc gca ccc gcc gcg gcc tcg gcc acg gcg acc ctc ccg gag gca ccc gca
leu ala ala ala ala pro ala ala ala ser ala thr ala thr leu pro glu ala pro ala
213  /   61                                 243   /   71
ccg gcg acg gcg tcg tcc gcc ccg gca gcg gca ccg gaa gcc gca ccg gcg gcc gcg ggc
pro ala thr ala ser ser ala pro ala ala ala pro glu ala ala pro ala ala ala gly
273  /   81                                 303   /   91
gac ctg gcc cgg acc ggg acg gcg acg gcc tcg cag cac cag gcg gac ggc gac ggc acg
asp leu ala arg thr gly thr ala thr ala ser gln his gln ala asp gly asp gly thr
333  /  101                                363   /  111
ttc ccc ccg gac gcg gcg atc gac ggc gac ccg gcg acc cgc tgg gcg agc ggc aac ggc
phe pro pro asp ala ala ile asp gly asp pro ala thr arg trp ala ser gly asn gly
393  /  121          27                    423   /  131                      17
ccg gac gcg gac gtc gag ttc acg gcc tgg ctc cag gtc gac ctc ggt gcg acg gcg tcg
pro asp ala asp val glu phe thr ala trp leu gln val asp leu gly ala thr ala ser
453  /  141                                483   /  151 RBS StuI
gtc gac cgg gtg gcg ctc gcg tgg gag gcg gcg tac gcg aag gcc tac cgg gtg cag gtc
val asp arg val ala leu ala trp glu ala ala tyr ala lys ala tyr arg val gln val
513  /  161                          28    543   /  171
gcg acg gcc gcc ccg cag gac ccg gcg tcg tgg acg acg gtg cac acc gag acg gcg ggc
ala thr ala ala pro gln asp pro ala ser trp thr thr val his thr glu thr ala gly
573  /  181                                603   /  191
gac gga ggg acg gac gac gtc acg ctc ccg acg ccg gcc gac gcg cgc tac gtg cgc atc
asp gly gly thr asp asp val thr leu pro thr pro ala asp ala arg tyr val arg ile
633  /  201                                663   /  211                  24
cag atg gac gcg cgc acg tcg ttc gac tgg gac gcc ccg acg ctg cac tgg tac ggc tac
gln met asp ala arg thr ser phe asp trp asp ala pro thr leu his trp tyr gly tyr
693  /  221      XhoI 31                   723   /  231
tcg ctg ttc gcg ctc gag gtc tac ggc acg ccg ggc gcg gtg gcg acg gcg ttc ggg acg
ser leu phe ala leu glu val tyr gly thr pro gly ala val ala thr ala phe gly thr
753  /  241                                783   /  251
agc ggt gtc cgg gtg ccg gcg ggc cag acc gcg cag gtg ccg gtc gtc ctc gcg gct ccg
ser gly val arg val pro ala gly gln thr ala gln val pro val val leu ala ala pro
813  /  261                                843   /  271                SmaI
gtg gcg cag gac acg acc gtg cgg gtc gcg tcg acg ggc ggc acg gcg gtg ccc ggg acg
val ala gln asp thr thr val arg val ala ser thr gly gly thr ala val pro gly thr
```

873 / 281 903 / 291
gac ttc acc gcg gtc gac gag acg ctc acg ttc ccg gcg ggc gcg acg acg gcc acg gtc
asp phe thr ala val asp glu thr leu thr phe pro ala gly ala thr thr ala thr val
933 / 301 963 / 311
gac gtg gtg acg acg gac cac ggc ccg ctg gcc ccg gtc cgg acg gtc gtg ctg gag ctg
asp val val thr thr asp his gly pro leu ala pro val arg thr val val leu glu leu
993 / 321 1023 / 331
acg gag ccg ggc gac ggc ctc gtc ctg ggc ggc cgc acg acg gcg acg gtg acg atc acg
thr glu pro gly asp gly leu val leu gly gly arg thr thr ala thr val thr ile thr
1053 / 341 1083 / 351
ccg cac cgg ccg ctg ccg gac gtc ggc gcg gtg acg gtg ctc gac gac tac gag gac ggc
pro his arg pro leu pro asp val gly ala val thr val leu asp asp tyr glu asp gly
1113 / 361 21-4 1143 / 371
gtg ccg gcg ggc tac acg acg tgg ggg agc ggc gca ccg gtg acg ccg gtg ctg agc acg
val pro ala gly tyr thr thr trp gly ser gly ala pro val thr pro val leu ser thr
1173 / 381 1203 / 391
acg acc acg gac cga ccg ggt gcg ccg gcg ggc agc cac gcg ctg gtc ggc acc gtg ggc
thr thr thr asp arg pro gly ala pro ala gly ser his ala leu val gly thr val gly
1233 / 401 1263 / 411
ggg ccg gcg gga ccc ggt gac tgg ttc ggg ctc acg cac gac ctg ccg ccg acg gac tgg
gly pro ala gly pro gly asp trp phe gly leu thr his asp leu pro pro thr asp trp
1293 / 421 1323 / 431
tcg gac cac gac ggg ttc acg ttc tgg ttc ctc ggc acg ggc ggc ggc ggg ctg ctg cgg
ser asp his asp gly phe thr phe trp phe leu gly thr gly gly gly gly leu leu arg
1353 / 441 1383 / 451
tac gag ctc aag agc ggc ggg cag ctg ttc gag acg tcg gtc gtg gac gac acg gcg ggc
tyr glu leu lys ser gly gly gln leu phe glu thr ser val val asp asp thr ala gly
1413 / 461 1443 / 471 SmaI
tgg cgc cgg gtg aac gtc gcg ttc ggc gcg cct gcg ctg aag aac gac ccg ggc agc gac
trp arg arg val asn val ala phe gly ala pro ala leu lys asn asp pro gly ser asp
1473 / 481 1503 / 491
gcg cgg ttc gac ccg acg gcg tcg acg ggc tgg gcg atc acg ctg acc gac ctg ggc gcg
ala arg phe asp pro thr ala ser thr gly trp ala ile thr leu thr asp leu gly ala
1533 / 501 1563 / 511
gcg tgg cag ctg gac gac ctc ggc ctg tac gac cgc gtg acg acg gtc gag gac gcg gag
ala trp gln leu asp asp leu gly leu tyr asp arg val thr thr val glu asp ala glu
1593 / 521 1623 / 531
ggc gac gtc ccc gct cgc gag ccg ggc agc acg gtc ggc ctg ttc acg tgg ggc tcg tcg
gly asp val pro ala arg glu pro gly ser thr val gly leu phe thr trp gly ser ser
1653 / 541 1683 / 551
ggc gct cag gtg tcg ctc ggc gtg acg cag cag gac cgc gag ggc ggt ccg gcg gac aac
gly ala gln val ser leu gly val thr gln gln asp arg glu gly gly pro ala asp asn
1713 / 561 1743 / 571 21-15
cac gtg ctc tcg ggg cgc cta cct ggt ccg tcg ggc ggc tgg ggc ggg ttc agc cag aac
his val leu ser gly arg leu pro gly pro ser gly gly trp gly gly phe ser gln asn
1773 / 581 1803 / 591
ctc gcc gcg ccg cag gac tgg agc tcg ttc cgc ggc atc cgg ctg ctc tgg tac gcg tcg
leu ala ala pro gln asp trp ser ser phe arg gly ile arg leu leu trp tyr ala ser
1833 / 601 1863 / 611
cag gac acg cgc ccc gcg tcg ccg acg gcc ggt gac gac atc aag gtc gag ctc aag gac
gln asp thr arg pro ala ser pro thr ala gly asp asp ile lys val glu leu lys asp
1893 / 621 1923 / 631
ggc ggc ccg gac ggc gag cac tcg gag ctg tgg gcg acg acg ttc aag gac aac tgg tcg
gly gly pro asp gly glu his ser glu leu trp ala thr thr phe lys asp asn trp ser
1953 / 641 1983 / 651
ccc gac ggc agc cgc tgg aag ctc gtc gag ctg ccg ttc gac cag ttc acg ctg ggc ggg
pro asp gly ser arg trp lys leu val glu leu pro phe asp gln phe thr leu gly gly
2013 / 661 2043 / 671
tac cag ccg ggt gac gcg cag acc cgc aac ggc acg ctc gac ctc acg tcg gcg tgg ggg
tyr gln pro gly asp ala gln thr arg asn gly thr leu asp leu thr ser ala trp gly
KpnI / 681 2103 / 691
tac gcc ctg acg ttc gtg ccg ggc acc gcg aac ccg gtg cgc tgg gcg gtc gac gac gtg
tyr ala leu thr phe val pro gly thr ala asn pro val arg trp ala val asp asp val
2133 / 701 2163 / 711
cag ctg tac ggc tcg gcg gtg ccc gcg ccg acg gcc gag gtc gct ccg gcc acc gac
gln leu tyr gly ser ala val pro ala pro thr ala glu val ala pro ala thr asp

reading frame and lacks expression of enzymatic activity. Also, the probe is fully protected in lanes 1 and 2 (Appendix 3) which suggests that an additional 5' end could be located upstream of the sequence covered by the probe used. As more sequencing data became available, the legitimate transcriptional start site of the *cbg*A gene should be located upstream of the *Stu*I site.

The deletions generated previously at the 5' end of the *cbg*A structural gene were localized by DNA sequencing of the fusion region. The sizes of the remaining insert DNA in clones Δ17, Δ31, Δ21-4 and Δ21-15 (Fig. 29) correlated roughly with the sizes of the major peptides detected by Western blot analysis (Fig. 27), showing that CbgA initiated several hundred bp upstream of the Δ17 fusion point. A putative GTG initiation codon, preceded by the typical Shine-Dalgarno sequence GAAGGAG, was located 33 bp downstream of the *Bam*HI site. The presence of a GTG codon for initiation of translation is uncommun in *E. coli* but is to be expected more frequently in microorganisms with high G+C genomic content like *C. fimi* and its *cen*C gene (Table 1) and *Streptomyces* (Bibb *et al.,* 1985). This GTG may correspond to the CbgA initiation codon. Furthermore, a typical prokaryotic purine rich RBS sequence GAAGGAG which is complementary to the 3'-OH end sequence CUCCUUA-OH of *E. coli* 16S rRNA subunit is located upstream of this GTG and has a spacing of about 5 bases which is within average distance for RBS (Hawley and McClure, 1983). Consequently, this putative translational initiation site would most certainly explain the detection of ß-glucosidase activity after translational frameshift of the *Bam*HI site at the 5' end of the insert. It is likely that upstream regulatory sequences have not been isolated within this insert.

The deduced amino acid sequence following the putative GTG start site shows striking similarities with the prokaryotic signal peptides of Cex and CenA, B and C (Table XI). The presence of CbgA in *C. fimi* culture supernatants could require some kind of secretory signal. By comparison with those *C. fimi* enzymes, it is suggested that a signal peptidase processing site may be present between alanine residues 42 and 43, whereas, a prediction of prokaryotic secretory signal sequence from PC GENE computer program has suggested a processing site

Table XI. Comparison of leader peptides from *C. fimi* cellulases

Protein	Residues
CbgA	ML-RARPT----------LLRRTRTT-AG-SPPDRHRPGVAAPLT
Cex	MP-RTTPAPGHPARGARTAL---RTT--------RRR---AATL-
CenA	MSTR-R-T----A-AA--LL-------AA-----------AA---
CenB	ML-R-----QVP-----------RTLVAGGS----------A-LA
CenC	MVSR-R-SSQ--ARG---AL-----T-A-----------VVATLA
CbgA	-ALA--LTV-P-L---ALA AAAPAAA*SAT
Cex	V-VGATV-VLP-----AQA*ATT
CenA	VAVGG-VTAL-T-TTAAQA*APG
CenB	VAVGV-L-VAP-LATGAAA*APT
CenC	LALA--L-AG-S-GT-ALA*ASP

Amino acid residues are indicated in single letter code. Space (-) indicates a gap left to improve the alignment. Underlined letters represent the first residues of the mature enzymes. Star (*) indicate the location of the leader peptidase cleavage site.

between residues 51 and 52.

The following region located between amino acids 43 and 79 is enriched in Pro and Ala residues representing about 78% hydrophobic residues. Regions enriched in Ala, Pro, Thr or Ser have been reported for various cellulolytic enzymes where the specific function of such sequences remains to be determined (Warren *et al.,* 1986; Hall *et al.,* 1989: Nakai *et al.,* 1988). The carboxy proximal segment of the endo-type semi-alkaline cellulase CasA leader sequence from the alkalophilic *Streptomyces* strain KSM-9 contains an Ala-Pro repeat region which shows a striking homology with the CbgA Pro-Ala rich region (Nakai *et al.,* 1988). The *cas*A gene product apparently initiates with a GTG codon which is followed by an unusual long leader sequence of 70 A.A. This Pro-Ala rich region could be an hinge region involved in secretion of CbgA by *C. fimi.*

The elevated expression of CbgA from *E. coli* carrying pUC13:62Δ17 and Δ31 plasmids could be the result of deletion of DNA encoding the Pro-Ala rich region from CbgA as *E. coli* would become Pro-Ala depleted if required to produce high levels of proteins rich in specific residues as previously mentioned (Hall *et al.,* 1989). It is well established that amino acid starvation shuts down synthesis of the translational machinery in *E. coli.* There is little explanation other than the large size of the CbgA protein for the lack of secretion of CbgA into the periplasm of *E. coli* carrying pUC13:62 plasmid. Nevertheless, some evidence that the CbgA leader peptide can promote export of a protein in the periplasmic space of *E. coli* comes from Western blot analysis data of osmotic shockate of *E. coli*/pUC18:*cbg*A-*cex*$_{CBD}$ expressing a CbgA:Cex$_{CBD}$ fusion peptide (section 3.10.4). This fusion peptide which is composed of part of the N-terminal region of CbgA is about 7 times smaller than CbgA and was detected in the periplasmic space of those cells.

3.10.2. N-terminal amino acid sequences of polypeptides related to CbgA

The N-terminal amino acid sequences of the p183, p137 and p60 polypeptides were determined (Table XII) because it was critical to

Table XII. N-terminal amino acid sequences of CbgA:Cex_{CBD}, p183, p137 and p60

Polypeptide	N-terminal amino acid sequence
CbgA:Cex_{CBD}	SATATLPEAPAPATA
p183*	GTPGAVATAF
(ß-Gal-CbgA)	SVYGTPGAVA
p137	GSAVPAPTA-VAPATD
p60	AAKDVKFGNDA-VKMLGVNVLADAVKVTL
(continue)	GPKG-NVVLDKSFGAPTITKDGVS-A-QIL-D

The amino acid residues are indicated in single letter code.

* Two residues were obtained for each cycle, indicating ragged ends as shown.

A gap (-) indicates the absence of detectable phenylthiohydantion in that cycle. Data were provided by Ms. S.Keilland from the University of Victoria. The DNA probes were synthesized according to the underlined sequence of the p60 polypeptide.

establish a relation between the DNA data and the protein work. The N-terminal amino acid sequencing data of the p183 peptide gave a sequence mainly composed of 2 residues per cycle. The location of the amino acid sequences obtained by Edman degradation of the p183 polypeptide which is a fusion peptide with 7 residues belonging to ß-gal polypeptide, were resolved by comparison of the corresponding amino acid sequence of the DNA sequencing data of the pUC13:62Δ31 fusion region with the amino acid sequencing data. By starting with the Gly residue of the first cycle, a first sequence was located following the predicted amino acid sequence and matching one residue at a time from each cycle. The remaining amino acid sequence of each cycle was then compared with the predicted sequence and was found to correspond to the sequence encoded in part by the fused ß-gal alpha-peptide, the Ser residue was located at the fusion point and was part of that peptide. At the fourth cycle, twice the amount of Gly residues were detected originating from the simultanous appearance of residues 229 and 232 of CbgA. These data indicate the presence of one polypeptide with different N-terminal sequences apparently processed by proteolytic cleavages, between residues Arg-6 and Ser-7 of the ß-gal alpha-peptide (not shown) and between residues Tyr-228 and Gly-229 of CbgA (Fig. 29) near the fusion point of Δ31. Furthermore, the p137 amino acid sequence (Table XII) was located at the end of the given sequence and was also processed between identical residues, Tyr-703 and Gly-704. A rather specific processing event was also observed in *E. coli* expressing Cex from *C. fimi* (Gilkes *et al.,* 1988). In an attempt to locate the p60 amino acid sequence data to the DNA, a set of DNA probes corresponding to part of the p60 amino acid sequence was used to screen the entire 7.2 Kb insert by Southern blot analysis. The probes were a mixture of 17-mers with the possible sequences AA(G/A)GA(C/T)GTNAA (G/A)TT(C/T)GG corresponding to residues K-D-V-K-F-G located at the beginning of p60 amino acid sequence (Table XII). This region was chosen because of the low number of variants for the synthesis of oligo-probes with multiple codons. No specific hybridization was obtained, therefore, the location of the amino acid sequence of the p60 polypeptide and its relatedness to CbgA could not be established. The only

interesting feature was an 83% homology between residues 18 to 56 of p60 and residues 39 to 77 of the 65 kDa cell wall protein antigen of *Mycobacterium tuberculosis* (Shinnick, 1987).

Proteolysis is most likely to be responsible for the apperearance of lower-Mr species from the various deletions analysed so far as well as from *C. fimi* where a peptide of approximatly 137 kDa in size was detected in most samples. This peptide could correspond to the previously isolated p137 peptide originating from processing between residues Tyr-703 and Gly-704 of CbgA isolated from *E. coli* carrying pUC13:62Δ31 plasmid. The reasons for the removal of more than 700 residues from the N-terminus of CbgA is not clear. These amino acids are not essential for activity since 1765 bp of DNA were deleted in the Δ21-15 clone without a strong decrease in the specific activity. The presence of long amino acid sequences between the signal sequences and the N-termini of the mature processed proteins have been reported for the secreted alkaline and neutral proteases of *B. amyloliquefaciens* (Vasantha *et al.*,1984). Similarly, proteolytic cleavage also occurs after a Tyr residue in the case of the apr[BamP] gene product (Vasantha *et al.*, 1984). A possible role of the pro-sequences of these proteases could be in secretion of these enzymes. The function of the pro-sequence from CbgA is unclear at present.

3.10.3. Mutation of the CbgA translational start site

To confirm the putative GTG start site of translation, a single base frameshift mutation was generated within the codon (Fig. 30). The frameshift did not reduce the expression of CbgA in *E. coli*, suggesting that translation was initiated at another codon further downstream. The polypeptide produced by the frameshift mutant was smaller than that produced by the wild-type (Fig. 31). This suggested the presence of an internal reinitiation site downstream from the putative GTG start site. Examination of DNA sequence revealed a putative Shine-Dalgarno sequence (CGAAGG) followed by a GTG codon located at position 504. The Δ27 deletion mutant which is out of frame with the CbgA coding sequence and also with the Plac reading frame has a similar specific activity to the wild-type but produces a number of peptides, all of which

Fig. 30. Mutation of the putative GTG start site of CbgA. The putative CbgA GTG start site (underlined) was mutated by primer extension. The 2 Kb fragment from the *C. fimi* DNA (hatched bars) subcloned into pUC19 giving rise to pUC19:1.9 plasmid was transferred into the vector pTZ19R. Consequently, pTZ19R-2 ssDNA was isolated and allowed to hybridize with the FP21 oligo prior to synthesis of dsDNA. Double stranded pTZ19R-2 plasmid containing the extra C base (pTZ19R-2FP21) was identified by DNA sequencing. The wild type 6.7 Kb *Stu*I-*Eco*RI fragment was subcloned into pTZ19R-2FP21 creating pTZ19R-7. The DNA sequence of the beginning of the insert is shown, the location of the putative RBS is indicated and star (*) shows the location of the C base added to the wild type DNA sequence. Vector DNA (solid bars) and convenient restriction sites are shown. Scale of drawing is indicated above.

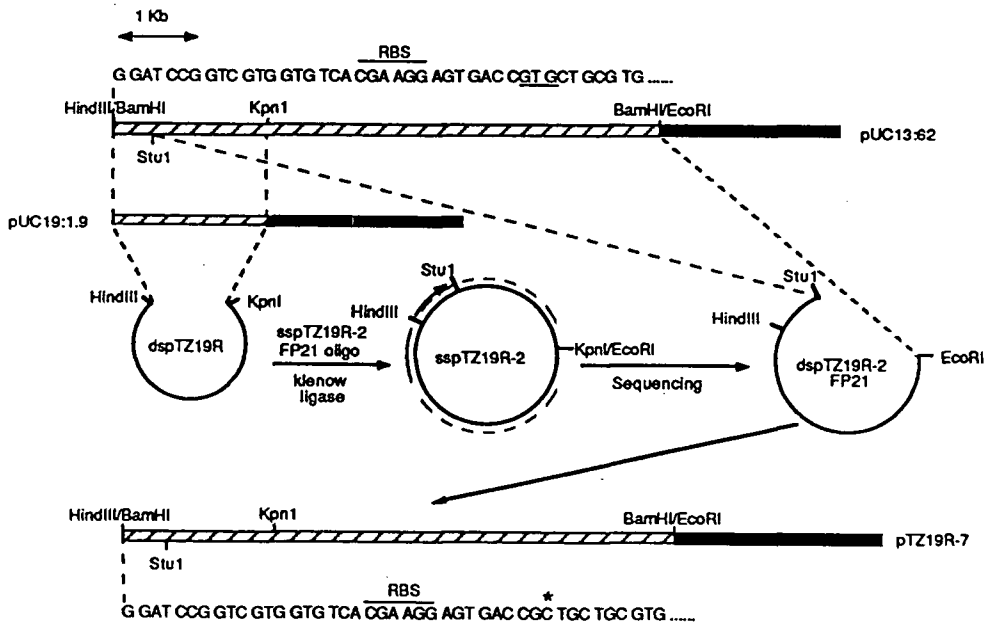
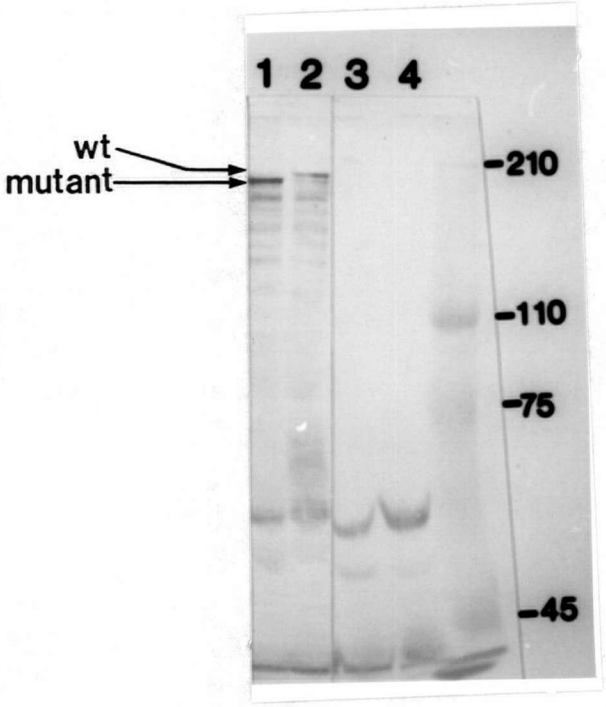


Fig. 31. Western blot analysis of the polypeptides produced by the *cbg*A translational frameshift mutant. Lane 1, 30 µg of *E. coli*/pTZ17R-7 total cell extract (or mutant); lane 2, 30µg of *E. coli*/pUC13:62 total cell extract (or wild-type); lane 3, 30 µg of *E. coli*/pTZ19R total cell extract (control); lane 4, 30 µg of *E. coli*/pUC13 total cell extract (control). Prestained Mr standards (kDa) are shown on the right. Antigenic polypeptides were detected using CbgA antiserum (see section 2.9). Arrows indicate the location of the relevant polypeptides.

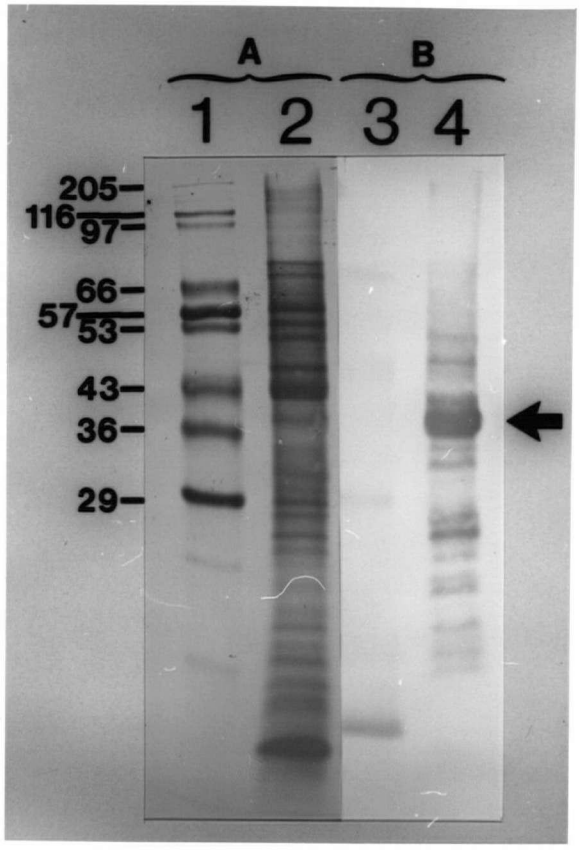


are smaller than the wild-type polypeptide. The end-points of the deletions in Δ27 and Δ28 are nucleotides 408 and 540, respectively. This supports the suggestion that the GTG at nucleotide 504 functions as an internal translation initiation site. The Δ28 deletion is also out of frame but it does not express ß-glucosidase activity. Internal reinitiation was also observed in the *xln*Z gene of *C. thermocellum* (Grépinet *et al.*, 1988) and in the *xln*A gene of *P. fluorescens* subspecies *cellulosa* (Hall *et al.*, 1989). In the case of *cbg*A, the initiation of translation from the upstream sequence adds 157 residues to CbgA and could allow secretion of the enzyme whereas initiation from the downstream GTG deletes the putative leader sequence and the Pro-Ala rich region and may force the enzyme produced to remain in the cytosol. CbgA was indeed detected in *C. fimi* cytosol and culture medium.

3.10.4. Cellulose affinity chromatography of a CbgA:Cex_{CBD} fusion peptide

To facilitate the determination of the N-terminal amino acid sequence of CbgA, a fusion protein was constructed comprising the N-terminal part of CbgA fused to the cellulose binding domain of Cex (Cex_{CBD}). The fusion product was purified by cellulose affinity column chromatography essentially as described (Gilkes *et al.*, 1988). Initially, a gene fusion was constructed using in-frame *Stu*I restriction sites within CbgA and Cex coding sequences. The fusion peptide CbgA:Cex_{CBD}, expressed in *E. coli* carrying the pUC18:*cbg*A-*cex*$_{CBD}$ plasmid, contained the first 154 residues of CbgA with its putative leader sequence and Pro-Ala rich region (Fig. 29) fused to the last 133 residues of Cex comprising the Pro-Thr box and the cellulose binding domain. The CbgA:Cex_{CBD} fusion protein reacted with antiserum to Cex and was detected in *E. coli* periplasmic space but its Mr of 37 kDa (Fig. 32) was 8.5 kDa larger than the expected value. The discrepancy between the predicted and observed Mr values could be attributed to the presence of the Pro-Thr box. Peptides derived from CenA and containing the Pro-Thr box migrate slower on SDS-PAGE (Gilkes *et al.*, 1989). The amino acid sequence of the N-terminal region of the fusion peptide matched perfectly well with residues 50 to 64 of the predicted sequence of CbgA

Fig. 32. Purification of the CbgA:Cex_{CBD} fusion peptide. The CbgA:Cex_{CBD} fusion peptide was purified from *E. coli*/pUC18:*cbgA-cex*$_{CBD}$ total cell extract by cellulose affinity chromatography on CF1 cellulose. The water eluate was concentrated by ultrafiltration and proteins were analyzed by (A) Coomassie blue staining and (B) Western blot of a 10% SDS-PAGE using antiserum to Cex protein. Lane 1, Mr standards (kDa); lane 2, 23 µg of protein; lane 3, prestained Mr standards; lane 4, 23 µg of protein. The location of the antigenic band corresponding to the CbgA:Cex_{CBD} fusion peptide is shown.



(Fig. 29). The observed Mr (37 kDa) of the fusion polypeptide is 13.5 kDa greater than predicted suggesting that the Pro-Ala rich region of CbgA may also affect migration of peptides on SDS-PAGE.

3.11. General discussion

The recombinant *C. fimi* ß-glucosidase A (CbgA) was characterized in *E. coli*. The gene was over expressed and the enzyme produced was purified. The kinetic parameters indicated that CbgA is an exo-ß-1,4-glucan glucohydrolase (EC3.2.1.74) because of its linear action and preference for soluble cellodextrins released during the hydrolysis of cellulose. Similar enzymes of fungal and bacterial origins acting on oligomeric substrates, such as cellodextrin glucohydrolase, have been reported elsewhere (see Schmid and Wandrey, 1989). So far, 13 different fungal exo-ß-1,4-glucosidases have been characterized and shown to hydrolyze cello-oligosaccharides more efficiently than cellobiose (Schmid and Wandrey, 1989). Only one similar bacterial exo-ß-1,4-glucosidase has been characterized but this enzyme hydrolyses pNPG more efficiently than the oligo-saccharides and is found cell-associated (Ait *et al.*, 1982).

The ß-glucosidase expressed in *E. coli* carrying the pUC13:62 plasmid is apparently a very large protein (M_r of 210 kDa). *E. coli* expressing *cbg*A does not appear to be affected by the production of this heterologous peptide suggesting that it can produce proteins of unusually high molecular weights. Only a few proteins with molecular masses greater than 100 kDa are found in *E. coli* (Tsung et al., 1989). The lack of secretion by *E. coli* of a normally secreted protein is often a problem that can be circumvented by the use of an alternative host. Apparently, CbgA can be processed in a specific and perhaps identical manner in *E. coli* and *C. fimi* suggesting that the structural features of the protein may dictate the way processing occurs. There is some evidence that CbgA forms aggregates either with itself, as do other ß-glucosidases, or with other proteins. The nature of the aggregation remains unknown at present. Interestingly, the aggregation of several different peptides that have various cellulolytic activities occurs in nature. These complex structures are known as "cellulosomes". The

corresponding *C. fimi* CbgA enzyme has been identified within a group of isoenzymes hydrolyzing various ß-glucosides. CbgA is secreted by *C. fimi* which is unusual for a bacterial ß-glucosidase.

The analysis of the 5' end of *cbg*A was found to be very informative regarding the mode of expression and structure of part of the protein indicating : (1) the presence of a putative leader sequence for secretion of the enzyme, composition of which was similar to those found in the other *C. fimi* cellulase leader peptides, (2) an internal translational initiation site presumably involved in the production of a peptide lacking its secretory signal peptide. This could preclude the secretion of that peptide into the culture medium. A need for the enzyme within the cell and also secreted into the medium can be rationalized if, during growth on cellulose, large cellodextrins must be hydrolysed in the culture medium whereas smaller ones enter the cells. One could also speculate that the expression of *cbg*A in *C. fimi* is regulated by a set of two different promoters. The first one catabolite repressed and inducible by cellulosic derivatives responsible for expression of *cbg*A from the most upstream region allowing secretion of CbgA in the culture medium where its activity is required. The second one constitutive and located within the structural gene, may prevent secretion of CbgA which could be involved in the hydrolysis of other ß-glucosides that may enter the cells or the generation of inducer molecules from transglycosylation activity. Interestingly, a sequence of 9 bp inverted repeats is located near a -35 region upstream from the second translational initiation site and may be some kind of regulatory sequence.

The analysis of this rather speculative mode of expression of *cbg*A in *C. fimi* would require cloning of more of the region upstream from the gene followed by transcriptional analysis of induced and non-induced states of the system. Mutants lacking CbgA activity are essential to define the precise function of CbgA in the hydrolysis of cellulose by *Cellulomonas fimi.* Finally, CbgA could be used in conjunction with an endo-glucanase to produce glucose directly from cellulose. *C. fimi* ß-glucosidase A may play an important role in the hydrolysis of cellulose. Its size may be indicative of the complexity of this role.

REFERENCES

Ait, N., Creuset, N. and Cattaneo, J. (1979) Characterization and purification of thermostable ß-glucosidase from *Clostridium thermocellum.* Biochem. Biophys. Res. Comm. 90 : 537-546.

Ait, N., Creuzetm, N. and Cattaneo, J. (1982) Properties of ß-glucosidase purified from *Clostridium thermocellum.* J. Gen. Microbiol. 128 : 569-577.

Alexander, J.K. (1968) Purification and specificity of cellobiose phosphorylase from *Clostridium thermocellum.* J. Biol. Chem. 243 : 2899-2904.

Armentrout, R.W. and Brown, R.D. (1981) Molecular cloning of genes for cellobiose utilization and their expression in *Escherichia coli.* Appl. Environ. Microbiol. 41 : 1355-1362.

Atkinson, T. and Smith. M. (1984) In; Oligonucleotide synthesis: a practical approach. Gait, N.J.(ed), IRL Press, Washington.

Bates, N.C. (1987) Characterization of *cbg* : a cloned gene encoding an extracellular ß-glucosidase from *Cellulomonas fimi.* M. Sc. Thesis, the University of British-Columbia.

Béguin, P., Gilkes, N.R., Kilburn, D.G., Miller, R.C.Jr., O'Neil, G.P. and Warren, R.A.J. (1987) Cloning of cellulase genes. CRC Critical Reviews in Biotechnology, vol 6 (2), pp.129-162.

Berghem, L.E.R. and Pettersson, L.G.)1973) The machanism of enzymatic cellulose degradation. Purification of a cellulolytic enzyme from *Trichoderma viride* active on highly ordered cellulose. Eur. J. Biochem. 37 : 21-30.

Birnboim, H.L. and Doly, J. (1979) A rapid alkaline extraction procedure for screening of recombinant plasmid DNA. Nucl. Ac. Res. 7 : 1513-1523.

Bisaria, V.S. and Ghose, T.K. (1981) Biodegradation of cellulosic materials: substrates, microorganisms, enzymes and products. Enzyme Microb. Technol. 3 : 90-104.

Blake, M.S., Johnston, K.H., Russel-Jones, G.J. and Gotschlich, E.C. (1984) A rapid, sensitive method for detection of alkaline phosphatase-conjugate anti-antibody on western blots. Analytical Biochem. 136 : 175-179.

Bousfield, I.J. and Callely, A.G. (eds) (1978) Coryneform bacteria. Academic Press. London.

Boyer, H.W., Betlach, M., Bolivar, F., Rodriguez, R.L., Heyneker, H.L., Shine, J. and Goodman, H.M. (1977) The construction of molecular cloning

vehicles, In Beers, R.F. and Basset, E.G. (eds.), Recombinant DNA molecules, 10 , Miles Symposium Series. Raven Press, New York, pp.9-19.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72 : 248-254.

Bronnenmeier, K. and Standenbauer, W.N. (1988) Purification and properties of an extracellular ß-glucosidase from the celllolytic thermophile *Clostridium stercorarium.* Appl. Microbiol. Biotechnol. 28 : 380-386.

Chassy, B.M., Mercenier, A. and Flickinger, J. (1988) Transformation of bacteria by electroporation. Elsevier Publication, Cambridge, Tibtech 6 : 303-310.

Cornish-Bowden, A. (1981) Fundamentals of enzyme kinetices. Butterworth & Co (Publishers) Ltd. UK.

Coughlan, M.P. (1985) The properties of fungal and bacterial cellulases with comment on their production and application. Biotechnology and Genetic Engineering Reviews, vol 3, pp.53-109.

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

Darveau, R.P. and Hancock, R.E.W. (1983) Procedure for isolation of bacterial lipopolysaccharide from both smooth and rough *Pseudomonas aeruginosa* and *Salmonella typhumurium* strains. J. Bact. 155 : 831-838.

Day. A.G. and Withers, S.G. (1985) The purification and characterization of a ß-glucosidase from *Alcaligenes feacalis.* Biochem. Cell. Biol. 64 : 914-922.

Dayhoff, M.O., Dayhoff, R.E. and Hunt, L.T. (1976) In "Atlas of protein sequence and structure". Dayhoff, M.O., Ed., National Biomedical Research Fondation : Washington, DC, 5 : 301-310.

Din, N., Beck, C.F., Miller, R.C.Jr., Kilburn, D.G. and Warren, R.A.J. (1990) Expression of the *Cellulomonas fimi* cellulase genes *cex* and *cen*A from the divergent *tet* promoters of transposon Tn10. Arch. Microbiol. 153 : 129-133.

Dinur, T., Osiecki, K.M., Legler, G., Gatt, S., Desnick, R.J. and Grabowski, G.A. (1986) Human acid ß-glucosidase: Isolation and amino acid sequence of a peptide containing the catalytic site. Proc. Natl. Acad. Sc. USA. 83 : 1660-1664.

Ehrenberg, L., Fedorcsak, I. and Lolymosy, F. (1976) Diethylpyrocarbonate in nucleic acid research. J. Mol. Biol. 16 : 189-262.

Ericksson, K.E. and Wood, T.M. (1985) in "Biosynthesis and Biodegradation of Wood components". (T. Higuchi, ed) Academic Press, New York, p.469.

Fan, L.T., Gharpuray, M.M. and Lee, Y.-H. (1987) Cellulose hydrolysis. Biotechnology Monographs, vol. 3,

Favoloro, J., Triesman, R. and Kamen, R. (1980) Transcription maps of polyoma virus-specific RNA: Analysis by two-dimentional nuclease S1 gel mapping. Methods Enzymol. 65 : 718-749.

Gerhardt, P., Murray, R.G.E., Costilow, R.N., Nester, E.W., Wood, W.A., Krieg, N.R. and Phillips, G.B. (1981) Manual of methods for general bacteriology., A.S.M. Washington, D.C. 20006.

Gilkes, N.R., Kilburn, D.G., Langsford, M.L., Miller, R.C.Jr., Wakarchuk, W.W., Warren, R.A.J., Whittle, D.J. and Wong, W.K.R. (1984) Isolation and characterization of *Escherichia coli* clones expressing cellulase genes from *Cellulomonas fimi.* J. Gen. Microbiol. 130 : 1377-1384.

Gilkes, N.R., Langsford, M.L., Kilburn, D.G., Miller, R.C.Jr. and Warren, R.A.J. (1984a) Mode of action and substrate specificities of cellulases from cloned bacterial genes. J. Biol. Chem. 259 : 10455-10459.

Gilkes, N.R., Kilburn, D.G., Miller, R.C.Jr. and Warren, R.A.J. (1984b) A mutant of *Escherichia coli* that leaks cellulase activity encoded by cloned cellulase genes from *Cellulomonas fimi.* Bio/Technology 2 : 259-263.

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

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

Gokhale, D.V., Soo-Han, E., Srinivan, V.R. and Dileep, N.D. (1984) Transfer of DNA coding for cellulases from *Cellulomonas species* to *Bacillus subtilis* by protoplast fusion. Biotechnology Letters 6 : 627-632.

Gokhale, D.V. and Deobagkar, D.N. (1989) Differential expression of xylanases and endoglucanases in the hybrid derived from intergeneric protoplast fusion between a *Cellulomonas* sp. and *Bacillus subtilis.* Appl. Environ. Microbiol. 55 : 2675-2680.

Gong, C.-S. and Tsao, G.T. (1979) Cellulase and biosynthesis regulation. Annual Reports on Fermentation Processes. 3 : 111-140.

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

Greenberg, N.M., Warren, R.A.J., Kilburn, D.J. and Miller, R.C.Jr. (1987b) Regulation and initiation of *cen*B transcripts of *Cellulomonas fimi.* J. Bact. 169 : 4674-4677.

Greenberg, N.M. (1988) Cellulase gene transcription in *Cellulomonas fimi* and an *Agrobacterium.* Ph. D. Thesis., the University of British-Columbia.

Greenwood, J.M., Gilkes, N.R., Kilburn, D.J., Miller, R.C.Jr. and Warren, R.A.J. (1989) Fusion to an endoglucanase allows alkaline phosphatase to bind to cellulose. FEBS letters 244 : 127-131.

Grépinet, O., Chebrou, M.-C. and Béguin, P. (1988) Nucleotide sequence and deletion analysis of the xylanase (*xyn*Z) of *Clostridium thermocellum.* J. Bact. 170 : 4582-4588.

Groleau, D. and Forsberg, C.W. (1981) Cellulolytic activity of rumen bacterium *Bacteroided succinogenes.* Can. J. Microbiol. 27 :517-530.

Gusakov, A.V., Sinitsyn, A.P., Goldstein, G.H. and Klyosov, A.A. (1984) Kinetics and mathematical model ot hydrolysis and transglycosylation catalysed by cellobiase. Enzyme Microbial Technology 6 : 275-282.

Hall, J., Hazlewood, G.P., Huskisson, N.S., Durrant, A.J. and Gilbert, H.J. (1989) Conserved serine-rich sequences in xylanase and cellulase from *Pseudomonas fluorescens* subspecies *cellulosa*: internal signal sequence and unusual protein processing. Molec. Microbiol. 3 1211-1219.

Han, Y.W. and Srinivasan, V.R. (1969) Purification and characterization of ß-glucosidase of *Alcaligenes faecalis.* J. Bact. 100 : 1353-1363.

Hsuanya, Y. and Laidler, K.J. (1984) Kinetics of the hydrolysis of cellobiose catalysed by ß-glucosidase. Can. J. Biochem. Cell. Biol. 63 : 167-175.

Hwang, J.-T. and Suzuki, H. (1976) Intracellular distribution and some properties of ß-glucosidases of a cellulolytic pseudomonad. Agr. Biol. Chem. 40 : 2169-2175.

Jobe, A. and Bourgeois, S. (1972) Lac repressor-operator interaction. VI. The natural inducer of the operon. J. Mol. Biol. 69 : 397-408.

Johnson, J.A., Wong, W.K.R. and Beatty, J.T. (1986) Expression of cellulase genes in *Rhodobacter capsulatus* by use of plasmid expression vectors. J. Bact. 167 : 604-610.

Jovin, T., Chrambach, A. and Naughton, M.A. (1964) An apparatus for temperature-regulated polycarylamide gel electrophoresis. Anal. Biochem. 9 : 351-369.

Katsumata, R., Ozaki, A., Oka, t. and Furuya, A. (1984) Protoplasts transformation of glutamate-producing bacteria with plasmid DNA. J. Bact. 159 : 306-311.

Kim, H.K. and Pack, M.Y. (1989) Cloning and expression of *Cellulomonas fimi* ß-glucosidase genes in *Escherichia coli.* Enz. Microb. Technol. 11 : 313-316.

Komagata, K., Yamada, K. and Ogawa, H. (1969) Taxonomic studies on coryneform bacteria. I. Division of bacterial cells. J. Gen. Appl. Microbiol. 15 : 243-259.

Kraft, R., Tardiff, J., Kranter, K.S. and Leinwand, L.A. (1988) Using mini-prep plasmid DNA for sequencing double stranded templates with sequenase^{tm}. Bio/Techniques 6 : 544-547.

Kricker, M. and Hall, B.G. (1984) Direct evolution of cellobiose utilization in *Escherichia coli* K12. Mol. Biol. Evol. 1 : 171-182.

Kristensen, T., Hartmut, V. and Wilhelm, A. (1987) A simple and rapid preparation of M13 sequencing templates for manual and automated dideoxy sequencing. Nucl. Ac. Res. 15 : 5507-5516.

Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. 154 : 367-382.

Lacks, S.A. and Springhorn, S.S. (1980) Renaturation of enzymes after polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate. J. Biol. Chem. 255 : 7467-7473.

Langsford, M.L., Gilkes, N.R., Wakarchuk, W.W., Kilburn, D.G., Miller, R.C.Jr. and Warren, R.A.J. (1984) The cellulase system of *Cellulomonas fimi.* J. Gen. Microbiol. 130 : 1367-1376.

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

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

Mandel, M. and Higa, A. (1970) Calcium dependent bacteriophage DNA infection. J. Mol. Biol. 53 : 154-

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

Martin, J.F., Santamaria, R., Sandoval, H., Real, G.D., Mateos, L.M., Gil, J.A. and Aguilar, A. (1987) Cloning systems in amino acid-producing corynebacteria. Bio/Technology Review 5 : 137-146.

Matthews, B.W.,Nicholson, H. and Becktel, W.J. (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. Proc. Natl. Acad. Sci. USA 84 : 6663-6667.

Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. 65 : 499-560.

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

Miller, G.L. (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Analyt. Chem. 31 : 426-428.

Miwa, K., Matsui, K., Terabe, M., Ito, K., Ishida, M., Tagaki, H., Nakamori, S. and Sano, K. (1985) Construction of novel shuttle vectors and a cosmid vector for the glutamic acid-producing bacteria *Brevibacterium lactofermentum* and *Corynebacterium glutamicum.* Gene 39 : 281-286.

Mizusaqa, S., Nishimura, S. and Seela, F. (1986) Improvement of the dideoxy chain terminator method of DNA sequencing by use of deoxy-7-deaza guanosine triphosphate in place of dGTP. Nucleic Acid Research 14 : 1319-1324.

Moldoveanu, N. and Kluepfel, D. (1983) Comparison of ß-glucosidase activities in different *Streptomyces* strains. Appl. Environ. Microbiol. 46 : 17-21.

Moser, B. (1988) Molecular cloning, characterization and expression of the endoglucanase C gene of *Cellulomonas fimi* and properties of the native and recombinant gene products. Ph. D. Thesis. The University of British-columbia.

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

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.

Nakamori, S., Ishida, M., Takagi, H., Ito, K., Miwa, K. and Sano, K. (1987)

Improved L-threonine production by the amplification of the gene encoding homoserine dehydrogenase in *Brevibacterium lactofermentum.* Agricultural and Biological Chemistry 51 : 87-91.

Nakamura, K., Misawa, N. and Kitamura, K. (1986) Cellulase genes of *Cellulomonas uda* CB4. I. Cloning and expression of ß-glucosidase genes in *Escherichia coli.* J. Biotechnol. 3 : 239-246.

Nisizawa, T., Suzuki, H., Nakayama, M. and Nisizawa, K. (1971) Inductive formation of cellulase by sophorose in *Trichoderma viride.* J. Biochem. 70 : 375-385.

Nossal, N.G. and Heppel, L.A. (1966) The release of enzymes by osmotic shock from *Escherichia coli* in exponential phase. J. Biol. Chem. 241 : 3055-3062.

O'Callaghan, C.H., Morris, A., Kirby, S.M. and Shingler, A.H. (1972d) Novel method for detection of ß-lactamase by using a chromogenic cephalosporin substrate. Antimicrob. Ag. Chemother., 1 : 283-288.

Oka, A., Sugisaki, H. and Takanami, M. (1981) Nucleotide sequence of the kanamycin resistance transposon Tn903. J. Mol. Biol. 147 : 217-226.

O'Neil, G.P. (1986) Structure of the gene encoding the exoglucanase of *Cellulomonas fimi* and its expression and secretion in *Escherichia coli.* Ph. D. Thesis, the University of British-Columbia.

O'Neil, G.P., Goh, S.H., Warren, R.A.J., Kilburn, D.J. and Miller, R.C.Jr. (1986) Structure of the gene encoding the exoglucanase of *Cellulomonas fimi.* Gene 44 : 325-330.

O'Neil, G.P., Kilburn, D.J., Warren, R.A.J. and Miller, R.C.Jr. (1986a) Overproduction from a cellulase gene with a high guanosine-plus-cytosine content in *Escherichia coli.* Appl. Environ. Microbiol. 52 : 737-743.

O'Neil, G.P., Warren, R.A.J., Kilburn, D.J. and Miller, R.C.Jr. (1986b) Secretion of *Cellulomonas fimi* exoglucanase by *Escherichia coli.* Gene 44 : 331-336.

Ong, E., Gilkes, N.R., Kilburn, D.J., Miller, R.C.Jr. and Warren, R.A.J. (1989) Enzyme immobilization using the cellulose-binding domain of a *Cellulomonas fimi* exoglucanase. Bio/Technology 7 : 604-607.

Osborn, M.J., Gander, J.E., Parisi, E. and Carson, J. (1972) Mechanism of assembly of the outer membrane of *Salmonella typhimurium.* Isolation and characterization of cytoplasmic and outer membranes. J. Biol. Chem. 247 : 3962-3972.

Osborn, M.J. and Munson, R. (1974) Separation of the inner (cytoplasmic) and

outer membranes of Gram-negative bacteria. Methods Enzymol. 31 : 642-667.

Owolabi, J.B. (1988) Characterization and expression of *Cellulomonas fimi* endoglucanase B gene and properties of the gene product from *Escherichia coli*. Ph. D. Thesis. The University of British-Columbia.

Owolabi, J.B., Beguin, P., Kilburn, D.J., Miller, R.C.Jr. and Warren, R.A.J. (1988) Expression in *Escherichia coli* of the *Cellulomonas fimi* structural gene for endoglucanase B. Appl. Environ. Microbiol. 54 : 518-523.

Pack, Y.-H., Hori, H., Suzuki, K.-I., Osawa, S. and komagata, K. (1987) Phylogenetic analysis of the coryneform bacteria by 5S rRNA sequences. J. Bact. 169 : 1801-1806.

Paradis, F.W., Warren, R.A.J., Kilburn, D.J. and Miller, R.C.Jr. (1987) The expression of *Cellulomonas fimi* cellulase genes in *Brevibacterium lactofermentum*. Gene 61 : 199-206.

Pentilla, M.E., Nevalainen, K.M.H., Raynal, A. and Knowles, J.C.K. (1984) Cloning of *Aspergillus niger* genes in yeast. Expression of the gene coding *Aspergillus* ß-glucosidase. Mol. Gen. Genet. 194 : 494-499.

Raynald, A. and Guérineau, M. (1984) Cloning and expression of the structural gene for ß-glucosidase of *Kluyveromyces fragilis* in *Escherichia coli* and *Saccharomyces cerevisiae*. Mol. Gen. Genet. 195 : 108-115.

Saddler, J.N. and Khan. A.W. (1981) Cellulolytic enzyme system of *Acetivibrio cellulolyticus*. Can. J. Microbiol. 27 : 288-294.

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

Sano, K., Amemura, A. and Harada, T. (1975) Purification and properties of a ß-1,6-glucosidase from *Flavobacterium*. *Biochemica et biophysica acta*. 377 : 410-420.

Santamaria, R., Gil, J.A., Mesas, J.M. and Martin, J. (1984) Characterization of an endogenous plasmid and development of cloning vectors and a transformation system in *Brevibacterium lactofermentum*. J. Gen. Microbiol. 130 : 2237-2246.

Santamaria, R.I., Gil, J.A. and Martin, J.F. (1985) High-frequency transformation of *Brevibacterium lactofermentum* protoplasts by plasmid DNA. J. Bact. 162 : 463-467.

Sato, M. and Takahashi, H. (1967) Fermentation of C^{14}-labelled cellobiose by *Cellulomonas fimi*. Agricult. Biologic. Chem. 31 : 470-474.

Schaefler, S. (1967) Inducible system for the utilization of ß-glucosides in *Escherichia coli.* J. Bact. 93 : 254-263.

Schimz, K.-L., Broll, B. and John, B. (1983) Cellobiose phosphorylase (EC 2.4.1.20) of *Cellulomonas*: occurence, induction, and its role in cellobiose metabolism. Arch. Microbiol. 135 : 241-249.

Schmid, G. and Wandrey, C. (1989) Characterization of a cellodextrin glucohydrolase with soluble oligomeric substrates: Experimental results and modeling of concentration-time-course data. Biotech. Bioengineer. 33 : 1445-1460.

Shaw, P-.C. and Hartley, B.S. (1988) A host-vector system for an *Arthrobacter species* . J. Gen. Microbiol. 134 : 903-911.

Shewale, J.G. and Sadana, J.C. (1981) Purification, characterization and properties of ß-glucosidase enzymes from *Sclerotium solfsii.* Archs Biochem. Biophys. 207 : 185-196.

Shewale, J.G. (1982) ß-glucosidase : its role in cellulase synthesis and hydrolysis of cellulose. Can. J. Biochem. 14 : 435-443.

Shinnick, T.M. (1987) The 65-kilodalton antigen of *Mycobacterium tuberculosis.* J. Bact. 169 : 1080-1088.

Singer, M.E. (1986) Plasmid transformation in *Arthrobacter* sp. H13A. In: Abstracts of the A.S.M. Annual Meeting, March 23-28, Washington, D.C.

Sinitsyn, A.P. and Klesov, A.A. (1981) Comparison of roles of exo-1,4-ß-glucosidase and cellobiase in enzymatic hydrolysis of celluloase. Biochem. 46 : 165-174.

Skipper, N., Sutherland, M., Davies, R.W., Kilburn, D.J., Miller, R.C.Jr., Warren, R.A.J. and Wong, R. (1985) Secretion of a bacterial cellulase by yeast. Science 230 : 958-960.

Smith, M.D., Flickinger, J.L., Lineberger, D.W. and Schmidt, B. (1986) Protoplast transformation in coryneform bacteria and introduction of an alpha-amylase gene from *Bacillus amuloliquefaciens* into *Brevibacterium lactofermentum.* Appl. Environ. Microbiol. 51 : 634-639.

Southern, E.M. (1975) Detection of specific sequences amoung DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98 : 503-517.

Sprey, B. and Lambert, C. (1983) Titration curves of cellulases from *Trichoderma reesei* : Demonstration of a cellulase-xylanase-ß-glucosidase-containing complex. FEMS Microbiol. Letters 18 : 217-222.

Stackebrandt, E. and Kandler, O. (1979) Taxonomy of the genus

Cellulomonas, based on phenotypic characters and DNA-DNA homology and proposal of seven neotype strains. Int. J. Syst. Bacteriol. 29 : 263-282.

Stuber, D. and Bujard, H. (1981) Organization of transcriptional signals in plasmids pBR322 and pACYC184. Proc. Natl. Acad. Sci. USA 78 : 167-171.

Stoppock, W., Rapp,P. and Wagner, F. (1982) Formation, localization and regulation of endo-1,4-ß-glucanases and ß-glucosidases from *Cellulomonas uda.* Appl. Environ. Microbiol. 44 : 43-53.

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

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

Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA 76 : 4350-4354.

Tsung, K., Inouye, S. and Inouye, M. (1989) Factors affecting the efficiency of protein synthesis in *Escherichia coli.* J. Biol. Chem. 264 : 4428-4433.

Umile, C. and Kubicek, C.P. (1986) A constitutive plasma-membrane bound ß-glucosidase in *Trichoderma reesei.* FEMS Microbiol. Letters 34 : 291-295.

Vaheri, M., Leisola, M. and Kauppinen, V. (1979) Transglycosylation products of cellulase system of *Trichoderma reesie.* Bio/Technology Lett 1 : 41-46.

Vansatha, N., Thompson, L.D., Rhodes, C., Banner, C., Nagle, C., Nagle, J. and Filpula, D. (1984) Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame between the regions coding for signal sequence and mature protein. J. Bact. 159 : 811-819.

Vieira, J. (1985) United States Biochemical Company technical literature accompaning the vectors pTZ 18/19.

Voller, A., Bidwell, D. and Barlett, A. (1976) Microplate immunoassays for the immunodiagnosis of virus infections, pp.506-512. In: N. Rose and H. Freedman (eds.), Manual of clinical immunology, first ed. A.S.M., Washington, D.C.

Wakarchuk, W.W., Kilburn, D.G., Miller, R.C. and Warren, R.A.J. (1984) The preliminary characterization of the ß-glucosidases of *Cellulomonas fimi.* J. Gen. Microbiol. 130 : 1385-1389.

Wakarchuk, W.W. (1987) The molecular cloning and characterization of a ß-glucosidase gene from an *Agrobacterium.* Ph. D. Thesis. The University of British-Columbia.

Warren, R.A.J., Beck, C.F., Gilkes, N.R., Kilburn, D.G., Langsford, M.L., Miller, R.C. Jr., O'Neill, G.P., Scheufeus, M. and Wong, W.K.R. (1986) Sequence conservation and region shuffling in an endoglucanase and an exoglucanase from *Cellulomonas fimi.* Proteins 1 : 335-341.

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

Woodward, J. and Wiseman, A. (1982) Fungal and other ß-D-glucosidases-their properties and applications. Enzyme Microb. Technol., 4 : 73-79.

Wong, R.W.K. (1986) The cloning and characterization of an endoglucanase gene of *Cellulomonas fimi.* Ph. D. Thesis, The University of British-Columbia.

Wong, R.W.K., Gerhard, B., Guo, Z-.M., Kilburn, D.G., Warren, R.A.J. and Miller, R.C.Jr. (1986) Characterization and structure of an endoglucanase gene *cen*A of *Cellulomonas fimi.* Gene 44 : 315-324.

Wong, R.W.K., Curry, C., Davies, R.W. and Kilburn, D.G. (1988) Wood hydrolysis by *Cellulomonas fimi* endoglucanase and exoglucanase coexpressed as secreted enzymes in *Saccharomyces cerevisiae.* Bio/Technology 6 : 713-719.

Worthington, C.C. (1988) Worthington Enzyme Manual, Worthington Biochemical Corporation, Freehold, New Jersey, USA 07728.

Yamada, K. and Komagata, K. (1972) Taxonomic studies on coryneform bacteria. V. Classification of coryneform bacteria. J. Gen. Appl. Microbiol. 18 : 417-431.

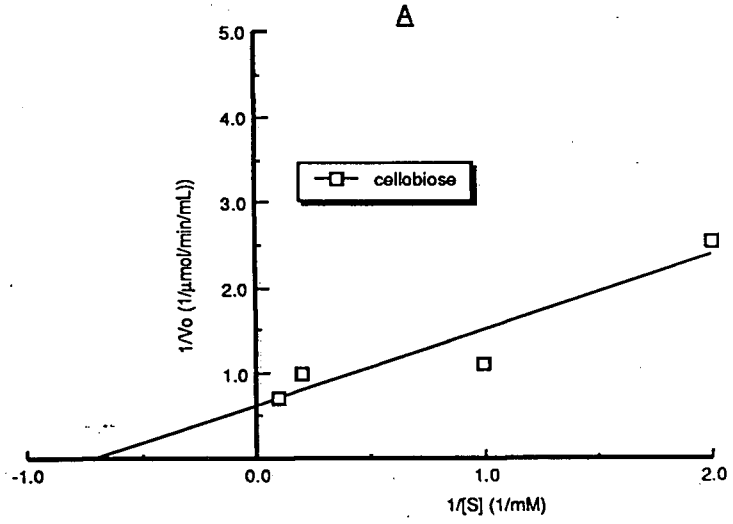
Yamada, K., Kinoshita, S., Tsunoda, T. and Aida, K. (1972) The microbial production of amino acids. Wiley, New York.

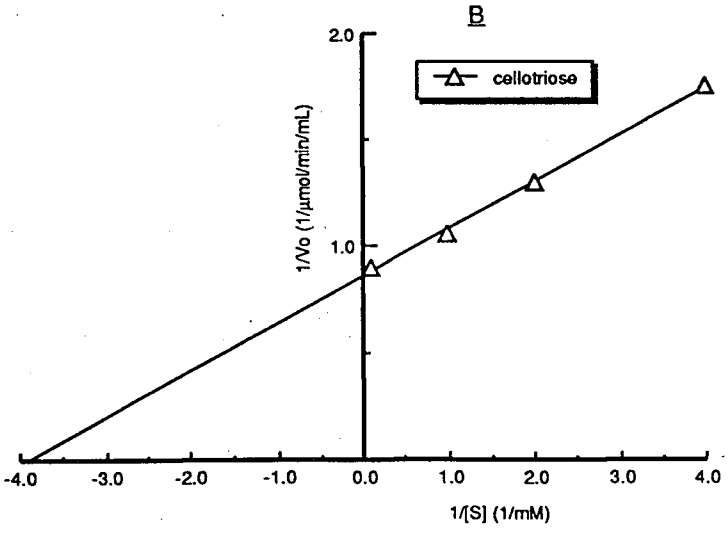
Yannish-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of M13 mp and pUC vectors. Gene 33 : 103-119.

Yeh, P., Oreglia, J., Prevost, F. and sicard, A.M. (1986) A shuttle vector system for *Brevibacterium lactofermentum.* Gene 47 : 301-306.

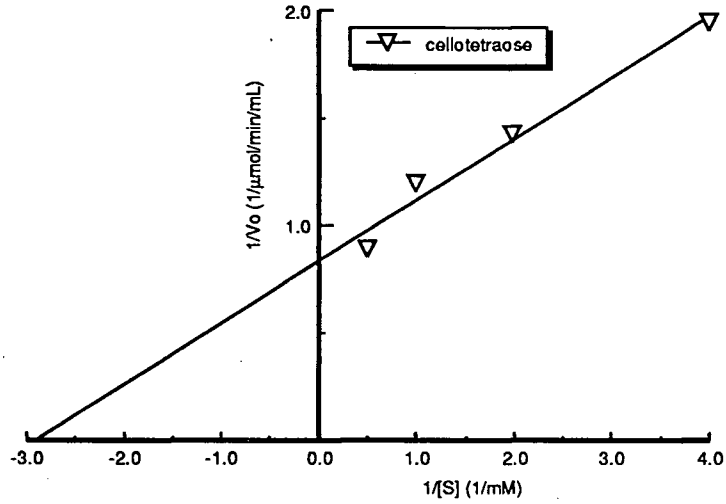
Yoshihama, M., Higashiro, K., Rao, E.A., Akedo, M., Shanabruch, W.G., Follettie, M.T., Walker, G.C. and Sinskey, A.J. (1985) Cloning vector system for *Corynebacterium glutamicum* . J. Bact. 162 : 591-597.

Appendix 1. The double-reciprocal plots of reciprocal rates (1/Vo) against reciprocal substrate concentration (1/[S]) are given for the hydrolysis of cellobiose (A) and higher cellodextrins (B, C, D) by the purified CbgA as described in section 2.7.2. The substrates are shown in shadowed legends. The intercept on the ordinate (1/Vo) axis corresponds to $1/V_{max}$ and the intercept on the abcissa (1/[S]) axis corresponds to $-1/K_m$. The inhibition of CbgA activity by glucose on the hydrolysis of pNPG at various substrate concentrations is shown by plotting the reciprocal rate (1/Vo) against glucose inhibitor concentration ([i]) (E). The crossing point of all three lines reflected on the abcissa is equal to $-K_i$. The hydrolysis of p-nitrophenyl derivatives (pNPG and pNPC) by CbgA is shown by plotting the continuous increase in optical density (410 nm) in function of time (F). The substrate and various concentrations are indicated in the legend.

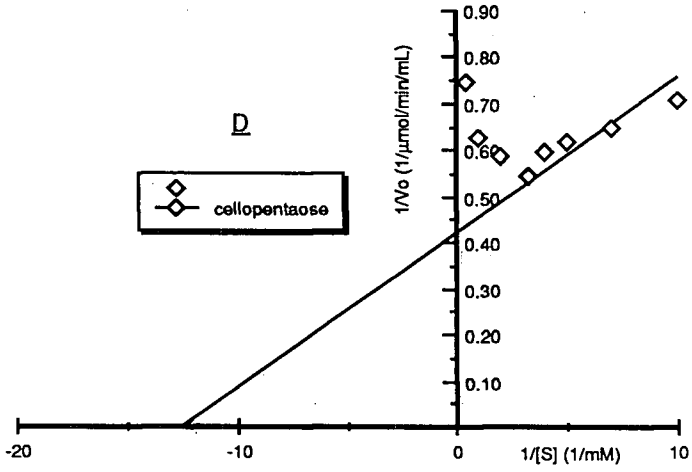




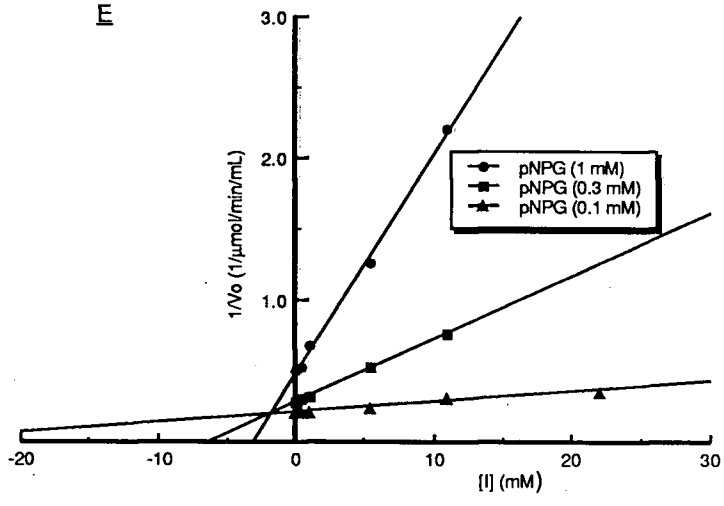
C


cellotetraose
1/Vo (1/µmol/min/mL)
2.0
1.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0
1/[S] (1/mM)

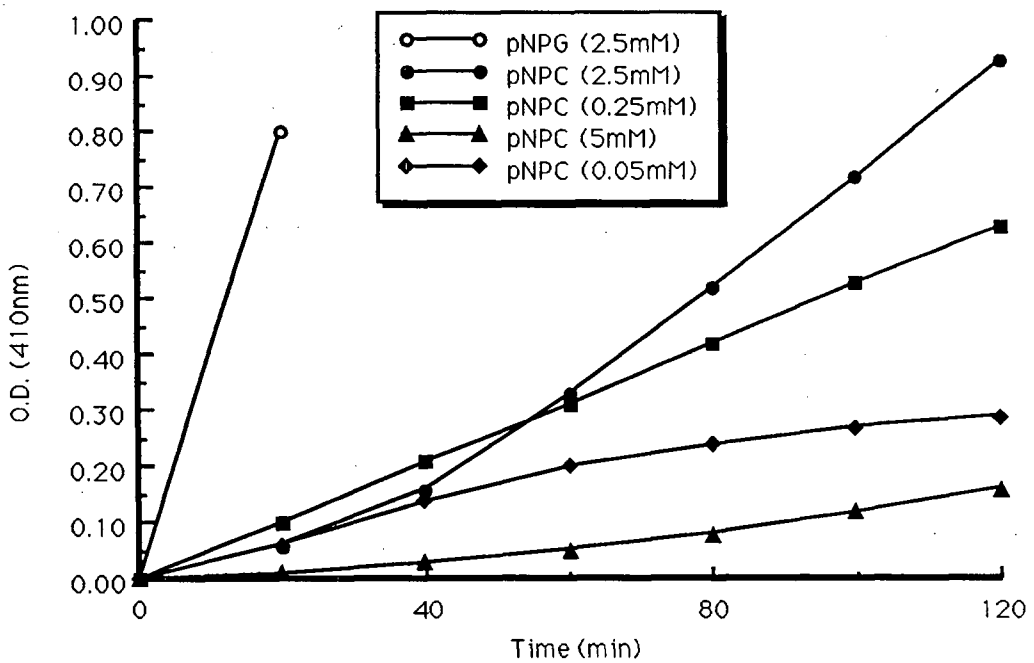
D


cellopentaose
1/Vo (1/µmol/min/mL)
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
-20
-10
0
10
1/[S] (1/mM)

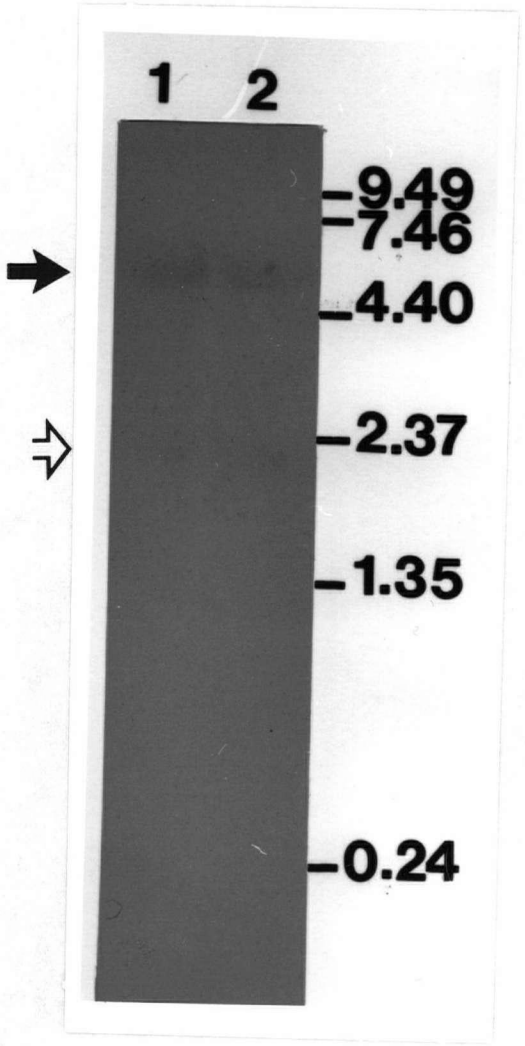
E


pNPG (1 mM)
pNPG (0.3 mM)
pNPG (0.1 mM)
1/Vo (1/µmol/min/mL)
3.0
2.0
1.0
-20
-10
0
10
20
30
[I] (mM)

E


pNPG (2.5mM)
pNPC (2.5mM)
pNPC (0.25mM)
pNPC (5mM)
pNPC (0.05mM)
O.D. (410nm)
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
40
80
120
Time (min)

Appendix 2. Detection of specific transcript by Northern blot analysis. RNA isolated from *C. fimi* CMC grown cells was denatured with formaldehyde, fractionated on a formaldehyde gel containing 1% (w/v) agarose and electrotransfered to a Biotrans membrane (Pall, Inc.). Hybridization was done with ^{32}P-labelled *Xho*I-*Stu*I probe (9.2 X 10^4 cpm) to : lane 1, 28.6 µg of RNA; lane 2, 14.3 µg of RNA. M_r markers are shown (nt). The membrane was exposed for 3 days at -70°C with intensifying screens. The location of the detected transcripts is shown by arrows.



Appendix 3. S1 nuclease protection analysis of *cbg*A transcript. After hybridisation of the *Stu*I-*Xho*I *cbg*A specific labeled DNA probe with RNA from *C. fimi* CMC-grown cells, the DNA-RNA hybrid was digested with various amount of S1 nuclease and analysed in a 8% polyacrylamide-7M urea containing gel alongside probe sequenced by the base-specific chemical cleavage method. The chemical sequencing ladders of the probe is shown. Protection of the probe by 30 µg of RNA digested with ; lane 1, 50 U; lane 2, 150 U; lane 3, 300 U of S1 nuclease. Lane 4, 30 µg of yeast tRNA digested with 300 U of S1 nuclease, lane 5, probe alone.