CELLULASE GENE TRANSCRIPTION

IN

CELLULOMONAS FIMI AND AN *AGROBACTERIUM*

by

NORMAN MICHAEL GREENBERG

B.Sc., the University of Toronto, 1982

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)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March, 1988

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
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date March 15, 1988

ABSTRACT

Transcriptional analysis was used to investigate the molecular mechanisms which effect cellulase gene expression in the gram-positive bacterium *Cellulomonas fimi* strain ATCC 484 and the gram-negative bacterium *Agrobacterium* sp. strain ATCC 21400. The *cen*A, *cex* and *cen*B genes of *C. fimi* encoding the extracellular β-1,4-endoglucanase, EngA (EC 3.2.1.4; M_r 48,700), the extracellular β-1,4-exoglucanase, Exg (EC 3.2.1.91; M_r 47,300) and the extracellular β-1,4-endoglucanase EngB (EC 3.2.1.4; M_r 110,000) respectively, were characterised. By northern blot analysis, *cen*A mRNA was detected in *C. fimi* RNA prepared from glycerol- and carboxymethylcellulose (CMC)-grown cells but not in RNA from glucose-grown cells. The *cex* mRNA was found only in RNA from CMC-grown cells. The *cen*B mRNA was found in all three preparations of RNA. Therefore, the expression of these genes is subject to regulation by the carbon source provided to *C. fimi*. High resolution nuclease S1 protection studies with unique 5'-labeled DNA probes and *C. fimi* RNA isolated *in vivo*, were used to map the 5' termini of *cen*A and *cex* mRNAs. Two *cen*A mRNA 5' ends, 11 bases apart, mapped 51 and 62 bases upstream of the *cen*A start codon, suggesting that *in vivo*, *cen*A transcription was directed from two promoters in tandem. The *cex* mRNA 5' end was found to map 28 bases upstream of the *cex* start codon. Using S1 mapping with unlabeled DNA probes and *C. fimi* RNA which had been isolated

in vivo but which had been 5'-labeled *in vitro* with vaccinia virus capping enzyme confirmed that true transcription initiation sites for *cen*A and *cex* mRNA had been identified. The S1 mapping revealed mRNA 3' termini 1,438, 1,449, and 1,464 bases from the major *cen*A start site, and one 3' terminus 1,564 bases from the major *cex* mRNA start site, in good agreement with the northern blot data. High resolution S1 studies were also used to show that abundant mRNA 5' ends mapped upstream of the *cen*B start codon in RNA prepared from CMC-grown cells, while less-abundant species mapped 52 bases closer to the ATG codon in RNA prepared from *C. fimi* grown on any one of the three substrates. These results seem to indicate a tandem promoter arrangement with an ATG-proximal promoter directing low-level constitutive *cen*B transcription and a more distal promoter directing higher levels of *cen*B transcription as a result of *C. fimi* growth on cellulosic substrate. Steady- state levels were determined for *cen*A, *cex* and *cen*B mRNAs with RNA prepared from glycerol-, glucose-, and CMC-grown cultures of *C. fimi* in slot-blot hybridisations with radiolabeled oligodeoxyribonucleotide probes. A *cex*-linked gene (*clg*) was identified by sequence inspection and S1 mapping.

Transcripts of the *abg* gene encoding the β-glucosidase (Abg, EC 3.2.2.21; M_r 50,000) of *Agrobacterium* sp. strain ATCC 21400 were also characterised. Northern blot analysis of *Agrobacterium* RNA revealed the size of the *in vivo abg*

mRNA was approximately 1,500 bases in length. High resolution S1 mapping determined *abg* mRNA 5' ends 22 bases upstream of the *abg* ATG codon and 3' ends 71 bases downstream of the *abg* stop codon.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS, NOMENCLATURE AND SYMBOLS

A	Adenine
aa	Amino acid(s)
abg	Gene encoding the *Agrobacterium* β- glucosidase
Abg	The *abg* gene product
ATCC	American Type Culture Collection
Ap	Ampicillin
bp	Base pair(s)
C	Cytosine
*cen*A	Gene encoding the endo-β-1,4-glucanase A of *Cellulomonas fimi*
*cen*B	Gene encoding the endo-β-1,4-glucanase B of *Cellulomonas fimi*
cex	Gene encoding the exo-β-1,4-glucanase of *Cellulomonas fimi*
CIAP	Calf intestinal alkaline phosphatase
CMC	Carboxymethylcellulose
DNA	Deoxyribonucleic acid
ds	Double-stranded
dNTP	deoxyribonucleoside triphosphate
EDTA	Ethylenediaminetetraacetic acid
EngA	The *cen*A gene product
EngB	The *cen*B gene product
Exg	The *cex* gene product

G	Guanine
HEPES	*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid
kb	1000 base pairs
kDa	1000 daltons
lacZ'	The first 78 amino acids of the *E. coli* β-galactosidase including the operator and promoter regions of the gene
mA	Milliamperes
MCS	Multiple cloning site
MOPS	Morpholinepropanesulfonic acid
M_r	Relative molecular mass
N	Any (deoxy-) ribonucleotide
nt	Nucleotides
PAGE	Polyacrylamide gel electrophoresis
PNK	Polynucleotide kinase
PPi	Pyrophosphate
r	Resistant
RBS	Ribosome binding site/sequence
SDS	Sodium dodecyl sulfate
ss	Single-stranded
SSC	Standard saline citrate
T	Thymine
TAE	Tris/Acetate/EDTA buffer
TE	Tris/EDTA buffer

ACKNOWLEDGEMENTS

I wish to thank Dr. R.C. Miller, Jr. for being my graduate supervisor and Drs. R.A.J. Warren, D.G. Kilburn, and P.P. Dennis for sharing their knowledge and for their advice throughout the course of this work. I also wish to thank Drs. N.R. Gilkes, J.T. Beatty, P. Beguin and G.B. Spiegelman for all their help. To my fellow graduate students, the staff and faculty of the Department of Microbiology - thanks for the memories. I am grateful to the Natural Sciences and Engineering Research Council of Canada for supporting the UBC cellulase research group and for my 1984-85 and 1985-86 postgraduate fellowships. This one's for the mishpocha.

1. INTRODUCTION

1.1. Background

Cellulolytic microorganisms elaborate enzymes, broadly classified as cellulases, which can hydrolyse β-1,4-glucosidic bonds. Many cellulolytic bacteria and fungi have now been identified and are being characterized (for recent reviews, see Beguin, *et al.*,1987 and Coughlan, 1985). Characterization of the individual enzymatic components involved in cellulolysis has been complicated by the multiplicity of cellulase activities most cellulolytic microorganisms produce and by the mechanisms which regulate cellulase expression and functions (Beguin, *et al.*, 1977; Coughlan, 1985; Duong *et al.*, 1983). On the basis of phenotypic observations and activity complementation studies, repression and induction have been suggested as mechanisms regulating cellulase biosynthesis and end-product inhibition has been suggested to affect enzyme activity (Beguin, *et al.*, 1977; Coughlan, 1985; Postma, 1986). It is generally believed that enhanced cellulase production, through genetic engineering, will facilitate the development of an economically feasible process of cellulose hydrolysis to produce fuels from biomass and agricultural cellulosic waste material (Beguin, *et al.*, 1987; Eveleigh, 1983; Ryu and Mandels, 1980). Therefore, the molecular cloning of cellulase encoding genes is currently the topic of intense

investigation (Beguin *et al.*, 1987).

While the structures of cellulase encoding genes, and the unique properties of their protein products are being determined (building the genetically enhanced cellulolytic system to convert cellulose to glucose is a driving force), the molecular mechanisms which govern their native expression have remained relatively uncharacterized. Since molecular cloning has been used to isolate cellulase genes from *Cellulomonas fimi* strain ATCC 484 (reviewed by Beguin, *et al.*, 1987) and a β-glucosidase gene from *Agrobacterium* sp. strain ATCC 21400 (Wakarchuk *et al.*, 1986), the availability of these cloned genes facilitates a study of their expression at the transcriptional level. Such a characterization is also of interest since transcription in either of these bacteria had not been previously investigated.

1.2. Cellulose and cellulases

Cellulose, the major cell wall and structural polysaccharide of plants, is the most abundant organic material on earth. Cellulose occurs as long chains of glucose residues, up to 14,000 units in length, held together by β-1,4-glucosidic bonds. The basic repeating unit of cellulose is the disaccharide cellobiose (see Fan *et al.*, 1980; Lehninger, 1975; Thomas, 1983,).

Chains of cellulose are organized in bundles of parallel

chains to form fibrils. These fibrils display both highly ordered structure, the crystalline regions, and less ordered structure, the amorphous regions. In its natural state, cellulose is found as semi-crystalline, insoluble fibrils cemented together within a complex organic matrix. This matrix is primarily composed of two components: hemicelluloses, which are β-1,4-linked polymers of D-xylose with side chains of arabinose and other sugars; and lignin, which is a polymer of aromatic alcohols (Gardner and Blackwell, 1974; Rees *et al.*, 1982).

The cellulolytic process appears to involve several types of enzymes, of which the endoglucanases, exoglucanases and β-glucosidases are perhapse the best characterized. The endoglucanases (E.C.3.2.1.4) hydrolyse the internal β-1,4-glucosidic linkages in the cellulose chains, generating new, non-reducing ends. The exoglucanases attack the non-reducing ends of the cellulose chains created by endoglucanase hydrolysis. Exoglucanases may be β-1,4-cellobiohydrolases (E.C. 3.2.1.91) which effect the release of cellobiose, or they may be β-1,4-glucan glucohydrolases (E.C. 3.2.1.74) which release glucose from the non-reducing ends of cellulose and cellodextrins. Certain endoglucanases and exoglucanases can act synergistically to hydrolyse semi-crystalline cellulosic substrates (Coughlan, 1985).

The short oligosaccharide products of endo- and exoglucanase hydrolysis can be further hydrolysed to glucose

through the action of β-glucosidases (E.C.3.2.1.21). While the endo- and exoglucanases are usually found as extracellular enzymes, β-glucosidases are primarily cell-associated. The action of β-glucosidases may be the rate limiting factor in the hydrolysis of cellulose to glucose; it may also relieve the end-product inhibition of the endo- and exoglucanases by cellobiose (Han and Srivinasan, 1969; Shewale, 1982). While a number of β-glucosidases have been characterized, only three have been both cloned and sequenced (Wakarchuk, 1987; Wakarchuk, *et al.*, 1986,1988; Kohchi and Toh-e, 1985).

While other components, both enzymatic and non-enzymatic, are associated with the cellulolytic systems of fungi and bacteria, their roles in cellulolysis are still not clearly understood and are beyond the scope of this introduction. Again the reader is refered to recent reviews (Coughlan, 1985, Enari and Niku-Paavola, 1987).

The following is a brief account of the *C. fimi* and *Agrobacterium* sp. genes which were subjects of this study.

1.3. The *cen*A, *cex* and *cen*B genes of *Cellulomonas fimi*

C. fimi is one of the most investigated cellulolytic bacteria (Beguin *et al.*, 1987). It is a gram-positive, rod-shaped mesophile with DNA of 72 mole% G+C (Bergey *et al*, 1974; Keddie, 1974; Stackebrant and Kandler, 1979). Endoglucanase, exoglucanase and cellobiase activities are induced when *C. fimi* is grown on cellulosic matrerial

(Langsford, *et al.*, 1984). To date, 5 genes encoding cellulolytic enzymes have been isolated from *C. fimi* and cloned into *E. coli* (Bates, 1987; Beguin *et al.*, 1987; Gilkes *et al.*, 1984; O'Neill, *et al.*, 1986; Owolabi *et al.*, 1988; Whittle, *et al.*, 1982; Wong *et al.*, 1986; B. Moser, personal communication). Of these five genes, *cen*A, *cex* and *cen*B, are the subjects of this study and are described below. The *cen*C gene encoding endoglucanase C is currently being characterized by Bernhard Moser and the *cbg* gene encoding a cellobiase is currently being characterized by Francois Paradis, both at UBC. All five genes were isolated as individual clones. The *cen*A, *cex* and *cen*B genes were not found to map together after screening a *C. fimi* lambda library. Whether or not the *cen*C and *cbg* genes map together or with the other three genes is not yet known. The products of these five genes are all believed to play a role in *C. fimi* cellulolytic growth.

The *cen*A gene encodes the major endoglucanase of *C. fimi* (Gilkes et al., 1984b; Wong, et al., 1986). Its nucleotide sequence has been determined (Wong *et al.*, 1986). It encodes a polypeptide, EngA, of 449 amino acids, initiating at an ATG codon and terminating with a TGA codon positioned 1347 nucleotides downstream. The native form of EngA is extracellular and the *cen*A nucleotide sequence predicts a leader peptide of 31 amino acids which functions to export EngA to the periplasm of *E. coli* (Wong *et al.*, 1986). Both

the native and recombinant forms of EngA have been purified to homogeneity (Gilkes *et al.*, 1984; Langsford, personal communication; Wong, 1986). The nucleotide sequence preceding the ATG start codon of the *cen*A structural gene contains a putative ribosome-binding site (Wong, et al., 1986; Shine and Dalgarno, 1974) which appears to function in *E. coli* clones harbouring the recombinant gene (Gilkes *et al.*, 1984a; Wong,et al., 1986 and W.K.R. Wong, personal communication). Transcription of the *cen*A gene in *E. coli* is not directed by the endogenous *C. fimi* 5' flanking sequence (Wong, et al., 1986).

The *cex* gene encodes the major exoglucanase of *C. fimi* (Gilkes *et al.*, 1984a,b; Langsford *et al.*, 1984; O'Neill *et al.*, 1986a,b,c). Its nucleotide sequence has been determined (O'Neill, 1986a). It encodes a polypeptide, Exg, of 484 amino acids, initiating with an ATG codon and terminating with a TGA codon 1452 bases downstream. The *cex* sequence predicts a 41 amino acid leader peptide which functions to export Exg to the periplasm of *E. coli* (Gilkes *et al.*, 1984a; O'Neill *et al.*, 1986b) Both the native and recombinant forms of Exg have been purified to homogeneity. The sequence preceding the ATG start codon of the *cex* structural gene contains a putative ribosome-binding site (O'Neill *et al.*, 1986a; Shine and Dalgarno, 1974) which appears to be function in *E. coli* clones harbouring the gene (O'Neill *et al.*, 1986a and G. O'Neill, personal

communication). Transcription of the recombinant *cex* gene in *E. coli* is not directed by the endogenous *C. fimi* DNA 5' flanking sequence. Replacement of the 5' flanking endogenous *C. fimi* DNA sequences with a heterologous *E. coli* promoter and ribosome-binding site has facilitated expression of Exg in *E.coli* to levels exceeding 20% of the total cellular protein (O'Neill *et al*, 1986b).

The *cen*B gene encodes the endoglucanase B (EngB) of *C. fimi* (Gilkes *et al.*, 1984a,; Owolabi *et al.*, 1987). Although the gene has been cloned in *E. coli*, only a partial DNA sequence has so far been determined. It encodes a polypeptide of about 110,000 KDa (Owolabi *et al.*, 1988) predicting a coding sequence of at least 3000 bp which appears to initiate with an ATG codon (Owalabi *et al.*, 1988). The nucleotide sequence determined for *cen*B also predicts a leader peptide of 33 amino acids which functions to export recombinant EngB to the periplasmic space in *E. coli* (Gilkes *et al.*, 1984a; Owolabi *et al.*, 1988). Attempts have been made to purify native and recombinant EngB to homogeneity (Owolabi *et al.*, 1988). The *C. fimi* DNA sequence preceding the ATG codon contains a putative ribosome-binding site which appears to function in *E. coli* (Owolabi, 1988). However, transcription of the *cen*B gene in *E. coli* is not directed from the endogenous *C. fimi* 5' flanking sequences (Greenberg *et al.*, 1987b; Owolabi *et al.*, 1988).

1.4. The *abg* gene of *Agrobacterium* sp. strain ATCC 21400

A gram-negative, rod-shaped mesophile, previously classified as *Alcaligenes faecalis* (Han and Srinivasan, 1969), was isolated from a mixed population of *Cellulomonas* and other species growing on cellulose (Han and Srinivasan, 1968,1969). This species, now classified by the ATCC as an *Agrobacterium*, produces a β-glucosidase which is very active on cellobiose (Day and Withers, 1986; Wakarchuk *et al.*, 1986). *C. fimi* produces at least two different β-glucosidases (Wakarchuk, *et al.*, 1984), of which the major one is probably a phospho-β-D-glucosidase. To date only one of the genes, *cbg*, encoding the cellobiase Cbg has been cloned and remains relatively uncharacterized. The second gene has not yet been isolated. However, the gene (*abg*) for the *Agrobacterium* β-glucosidase was cloned (Wakarchuk *et al.*, 1986) with the hope of supplementing the cloned *C. fimi* endo- and exocellulase genes in a reconstructed, genetically enhanced cellulolytic system. Whereas the *C. fimi* cellobiase gene(s) had not yet been well characterized at the molecular level, the *abg* gene had been well studied making *abg* more suitable for these investigations.

The *abg* gene encodes a polypeptide, Abg, of 459 amino acids initiating with an ATG codon and terminating with a TGA codon 1377 nucleotides downstream (Wakarchuk *et al.*,1988). Both the native and recombinant forms of the enzyme have been purified to homogeneity, and the enzyme has been found not to

possess a leader peptide. The sequence preceding the ATG codon contains a putative ribosome-binding site which functions in *E. coli*. The adjacent *Agrobacterium* promoter sequences appear to function in *E. coli* (Wakarchuk *et al.*, 1986 and 1988).

The general features of the *cen*A, *cex*, *cen*B and *abg* genes are summarized in Table 1.

TABLE I. Characteristics of the *cen*A, *cex*, *cen*B and *abg* structural genes*

Source of gene	Gene	Start codon	ORF	Stop codon	#Amino Acids† Leader peptide	#Amino Acids† Mature peptide	Mr†† Predicted	Mr†† Observed
Cellulomonas fimi								
ATCC 484								
	*cen*A	ATG	1347	TGA	31	418	43,800¥	48,700¥
	cex	ATG	1452	TGA	41	443	47,100¥	47,300¥
	*cen*B	ATG	3000	N/D	33	1000±	N/D	110,000¥
Agrobacterium sp.								
ATCC 21400								
	abg	ATG	1377	TGA	N/D	459	50,980	50,000

* for references, see text

† values determined from nucleotide and protein sequence data

†† i) predicted values for mature form proteins based on the *cen*A, *cex* and *abg* gene sequences
ii) observed values from SDS PAGE of purified (mature form) recombinant proteins (N.R. Gilkes, unpublished observations)

¥ mature form (i.e. lacking leader peptide signal sequence)

± predicted values (see text)

N/D not determined

2. MATERIALS AND METHODS

2.1. Bacterial strains, phages and plasmids

A list of the bacterial strains, plasmids and phages used in these studies is given in Table II. Stock cultures of bacteria were maintained at -20°C or -80°C in LB or M9-glucose medium containing 15% glycerol.

2.2. Enzymes and reagents

Restriction endonucleases and DNA and RNA modifying enzymes were purchased from Bethesda Research Laboratories, Inc., Pharmacia P-L Biochemicals, Boehringer-Mannheim Inc., and New England Biolabs. Radionuclides were from New England Nuclear (Dupont-NEN) and Amersham Inc. All other chemicals were of reagent grade or higher and were purchased from commercial suppliers.

2.3. Media and growth conditions

C. fimi Strain ATCC 484 was grown in basal medium (Stewart and Leatherwood, 1976) supplemented with 0.2% (w/v) glycerol, 0.2% (w/v) glucose, 1.0% (w/v) CMC (Sigma; low viscosity) as carbon source(s). *E. coli* strains were grown in 2 X YT medium (Messing, 1983) or M9-glucose medium (Miller, 1972). *Agrobacterium sp.* Strain ATCC 21400 was grown in low salt-LB (LSLB) medium which contained per litre: 0.5 g Nacl, 5 g Bacto-yeast extract and 10 g Bacto-tryptone.

Table II . Bacterial strains, phages and plasmids

Bacterial strain	Genotype	Reference
ATCC 484	cellulose utilization	Stackebrandt and Kandler,1979.
ATCC 21400	abg	Han and Srivinasan, 1969
E. coli JM83	ara Δ(lac-proAB) rpsL ø80 lacZΔM15	Yanisch-Perron et al., 1985
E. coli JM101	supE thi Δ(lac-proAB) [F' traD36 proAB lacI^q lacZΔM15]	Yanisch-Perron et al., 1985

Phage	Genetic characteristics	Reference
M13mp8::Bam0.8kb	cenA	Wong, 1987
M13mp11	lac	Messing, 1983
λcI857	cIindIts857 Sam7	Maniatis et al., 1982

Plasmid	Genetic characteristics	Reference
pABG5	Ap^r lacZ' abg	Wakarchuk et al., 1988
pBR322	Ap^r Tc^r	Bolivar, et al. 1977
pEC2.1	Ap^r cenA	Wong et al., 1986
pTZ19U-B	Ap^r lacZ' orif1 abg	Wakarchuk, 1987
pUC12	Ap^r lacZ'	Messing, 1983
pUC12A25	Ap^r lacZ' cex2585	O'Neill, 1986
pUC13	Ap^r lacZ'	Messing, 1983
pUC13Bam31	Ap^r lacZ'Δcex397	O'Neill, 1986
pUC13Δ9R5	Ap^r lacZ'abg	Wakarchuk, 1987
pUC18	Ap^r lacZ'	Yanisch-Perron et al., 1985
pUC19	Ap^r lacZ'	Yanisch-Perron et al., 1985
pUC19C3PS	Ap^r lacZ'cenB	Owolabi, 1988

All strains were grown at 30°C, except where otherwise indicated. When solid medium was required, agar (Difco Laboratories) was added to 1.5% (w/v) except for basal medium containing CMC, in which 1.0% agar was used. When appropriate, ampicillin (Sigma) was added to 100 μg ml^{-1} in liquid or solid medium.

2.4. RNase-free work

Chemicals and reagents used for RNA work were purchased soley for this purpose and were kept separate from regular laboratory supplies. All glasssware used for RNA work was either baked at 300°C for 3 h or was bought as disposable labware. When appropriate, solutions were treated with 0.2% (v/v) diethylpyrocarbonate as described previously (Ehrenberg *et al.*, 1976; Maniatis *et al.*, 1982). All plastics (pipette tips and microfuge tubes) were sterilized by autoclaving without further pretreatment.

2.5. RNA extraction

RNA was prepared from all bacterial strains by a modification of published procedures (see Miller *et al.*, 1981; Kennell and Bicknell, 1973; Greenberg *et al.*, 1987). Briefly, cultures (up to 100 ml) were rapidly chilled on ice and were transfered to pre-chilled centrifuge tubes (-20°C). Cells were recovered by centrifugation for 5 min at 6,000 x *g*. Cells were then resuspended in 1/25 to 1/10 volume of 50

mM Tris-HCl (pH 6.8 at 20°C)-2 mM EDTA-1.0% SDS, transfered to a clean centrifuge tube and placed immediately into a boiling water bath for up to 2 min. The tubes were chilled on ice for 5 min and 1/2 vol of ice-cold 5M NaCl was added and mixed briefly on a Vortex mixer. After 5 min on ice, the resultant slurry was centrifuged for 10 min at 10,000 x *g* and the cleared supernatant fluid was decanted to a 30 ml Corex (Corning Glass Works) glass tube. The nucleic acids were precipitated with 2.5 vols of 95% ethanol at -20°C for 12 to 16 h and recovered by centrifugation for 20 min at 10,000 x *g*. The pellets were washed with 70% ethanol at -20°C and redissolved in 0.4 to 2.0 ml of 10 mM Tris-HCl (pH 7.5)-40 mM Nacl-5 mM $MgCl_2$. Samples were treated with 5 units 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. The organic phases were combined and back extracted with 0.2 to 2 ml of TE (pH 7.5). The aqueous phases were pooled and RNA was recovered by precipitation with 2.5 to 3 vols of 95% ethanol and centrifugation for 10 min at 10,000 x *g*. The pellets were washed with 70% ethanol and redissolved in 20 mM $NaPO_4$ (pH 6.5)-1 mM EDTA (RNA storage buffer). The RNA preparations were compared for similar banding patterns after analytical electrophoresis on agarose gels and subsequent staining with ethidium bromide. RNA concentrations were determined by A_{260}, and samples were divided into aliquots and stored at

-70°C.

2.6. DNA extraction and purification

Plasmid DNA was isolated by a modification of the alkaline-lysis procedure (Birnboim and Doly, 1979). When required for the preparation of high specific-activity probes, DNA was further purified by centrifugation to equilibrium in CsCl density gradients containing ethidium bromide (Maniatis *et al*, 1982).

2.7. Preparation of ^{32}P labeled DNA

To end-label DNA restriction fragments, plasmid DNA was digested with restriction enzyme at the temperature recommended by the supplier. Digestions were monitored by agarose gel electrophoresis. Digested DNA was extracted twice with phenol- chloroform (1:1) and precipitated with 95% ethanol. For 5' end-labeling reactions, the DNA was treated with calf intestinal alkaline phosphatase (CIAP) and then labeled with [γ-^{32}P]-ATP (3,000-7,000 Ci mmol^{-1}) and T4 polynucleotide kinase (PNK) as described previously (Maniatis *et al.*, 1982). The 3' ends were labeled with an appropriate [α-^{32}P] dNTP (3,000 Ci mmol^{-1}) and the Klenow fragment of DNA polymerase I as described previously (Maniatis *et al.*, 1982). For nick- translations, plasmid DNA was incubated with DNase I and [α-^{32}]dNTPs and DNA polymeraseI and recovered from Sephadex G50 columns as described previously (Maniatis *et*

al., 1982). Incorporation of label was monitored by liquid scintillation spectrometry in an ISOCAP-300 (Nuclear Chicago).

2.8. Preparation of hybridisation probes

^{32}P-end-labeled dsDNA was digested with an appropriate restriction endonuclease to liberate fragments uniquely labeled at one end (strand specific). The digestions were routinely performed under conditions recommended by the suppliers. The hybridisation probes were purified by electrophoresis in 5% polyacrylamide gels, electroelution and precipitation with ethanol as described previously (Maniatis *et al.*, 1982). In some instances, yeast tRNA (20 $\mu g\ ml^{-1}$) was added as carrier in the final precipitation. Pellets were washed with 70% ethanol, dried briefly in air, and redissolved in TE (pH 7.5). Samples were removed for quantitation by liquid scintillation counting. Typical specific activities were 5×10^6 to 1×10^7 cpm μg^{-1} of DNA probe.

2.9. DNA Sequencing and sequence analysis

DNA was sequenced by the base-specific chemical cleavage method (Maxam and Gilbert, 1980). Sequences were analysed with the SEQNCE program developed by Delaney Software (Vancouver, British Columbia, Canada) on an Amdahl™- V8 Mainfraim (UBC Computing Centre) or with the DNA Inspector II program (Textco, West Lebanon, N.H., USA) on an Apple™

Macintosh™ SE microcomputer.

2.10. DNA molecular weight standards

To prepare dsDNA molecular weight standards, lambda DNA was digested with *Hind*III and labeled at the 3'-termini as described above. To prepare ssDNA molecular weight standards, M13mp11 ssDNA was digested with *Hae*III (see von Gabain *et al.*, 1983) and then 5'-end labeled as described above. Fragment sizes were determined from published nucleotide sequences (van Wezenbeek *et al.*, 1980; Yanisch-Perron *et al.*, 1985). The 525-base fragment of M13mp11 arises from partial digestion (Greenberg *et al.*, 1987a,b).

2.11. Northern blot analysis

For Northern (RNA blot) analysis, 20 µg of total RNA were precipitated with ethanol, redissolved in 10 µl of 30 mM MOPS-1 mM EDTA-5mM sodium acetate (running buffer [pH 7] with 50% formamide and 2.2 M formaldehyde, heated for 15 min at 68°C, and cooled briefly on ice. Loading dye was added (to give 3% [w/v] Ficoll [Pharmacia] and 0.02% [w/v] bromphenol blue and xylene cyanol) and samples were electrophoresed alongside ^{32}P-labeled molecular weight markers in 1.0% agarose-6.6% formaldehyde gels at 20 to 40 mA with recirculation of running buffer. Nucleic acids were blotted to BioTrans membranes (Pall, Inc.) in 20X SSC (1X SSC is 0.15 M NaCl plus 0.15 M sodium citrate) for 12 to 16 h (Southern,

1975), and the membranes were allowed to dry in air and then baked at 80°C for 1 to 2 h. Prehybridisations and hybridisations were performed essentially by the protocols supplied with the membranes. Briefly, prehybridisations were done in 5X SSC-50% formamide-4mM PP_i-5X Denhardt buffer (Maniatis *et al.*, 1982) 10% dextran sulfate-250 µg of heat denatured (100°C for 5 min) salmon sperm DNA ml^{-1} . Incubations were for 1- 2 h at 42°C with constant agitation. For hybridisations, the probe DNA (2 to 10 ng ml^{-1}) and carrier were denatured together by heating. The probes were allowed to hybridise to filters for 16 to 24 h at 42°C with constant agitation. Blots were washed at 22°C with three changes of 2X SSC- 0.1% SDS and then at 60°C with three changes of 0.1X SSC- 0.1% SDS, dried to Whatman 3MM paper under vacuum and exposed to X-ray film (Eastman Kodak Co.) at -70°C with intensifying screens.

2.12. In vitro cap labeling of RNA

To characterise primary transcripts, total *C. fimi* RNA from CMC-grown cultures was labeled *in vitro* at the 5' end by using the vaccinia virus capping enzyme, guanylyltransferase, as described previously (Moss, 1981; Wich *et al.*, 1986). Briefly, up to 60 µg of total *C. fimi* RNA were labeled in 0.1 ml mixtures containing 25 mM Tris-HCl (pH 7.5)-2 mM $MgCl_2$-1 mM dithiothreitol-250 µCi of [α-^{32}P]-GTP (3,000 Ci $mmol^{-1}$)-10 to 25 units of guanylyltransferase. After 30 min at 37°C

the reaction was stopped by the addition of EDTA to 4mM and SDS to 0.2%. The RNA was extracted twice with phenol-chloroform (1:1) and precipitated twice from ethanol in the presence of 2M ammonium acetate. The RNA was finally recovered by ethanol precipitation from 0.3M sodium acetate. Typical specific activities were 1 X 10^5 to 3 X 10^5 cpm μg RNA^{-1}.

2.13. Hybrid protection analysis

The mRNA 5' and 3' termini were mapped with labeled DNA probes essentially as previously described (Favaloro *et al.*,1980; Berk and Sharp, 1977, 1978; Weaver and Weissman, 1979). Up to 30 μg of RNA were precipitated with end-labeled DNA probe, redissolved in 30 μl of hybridisation buffer (0.4 M NaCl-0.04 M sodium phosphate [pH 6.5]-0.4mM EDTA-80% formamide), heated for 15 min at 85°C, and held at 60°C (*C. fimi*) or 49°C (*Agrobacterium*) for 3 h. Samples were rapidly diluted with 300 μl of ice-cold S1 buffer (30 mM sodium acetate [pH 4.5]-28 mM NaCl-4.5 mM $ZnSO_4$) and treated with about 1000 units of S1 nuclease for 30 min at 37°C. The reactions were terminated by the addition of 75 μl of stop buffer (2.5mM ammonium acetate-50mM EDTA), and 20 μg yeast tRNA were added. The undigested nucleic acids were precipitated with 400 μl of isopropanol and recovered by centrifugation.

When capped RNA and unlabeled DNA probes were used in the mapping experiments, the procedure was modified as follows

(Which *etal.*, 1986): up to 50 µg of capped RNA were precipitated with up to 500 ng of unlabeled DNA probe, redissolved in hybridisation buffer, heated to 85°C, held at the appropriate hybridisation temperature for 3 h and then treated with S1 nuclease as described above. After the S1 treatment, the 0.33 ml samples were incubated with 25 ng of RNaseA for 15 min at 22°C to reduce the background of unhybridised, labeled RNA. This reaction was terminated by the addition of SDS to 0.25% and two extractions with phenol-chloroform (1:1). Trimmed hybrids were recovered by precipitation with ethanol.

After either of these procedures, pellets were dissolved in sequencing dye buffer (90% formamide, 0.02% [w/v] bromphenol blue and xylene cyanol) (Maniatis *et al.*, 1982) and heated to 90°C for 2 min. The redissolved samples were fractionated in polyacrylamide gels with appropriate size markers (see figure legends). The gels were dried to Whatman 3MM filter paper and exposed to X-ray film (Eastman Kodak) at -70°C with intensifying screens.

2.14. Synthetic oligodeoxyribonucleotide hybridisation probes

The oligodeoxyribonucleotide 30-mers used in these studies (Table III) were synthesised chemically on an Applied Biosystems 380A DNA Synthesiser by Tom Atkinson using phosphite triester chemistry, essentially as described (Atkinson and Smith, 1984). The oligodeoxyribonucleotide 30-

TABLE III. Oligodeoxyribonucleotide hybridisation probes

Gene	Oligodeoxyribonucleotide probe sequence
*cen*A	5' CAGCGCTGCGGCGGTTCTGCGGGTGGACAT 3'
cex	5' GTGGCCGGGTGCGGGCGTGGTCCTAGGCAT 3'
*cen*B	5' GACGAGCGTGCGTGGGACTTGGCGGAGCAT 3'

mers were separated from incomplete synthesis products by electrophoresis in a 16% polyacrylamide-7 M urea sequencing gel, located by UV-shadowing, and extracted from the gel by the crush and soak method (Atkinson and Smith, 1984). Oligomers were then further purified by reverse-phase chromatography using Sep-Pak C_{18} cartridges (Millipore/Waters Assoc., Milford, MA), and elution with 20% acetonitrile-80% water. The purified oligomers were lyophilised then stored at -20°C. For 5' end-labeling, 20 µl reaction mixtures containing 250 ng of oligomer in sterile distilled water (about 10 µl), 2 µl buffer (500 mM Tris-HCl [pH 8]-500 mM NaCl-100 mM $MgCl_2$), 250 µCi [γ-^{32}P]-ATP and 10 units of T4 polynucleotide kinase were incubated at 37°C for 30 min. The labeled oligomers were recovered from Sephadex G-50 columns and stored frozen at -20°C until needed. Specific activities were determined by liquid scintillation spectrometry in an ISOCAP 300 (Nuclear Chicago).

2.15. Slot Blot Hybridisations

The Schliecher & Schuell Minifold™ II (SRC 072/0) micro-sample filtration manifold was used for the quantitation of specific RNAs in samples. All samples were dissolved in 100 µl of DEPC-treated dH_2O and 300 µl of a solution of 6.15M formaldehyde-10X SSC, then incubated at 65°C for 15 min. Plasmid DNA standards (diluted with 10X SSC-6.15M formaldehyde) were brought to 4 µg total weight with carrier

RNA. Each 4 µg sample was loaded into a well of the manifold under vacuum (according to the method of G. Wahl, technical bulletin #371 accompanying the Minifold II, S&S™, 1983). Samples were washed with 400 µl of 10X SSC and the BioTrans membrane filter (Pall, Inc.) was air dried and then baked at 80°C for 1 to 2 h. Prehybridisations were for 1 to 2 h at 55°C in 1 ml of 6X SSC-5X Denhardt's solution (Maniatis *et al.*, 1982) per 25 cm^2 of membrane. Hybridisations were for 16 to 24 h at 55°C with 1 X 10^6 cpm of ^{32}P-labeled oligonucleotide probe ml^{-1}. The prehybridisations and hybridisations were perfomed in heat sealed Seal-a-meal™ bags (Sears, Inc.) under a sponge in a water- filled polypropelyne box submerged in a water bath. Filters were washed at room temperature (22°C) with 6X SSC followed by washings at 60°C with one change each of 2X SSC/ 0.1% SDS, 1X SSC/ 0.1% SDS and 0.1X SSC/ 0.1% SDS. Filters were wrapped in Saran-Wrap™ and exposed to X-ray film (Kodak) with intesifying screens at -70°C. Developed autoradiograms were scanned with a Helena Industries Quick Scan integrating densitometer (model 1111).

3. Results

3.1. Characterization of the *cenA* transcripts of *C. fimi*.

3.1.1. Regulation by carbon source and approximate length of *cenA* mRNA.

A qualitative northern blot analysis was used to determine the approximate length of *cenA* mRNA and the influence of the carbon source provided for *C. fimi* growth on *cenA* transcription. For these experiments, *C. fimi* RNA was prepared from cultures grown in basal medium supplemented with either 0.2%(w/v) glycerol (0.214 generation hr^{-1}), 0.2%(w/v) glucose (0.226 generation hr^{-1}) or 1.0%(w/v) CMC (0.193 generation hr^{-1}). The intragenic 123 bp *Sst*1-*Sal*1 fragment of *cenA* (Fig. 1, A) was used as the hybridisation probe. The probe, isolated from plasmid pNG101 (Fig. 2) was 5' end-labeled at the *Sal*1 site (Fig. 3).

A very intense signal was observed in hybridisations between *C. fimi* RNA extracted from CMC-grown cells and the labeled probe. The species of RNA detected by the probe was approximately 1400 bases in length (Fig. 4, lane 3). A less intense signal was detected in hybridisations between the probe and RNA from glycerol-grown cells (Fig. 4, lane 1). This signal also corresponded to a species of RNA of about 1400 bases in length. No signal was detected in hybridisations between the probe and RNA from glucose-grown

FIGURE 1. Partial restriction map of the *cen*A gene. Representation of the cloned 2.2-kilobase *Bam*H1- *Sma*1 segment of *C. fimi* DNA containing the *cenA* gene. The structural gene is shown as a boxed region, and is translated from left to right (5' -> 3'). A, *Sst*1-*Sal*1 northern blot probe; B, *Sma*1-*Sal*1 5' S1 probe; C, *Bgl*II-*Sma*1 3' S1 probe. The restriction endonucleases are abbreviated as follows: Bg, *Bgl*II; Bm, *Bam*H1; Sa, *Sal*1; Sm, *Sma*1; Ss, *Sst*1.

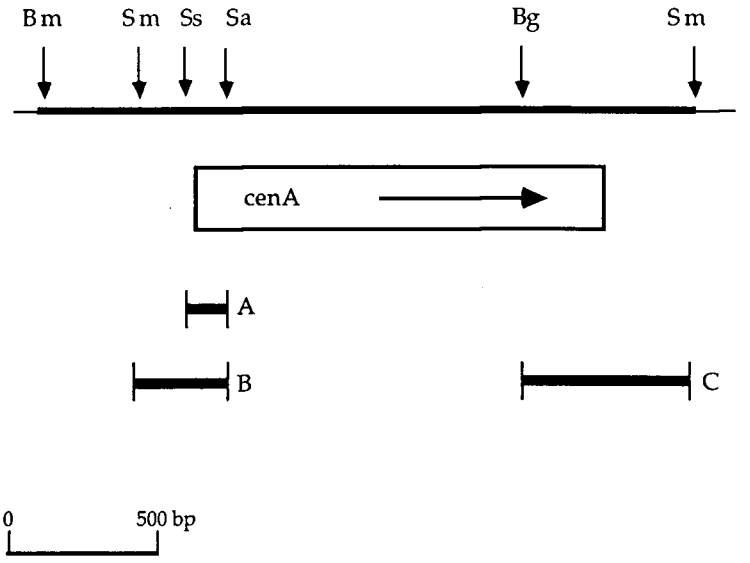


FIGURE 2. Subcloning 5' flanking and terminal *cenA* DNA. To facilitate the isolation of *cenA* 5'-specific hybridisation probes (Fig.1, A and B), a portion of *C. fimi* DNA was subcloned from M13mp8::Bam 0.8kb (Wong, 1986) to the vector pUC18. The M13mp8::Bam 0.8kb dsDNA was digested with *Bam*H1, fractionated in a 1.0% low melting point (l.m.p.) agarose gel and the 0.75 kb *Bam*H1 fragment was recovered (Maniatis *et al.*, 1982). This was then digested with *Sal*1 and ligated to *Bam*H1-*Sal*1 cut pUC18 DNA. Ampicillin resistant (Ap^r) Lac^- JM101 transformants were screened by mini-lysate plasmid isolation and subsequent restriction enzyme analysis. A plasmid carrying the 0.64 kb *Bam*H1-*Sal*1 fragment of *cenA* in the multiple cloning site (MCS) of pUC18 was isolated and designated pNG101. Restriction endonucleases are abbreviated as follows: Bm, *Bam*H1; H3, *Hind*III; R1, *Eco*R1; Sa, *Sal*1; Sm, *Sma*1; Ss, *Sst*1.

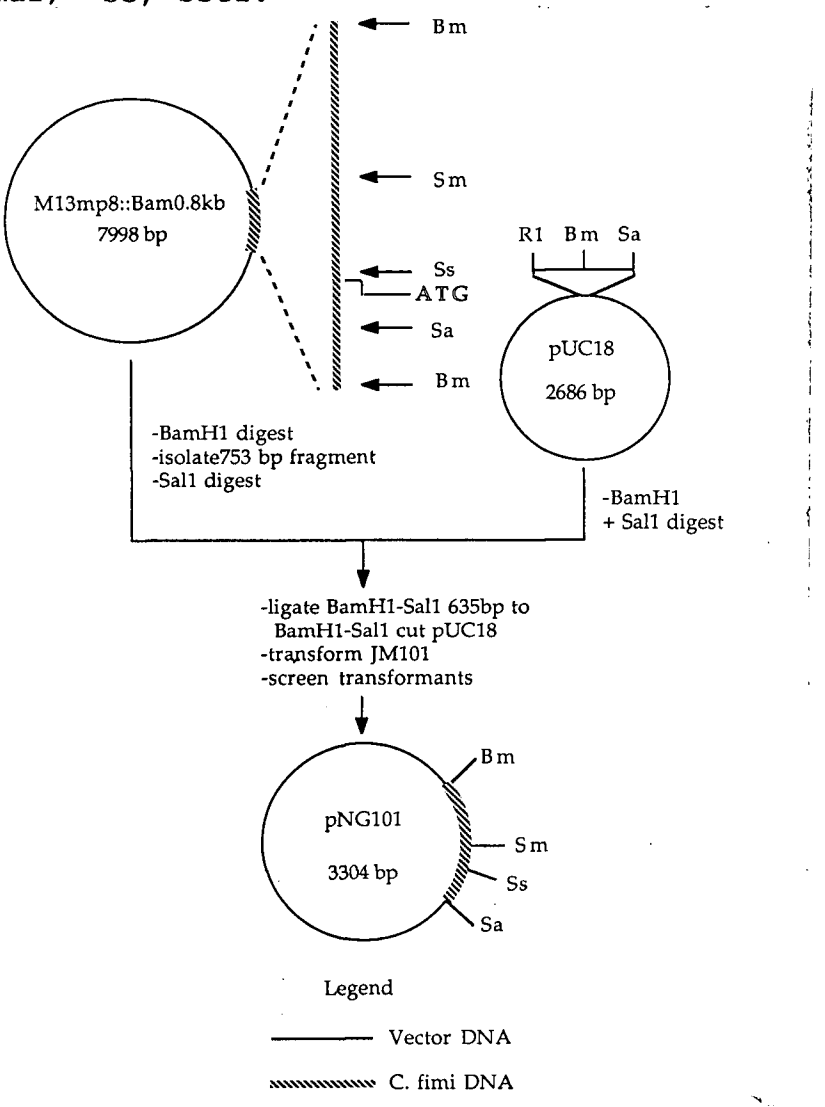


FIGURE 3. Isolation of a *cenA* specific northern blot probe. Plasmid pNG101 was digested with *Sal*1, treated with CIAP and 5'-end labeled with [γ-^{32}P]-ATP and T4 PNK. Linear, end-labeled plasmid DNA was digested with *Sst*1 and fractionated by electrophoresis in a 5% polyacrylamide gel. After autoradiography, a gel slice with the 123 bp *Sst*1-*Sal*1 probe was excised. The probe was then electroeluted into TAE buffer, precipitated with ethanol and recovered by centrifugation.

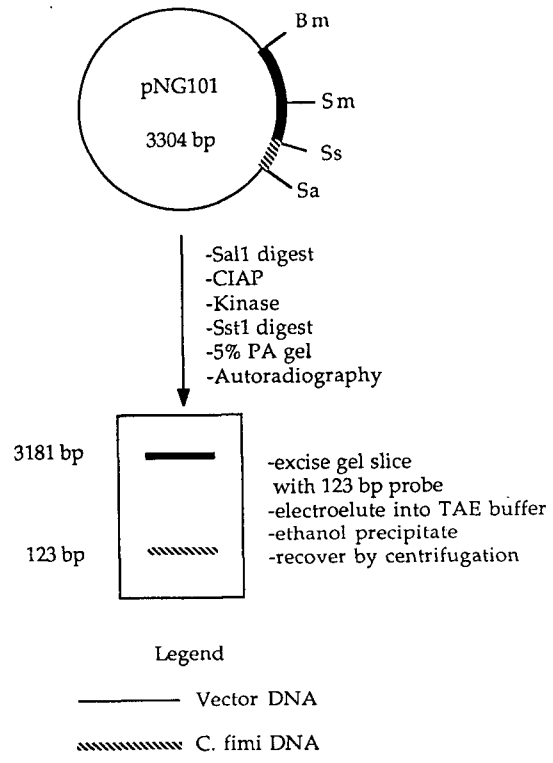
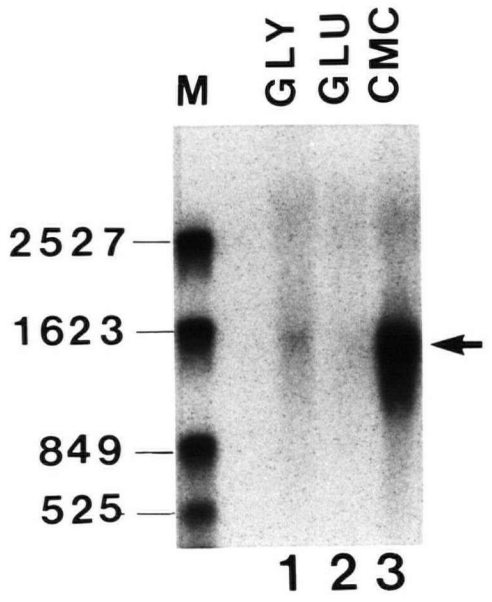


FIGURE 4. Northern blot analysis of *cenA*-specific transcripts. RNA was extracted from *C. fimi* cultures grown in basal medium supplemented with glycerol, or glucose, or CMC, fractionated in a formaldehyde gel containing 1.0% agarose, and blotted to a Biotrans membrane. The blot was then hybridised with the *cenA* intragenic *Sst*1-*Sal*1 probe (Fig. 1, A) labeled at the 5' Sal1 site (Fig. 3). Lanes: M, *Hae*III restriction fragments of single-stranded M13mp11, 5'-labeled with ^{32}P (sizes in nucleotides are indicated on the left); 1, RNA from glycerol-grown cells; 2, RNA from glucose-grown cells; 3, RNA from CMC-grown cells. Arrow indicates the major hybrid.



cells (Fig. 4, lane 2).

3.1.2. Mapping the *cenA* mRNA 5' ends.

Two complementary hybrid protection studies were used to confirm the direction of *cenA* transcription and to determine the 5' ends of *cenA* mRNA.

3.1.2.1. Mapping the *cenA* 5' ends with a 5'-labeled DNA probe.

Transcripts synthesized *in vivo* were analysed by high resolution S1 nuclease mapping. The 315 bp *Sma*1-*Sal*1 DNA restriction fragment (Fig. 1, B) isolated from the plasmid pNG101 (Fig. 5) was used as the hybridisation probe. The probe, labeled at the 5' *Sal*1 site, was denatured in solution and hybridised with total *C. fimi* RNA and S1 nuclease was used to degrade the unprotected portions of the DNA probe in any of the resulting RNA-DNA hybrids. The lengths of the protected probe species still bearing the ^{32}P labeled 5'-*Sal*1 terminus were then determined by gel electrophoresis under denaturing conditions. As seen in Fig. 6, an autoradiograph of the analytical gel revealed four distinct species of protected probe (lane 6). These species all mapped upstream of the *cenA* translation initiation codon. Three of these species were closely spaced (Fig.6; -1, +1, +2), the fourth migrated slightly slower (Fig.6; -11).

FIGURE 5. Isolation of a *cenA* 5' mRNA-specific S1 probe. The plasmid pNG101 was digested with *Sal*1, treated with CIAP and 5'-end labeled with [γ-^{32}P]-ATP and T4 PNK. Linear 5'-end labeled plasmid DNA was then digested with *Sma*1 and fractionated by electrophoresis on a 5% polyacrylamide gel. After autoradiography, a gel slice with the 315 bp *Sma*1-*Sal*1 probe was excised. The probe was then electroeluted into TAE buffer, precipitated with ethanol and recovered by centrifugation.

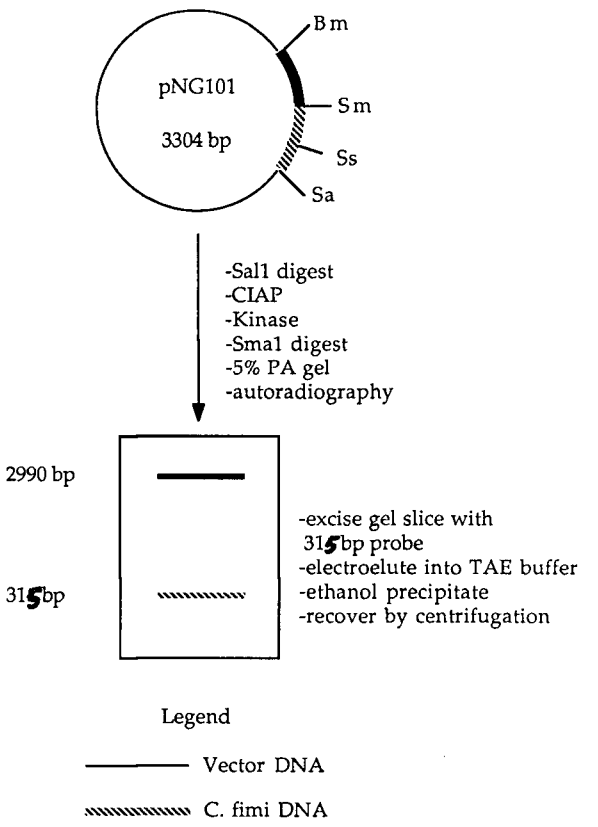
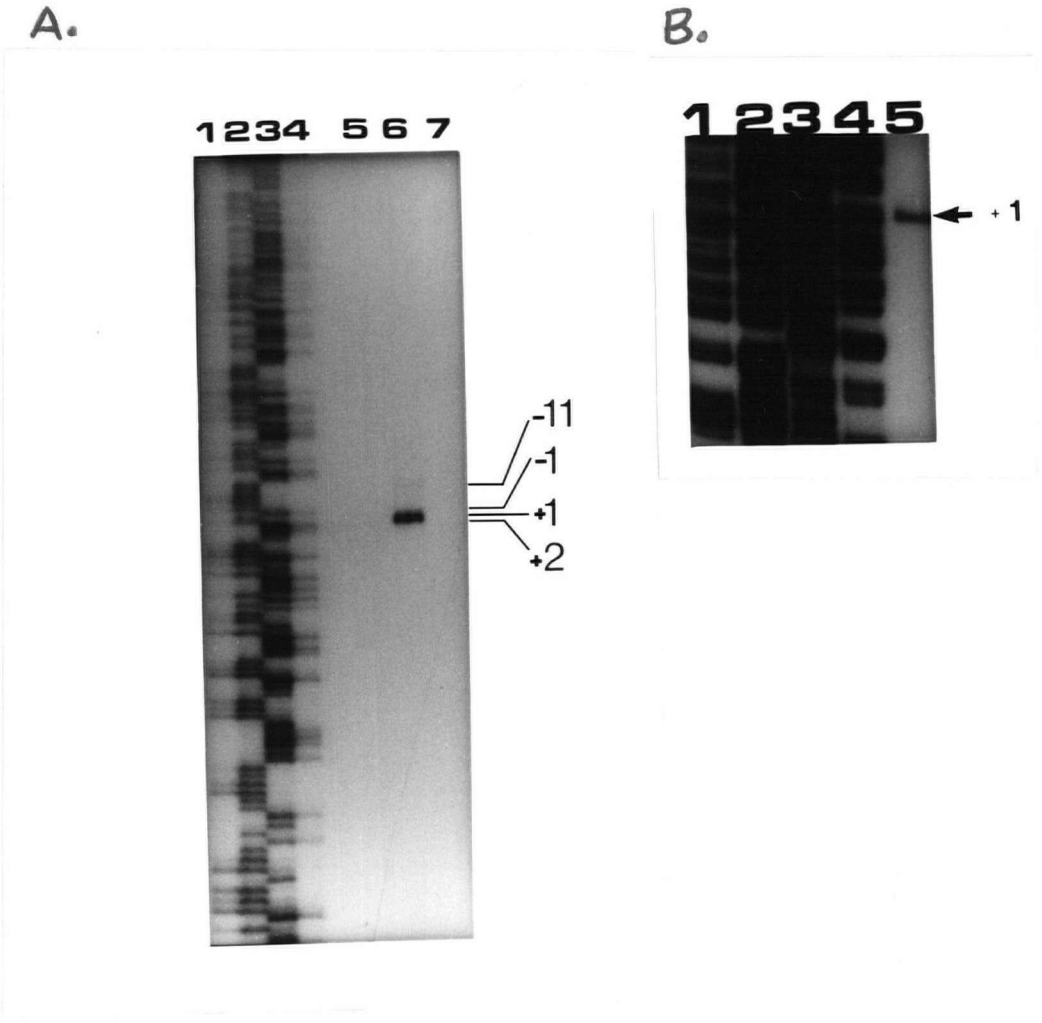


FIGURE 6. Mapping the 5' ends of *cenA* mRNA . After hybridisation with RNA and treatment with S1 nuclease the remaining *Sma*1-*Sal*1 probe (labeled at the 5' *Sal*1 site) was analysed in an 8% polyacrylamide-7M urea sequencing gel alongside probe sequenced by the base-specific chemical cleavage method (Maxam and Gilbert, 1980). Lanes 1 through 4 contained the chemical sequencing ladders of the probe: G>A, G+A, T+C, and C>T, respectively. Numbers on the right identify species of protected probe. A. Protection of the *Sma*1-*Sal*1 probe by: RNA from glucose-grown *C. fimi* (lane 5), RNA from CMC-grown *C. fimi* (lane 6), and yeast tRNA (lane 7). B. Protection of the *Sma*1-*Sal*1 probe by RNA from glycerol-grown cells (lane 5).



Of the four species of probe observed when RNA from CMC-grown *C. fimi* was used in the analysis (Fig.6A, lane 6), only the most prominent species (+1) was also observed, albeit as a relatively weaker signal, in mapping studies with RNA from glycerol-grown *C. fimi* (Fig. 6B, lane 5). The labeled probe was not protected in mapping studies with RNA from glucose grown *C. fimi* (Fig.6A, lane 5) or with yeast tRNA (Fig.6A, lane 7). These results confirmed those of the northern blot experiments (see section 3.1.1.) which had detected *cenA* mRNA in the RNA from CMC-gromwn and glycerol-grown *C. fimi*, but not from glucose-grown *C. fimi*. The most prominent protected species, +1, mapped to a G residue, 51 bases from the translation inititation codon of the unprocessed *cenA* gene product (Fig. 7; Wong *et al.*, 1986) and 164 bases from the ^{32}P-labeled 5' end of the probe.

3.1.2.2. Mapping the *cenA* mRNA 5' terminus with capped RNA.

The hybrid protection study with the *Sma*1-*Sal*1 DNA probe labeled at the 5' *Sal*1 site (section 3.1.2.1. above) could have identified the 5' ends of *cenA* mRNAs which were either intact or partially degraded at their 5' termini. Therefore, a second independent approach was taken to confirm that primary transcript initiation sites had been identified for *cen*A mRNA. Total RNA from CMC-grown *C. fimi* was labeled *in vitro* with the vaccinia virus guanylyltransferase enzyme which only recognizes those RNAs possessing 5' di- or tri-

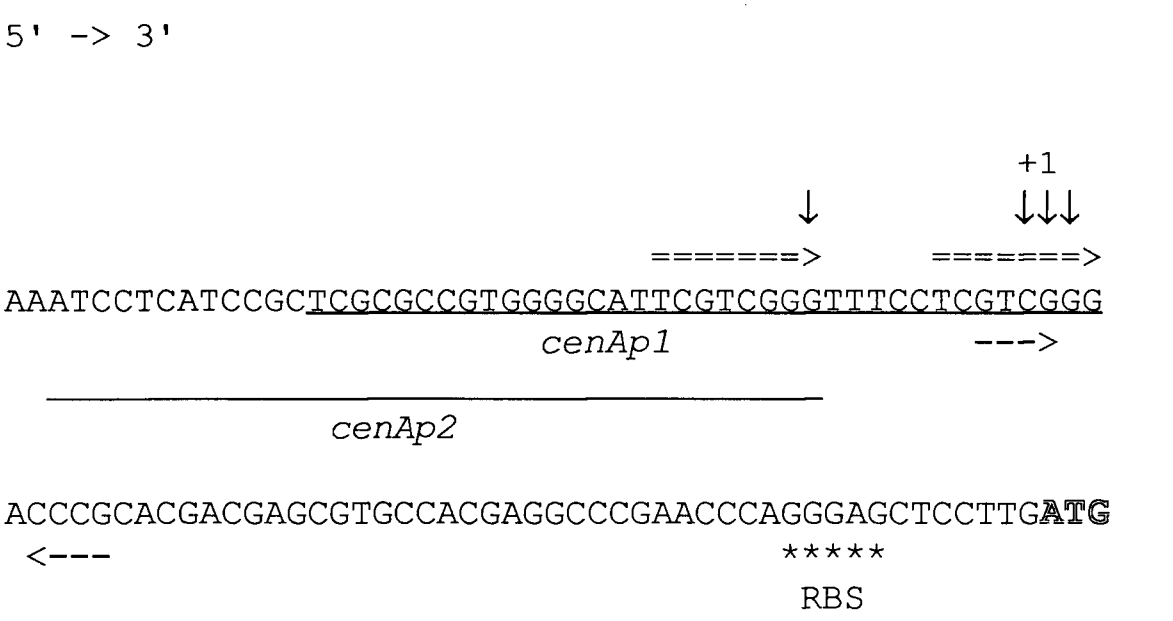
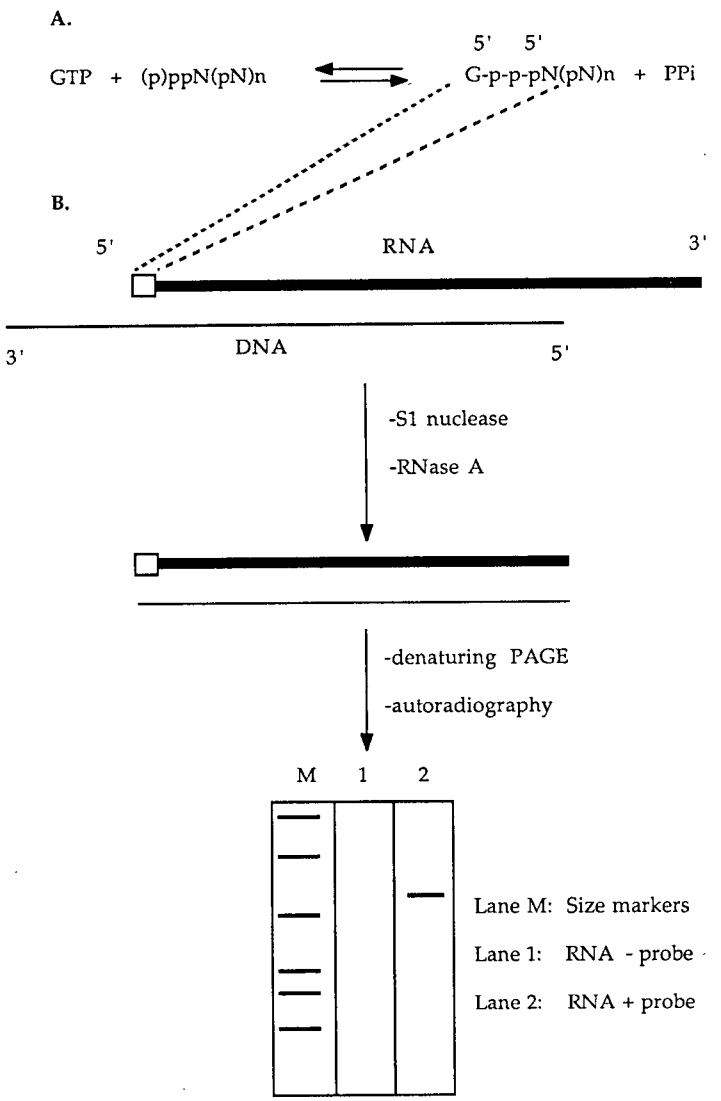


FIGURE 7. DNA sequence corresponding to the 5'- terminal region of *cen*A mRNA. Only the sense strand is shown. The vertical arrows denote the 3' nucleotides of the protected fragments of the 5' S1 probe. The ATG codon is outlined. A putative ribosome binding site (RBS) is underscored with asterisks. The putative *cenAp1* and *cenAp2* promoter regions are underlined. A direct repeat is overscored with thick arrows and an inverted repeat is underscored with thin arrows.

phosphates (i.e. the intact 5' ends of primary transcripts) as suitable receptors in a capping reaction (Fig.8). By using [α-^{32}P]-GTP as the donor in the capping reaction, primary transcripts which had been isolated *in vivo*, were thereby 5' labeled *in vitro*. The RNA prepared from CMC-grown *C. fimi* was used in the capping experiments since such preparations had been shown through northern blot and S1 studies to be enriched for *cenA* mRNA.

To map the *cenA* mRNA 5' end by exploiting the substrate specificity of the capping enzyme in the RNA labeling reaction, the *Sma*1-*Sal*1 probe (Fig.1, B) was tested for its ability to protect about 164 bases of the 5' end of a capped RNA species from nuclease digestion. This size was predicted from the distance between the +1 site identified in Fig. 6 and the *Sal*1 end of the probe. A protected fragment of about this size was indeed observed after nuclease digestion of the RNA-DNA hybrids (Fig. 9, lane 2) which was conspicuously absent in the control (Fig. 9, lane 1, RNA without probe). The slightly slower migration of the capped transcript may be explained by the presence of the 5' cap structure ($G^{5'}ppp^{5'}Np..$). Since in this experiment a species of partially protected capped RNA corresponding to the higher (-11) band observed in Fig. 6 was not detected, the S1 mapping data at best suggests the possibility of a -11 start site (see discussion section).

FIGURE 8. Hybrid protection with capped RNA. A. The reaction catalysed by the vaccinia virus guanylyltransferase enzyme between a [α-^{32}P]-GTP donor and a typical procaryotic RNA primary transcript. This results in a 'capped' RNA species, uniquely labeled at the 5' terminus (adapted from Moss, 1981). B. Schematic representation of a hybrid protection analysis using a suitable DNA probe to map the 5' terminus of a capped RNA. The capped RNA is allowed to hybridise to the DNA probe. After treatment with S1 nuclease and RNaseA the resulting hybrids are subjected to electrophoretic analysis in a denaturing polyacrylamide gel. Autoradiography reveals the size of the protected capped RNA relative to markers run in parallel. This determines the distance from the known position of the 5' end of the complementary DNA probe to the position of the 5' (capped) end of the RNA.



3.1.3. Mapping the 3' end of *cenA* mRNA.

To identify the 3' end of *cenA* mRNA, transcripts synthesised *in vivo* were analysed by S1 nuclease mapping with a 575 bp *Bgl*II-*Sma*1 DNA probe (Fig. 1, C). The probe, isolated from the plasmid pNG102 (Fig. 10) was labeled at the 3' *Bgl*II site with ^{32}P (Fig. 11). The probe was denatured in solution and hybridised with total *C. fimi* RNA and S1 nuclease was used to degrade the unprotected portions of the DNA probe in any of the resulting RNA-DNA hybrids. The lengths of the protected probe species still bearing the ^{32}P labeled 3'-*Bgl*II terminus were then determined by gel electrophoresis under denaturing conditions. As seen in Fig. 12, an autoradiograph of the analytical gel revealed three distinct species of protected probe (lane 6). These species mapped to positions 1438, 1449, and 1464 bases from the +1 site of *cenA* mRNA, and all three were downstream of the *cenA* translation termination codon (Fig. 13; Wong *et al.*, 1986). These species were not observed in mapping studies with RNA from glucose-grown *C. fimi* (Fig. 12, lane 5) or with yeast tRNA (not shown). The *cenA* mRNA 3' ends were found to map to a region of inverted repeats.

FIGURE 9. Mapping the *cen*A mRNA 5' terminus with capped RNA. The *C. fimi* RNA labeled with guanylyltransferase and [α-^{32}P]-GTP and the *cenA Sma*1-*Sal*1 probe (Fig.1, B) were hybridised in solution and treated with S1 nuclease and RNaseA. The hybrids were then analysed in a 5% polyacrylamide-7M urea gel. The numbers on the left indicate the size and migration of 5'-labeled M13mp11 *Hae*III fragments. The arrow on the right indicates the specific probe-protected RNA species (lane 2, RNA + probe). The results of the parallel negative control experiment without probe added to the labeled RNA are also shown (lane 1, RNA - probe).

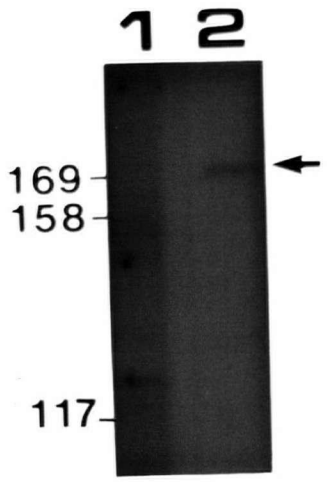


FIGURE 10. Subcloning the 3' terminal and flanking DNA of the *cenA* structural gene. The *cenA* 3'-specific hybridisation probe (Fig.1, C) was isolated from a portion of *C. fimi* DNA which had been subcloned from pEC2.1 (Wong *et al.*, 1986) to the MCS of pUC18. The pEC2.1 DNA was digested with *Sma*1, fractionated in a 1% low melting point agarose gel and the 939 bp *Sma*1 fragment was recovered (Maniatis *et al.*, 1982). This was ligated to pUC18 which had been digested with *Sma*1 and treated with CIAP. Ampicillin resistant (Ap^r), Lac^- JM101 transformants were screened by mini-lysate plasmid isolation and subsequent restriction enzyme analysis for a plasmid bearing the 939 bp *Sma*1 fragment in the MCS of pUC18. The plasmid isolated was designated pNG102.

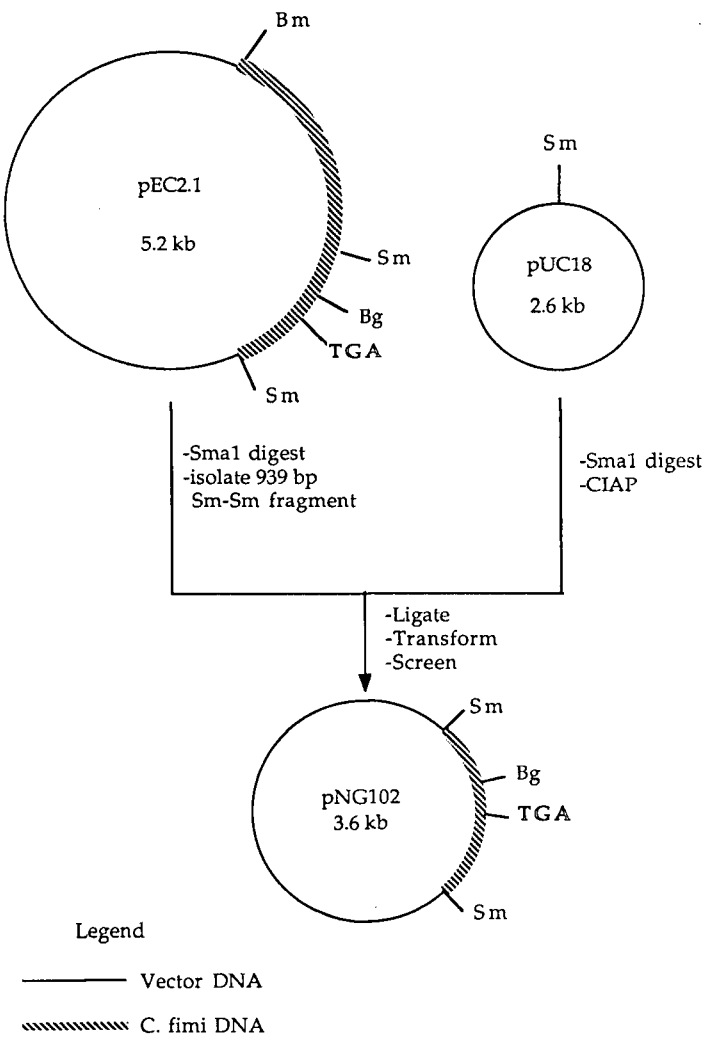


FIGURE 11. Isolation of a 3' *cen*A mRNA specific S1 probe. The plasmid pNG102 was digested with *Bgl*II and 3'-end labeled with [α-^{32}P]-dGTP and the Klenow fragment of *E. coli* DNA polymerase I. Linear 3' end-labeled plasmid DNA was then digested with *Sma*1 and fractionated by electrophoresis in a 5% polyacrylamide gel. After autoradiography, a gel slice with the 575 bp *Bgl*II-*Sma*1 probe was excised. The probe was then electroeluted into TAE buffer, precipitated with ethanol and recovered by centrifugation.

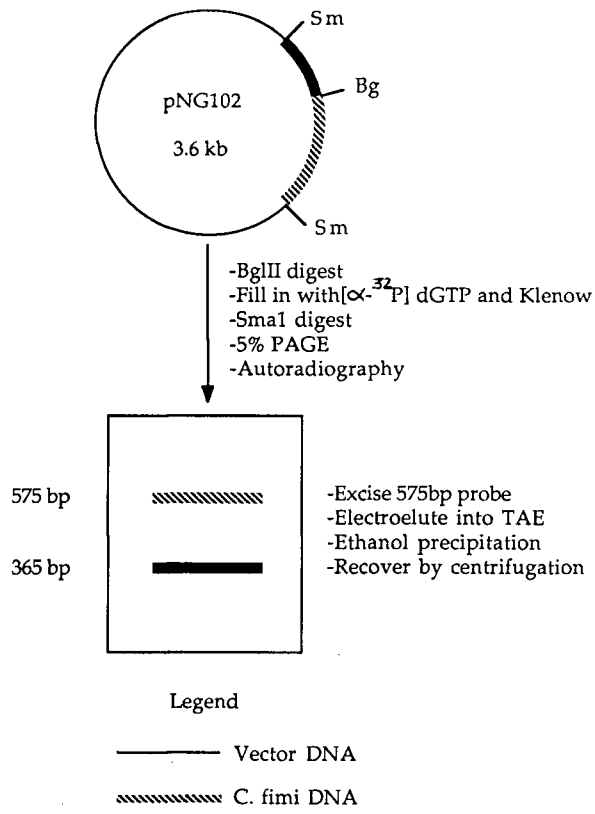
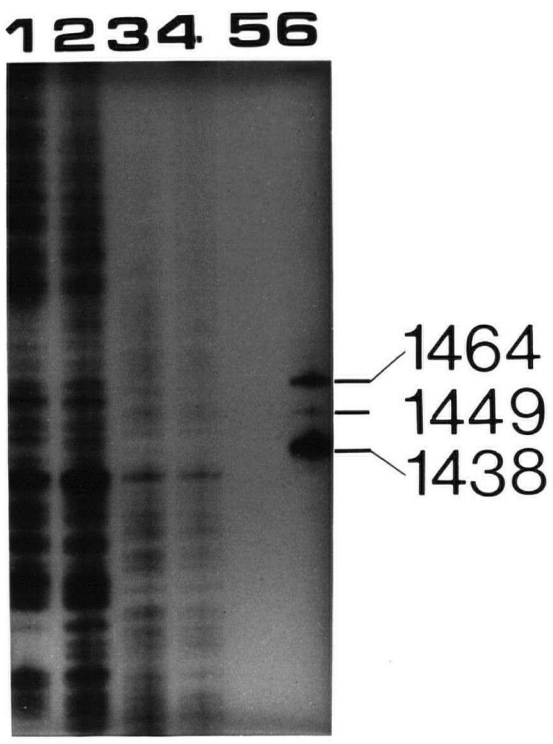


FIGURE 12. Mapping the 3' end of *cenA* mRNA. After hybridisation with *C. fimi* RNA and treatment with S1 nuclease, the *cenA*-specific DNA probe was analysed in a 5% polyacrylamide-8.3M urea sequencing gel alongside probe sequenced by the chemical cleavage method (Maxam and Gilbert, 1980). The *cenA Sma*1-*Bgl*II (site of the 3' end labeled with ^{32}P) probe (Fig. 1, C) was hybridised with RNA from glucose-grown cells (lane 5) and CMC-grown cells (lane 6). Lanes 1 through 4 contained chemical sequencing ladders G>A, G+A, T+C, and C>T, respectively. Numbers on the right denote distance (in bases) from the +1 site (Fig. 6).



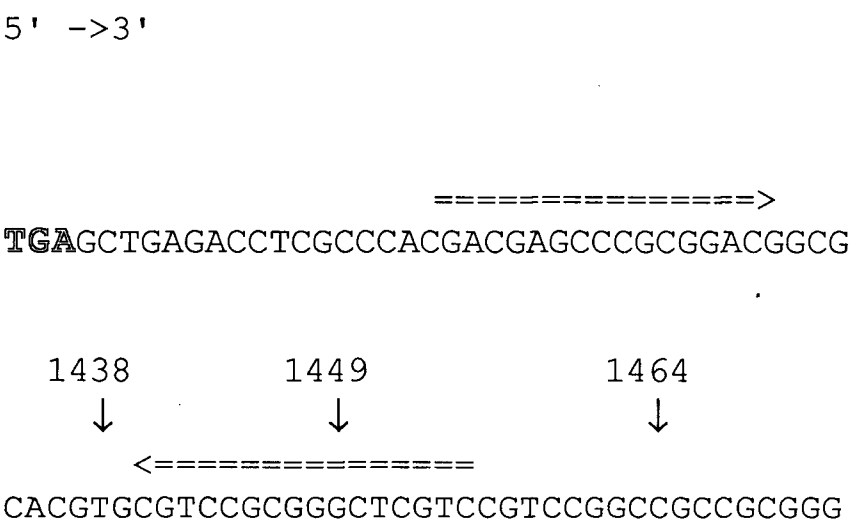


FIGURE 13. DNA sequence corresponding to the 3'-terminal region of *cen*A mRNA. Only the sense strand is shown. The vertical arrows denote the 5' nucleotides of the partially protected fragments of the 3' S1 probe. The TGA stop codon is outlined. The nucleotide sequences which may form stem loop structures are shown as opposing overlined horizontal arrows.

3.1.4. Steady state levels of *cen*A mRNA.

The effects of the carbon sources provided to *C. fimi* on *cen*A expression were further characterized by hybridisation analysis to determine the steady state levels of *cen*A mRNA. In these experiments, RNA was isolated from glycerol-, glucose-, or CMC-grown mid-log phase cultures of *C. fimi* and analysed *in vitro* by quantitative filter hybridisations. Synthetic oligodeoxyribonucleotides, 30 bases in length and complementary in sequence to the first 10 codons of the *cen*A structural gene, were 5' labeled with ^{32}P and used as hybridisation probes. This approach was followed since it allowed for the rapid detection and quantitation of multiple RNA samples. A series of 2-fold dilutions of plasmid pNG101 DNA bearing the target sequence (Fig. 2) served as the internal standards for quantitiative determinations. The results of these determinations are summarised in Table IV. As expected, the steady state levels of *cen*A mRNA in exponentially growing cells were found to be affected by the carbon source. Whereas the RNA from CMC-grown cells showed about 5-fold more *cen*A mRNA than RNA from glycerol-grown cells by this analysis, the northern and S1 data suggested even a greater difference. This discrepency probably results from transcripts which were not full length yet could still bind the probe and thereby contribute to the signal in the slot blots but which would be unable to contribute to the signal in the northern and S1 analysis.

The RNA from glucose-grown cells had less *cen*A mRNA than either of the other two preparations. These results, together with the data obtained through northern blot and S1 fine-mapping studies showed *cen*A expression in *C. fimi* to be affected as a function of the carbon source provided during growth. It also appeared that *cen*A expression was induced by growth on soluble cellulosic substrate (CMC) and repressed by growth on glucose. It should be noted that the analysis performed detected only steady state *cen*A mRNA levels, and the rates of *cen*A transcript initiation were not measured. Therefore the possibility exists that the regulation of *cen*A expression could also be post-transcriptional, for example at the level of RNA stability and turnover, which could also account for the observed changes in the relative levels of mRNA.

Carbon source	*cen*A mRNA†
Glycerol	48±6
Glucose	13±4
CMC	257±8

† amol *cen*A mRNA per μg total *C. fimi* RNA

TABLE IV. Steady state *C. fimi cen*A mRNA levels. Total RNA was prepared from exponentially growing cultures of *C. fimi* provided with basal medium supplemented with glycerol (0.2%), glucose (0.2%) or CMC (1.0%) as carbon source. Samples of RNA were immobilized on nylon membranes and then hybridised with 5' ^{32}P-labeled oligodeoxyribonucleotide probes 30 bases in length and complementary to the first 10 codons of the *cen*A structural gene (specific activity 5 X 10^7 cpm μg^{-1}). Following hybridisations at 52°C (10 ng probe per ml; 1 ml per 100 cm^2 membrane), the filters were washed at 52°C with three changes of 2X SSC / 0.1 % SDS and one change of 1X SSC / 0.1 % SDS. The filters were then washed at 60°C with 0.1X SSC / 0.1 % SDS and exposed to X- ray film at -70°C with intensifying screens. Autoradiographic signals were quantitated with a Helena QuickScan densitometer equipped with an integrator. Titrated amounts of the plasmids pNG101, pNG202 and pNG301 specific for the *cen*A, *cex* and *cen*B probes, respectively, served as the internal standards. Alkali- treated (0.2N NaOH for 2 h at 37°C) RNA samples served as the negative controls. All samples applied to the membranes contained 4 μg of total nucleic acid. The results are expressed as the mean (± SD) from two determinations in atom moles (amol) of *cen*A mRNA per μg of total *C. fimi* RNA.

3.2. Characterization of the *cex* transcripts of *C. fimi*.

3.2.1. Regulation by carbon source and approximate length of *cex* mRNA.

A qualitative northern blot analysis was used to determine the approximate length of *cex* mRNA and the influence of the carbon source for *C. fimi* growth on *cex* expression. For these experiments, *C. fimi* RNA was prepared from cultures grown in basal medium supplemented with either 0.2%(w/v) glycerol, 0.2%(w/v) glucose or 1.0%(w/v) CMC. The intragenic 122 bp *Sty*1-*Sal*1 fragment of *cex* (Fig. 14, A) was used as the hybridisation probe. The probe, isolated from the plasmid pUC12A25 (Fig 15; O'Neill, 1986) was 5' labeled at the *Sal*1 site (Fig 16).

A strong signal was observed in hybridisations between *C. fimi* RNA extracted from CMC-grown cells and the labeled *Sty*1-*Sal*1 probe. The species of RNA detected by the probe was approximately 1500 bases in length (Fig 17, lane 3). No signals were detected in the hybridisations between the probe and RNA from glycerol- (lane 1) or glucose- (lane 2) grown cells.

3.2.2. Mapping the 5' ends of *cex* mRNA.

Two complementary hybrid protection studies were used to confirm the direction of *cenA* transcription and to determine the 5' ends of *cenA* mRNA.

FIGURE 14. Partial restriction map of the *cex* gene. Representation of the cloned 2.6 kb *Bam*H1-*Sal*1 segment of *C. fimi* DNA containing the *cex* gene. The structural gene is shown as a boxed region and is translated from left to right (5' -> 3'). A, *Sty*1-*Sal1* northern blot probe; B, *Pst*1-*Ban*1 5' S1 probe; B', *Sau*3A1 fragment used in the 5' mapping experiments with radiolabeled RNA; C, *Sau*3A1-*Sal*1 3' S1 probe. The restriction endonucleases are abbreviated as follows: Bm, *Bam*H1; Bn, *Ban*1; Ps, *Pst*1; S3, *Sau*3A1; Sa, *Sal*1; St, *Sty*1.

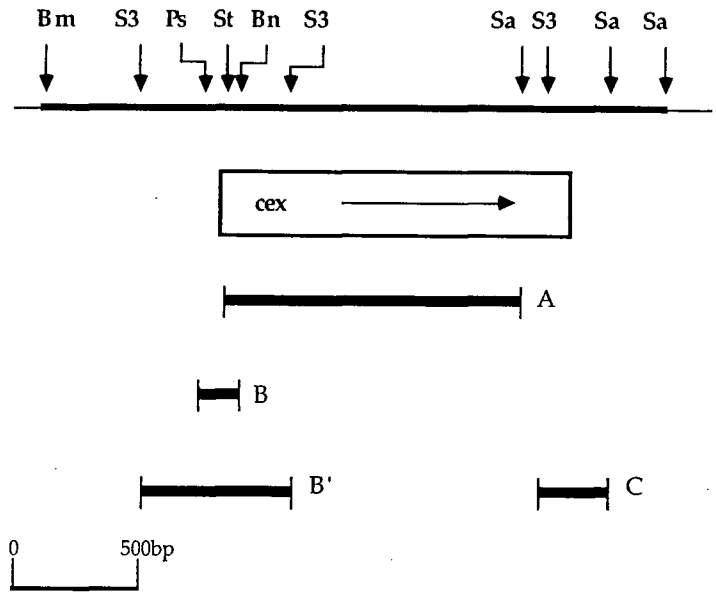


FIGURE 15. A partial restriction map of the plasmid pUC12A25. The plasmid carries 2.6 kb of *C. fimi* DNA with the *cex* gene and its flanking regions within the MCS of pUC12. The restriction endonucleases are abbreviated as follows: Bm, *Bam*H1; H3, *Hind*III; N, *Nar*1; Ps, *Pst*1; Sa, *Sal*1; S3, *Sau*3A1, St, *Sty*1.

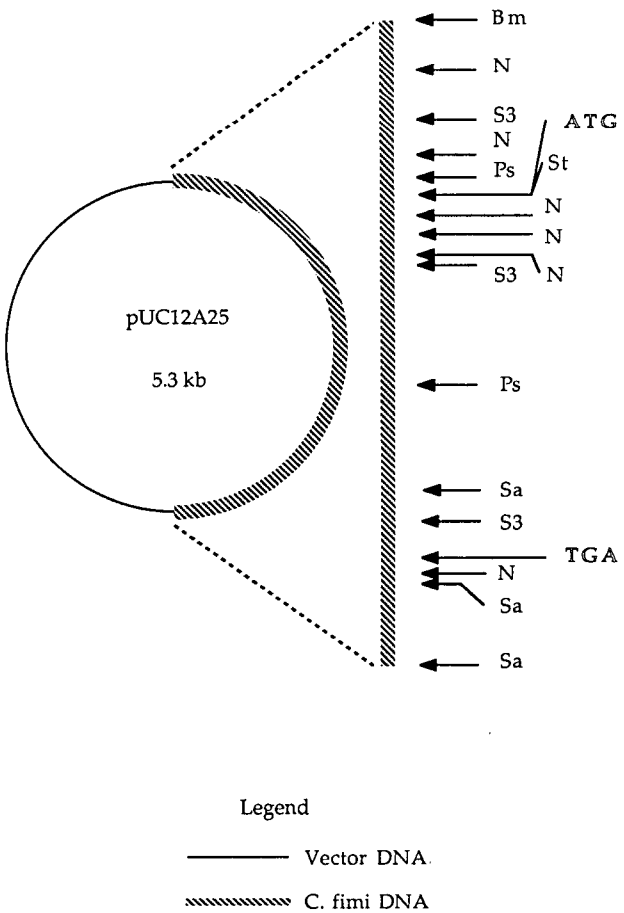


FIGURE 16. Isolation of a *cex*-specific northern blot probe. The plasmid pUC12A25 was digested with *Sal*1, treated with CIAP and 5'-labeled with [γ-^{32}P]-ATP and T4 PNK. Linear 5' end-labeled plasmid DNA was then digested with *Sty*1 and fractionated in a 5% polyacrylamide gel. After autoradiography, a gel fragment with the 1222 bp *Sty*1-*Sal*1 probe was excised. The probe was electroeluted into TAE buffer, precipitated with ethanol and recovered by centrifugation.

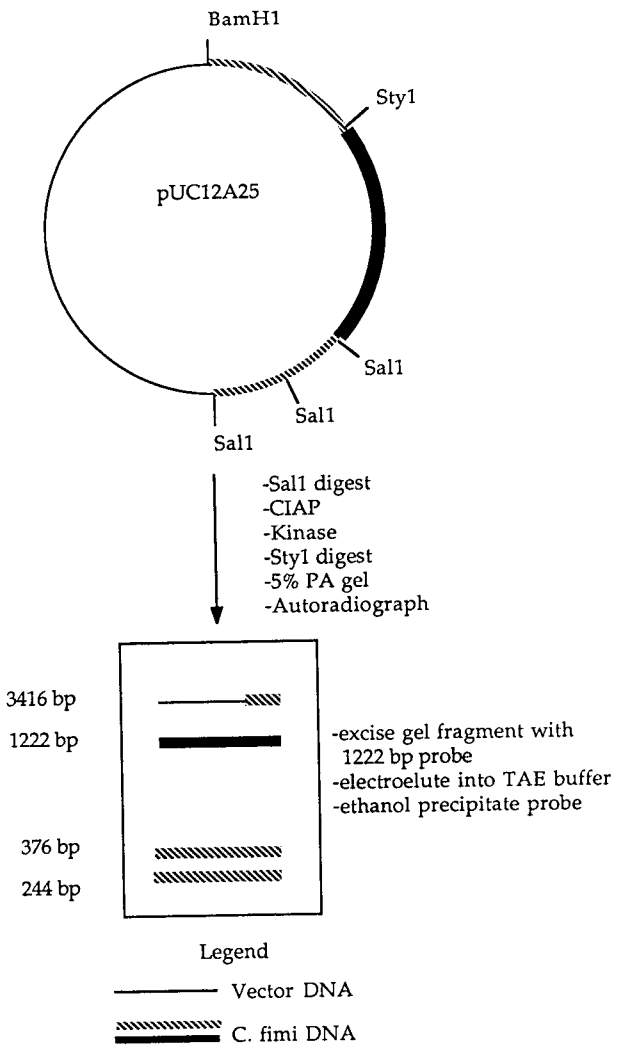
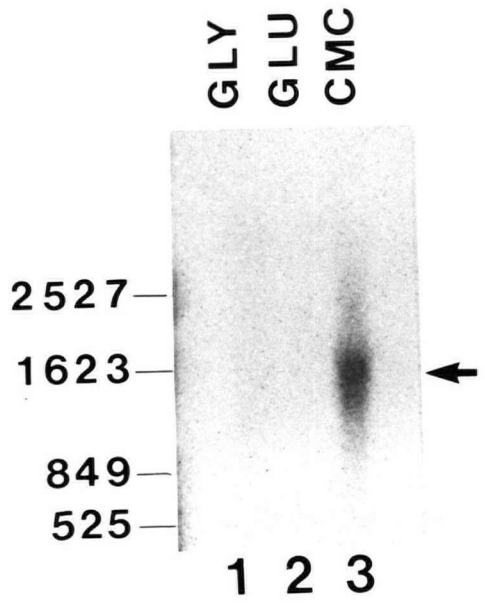


FIGURE 17. Northern blot analysis of *cex*-specific transcripts. RNA was extracted from *C. fimi* cultures grown in basal medium supplemented with glycerol, or glucose, or CMC, fractionated in a formaldehyde gel containing 1.0% agarose, and blotted to a Biotrans membrane. The blot was then hybridised with the *cex* intragenic *Sty*1-*Sal*1 probe (Fig. 15, A) labeled at the 5' *Sal*1 site. Lanes: 1, RNA from glycerol- grown cells; 2, RNA from glucose- grown cells; 3, RNA from CMC- grown cells. The numbers on the left designate the size and migration of *Hae*III restriction fragments of single- stranded M13mp11 DNA.



3.2.2.1. Mapping the *cex* mRNA 5' ends with a 5' labeled DNA probe.

Transcripts synthesized *in vivo* were analysed by high resolution S1 nuclease mapping. The 136 bp *Pst*1-*Ban*1 DNA restriction fragment was used as the hybridisation probe (Fig 14, B). To isolate this probe, a 136 bp *Pst*1-*Nar*1 fragment was first isolated from pUC12A25 and subcloned to pUC18 generating pNG200. A 216 bp *Sma*1-*Pvu*II fragment containing two extraneous *Ban*1 sites was then deleted, generating pNG201 (Fig. 18), from which the *cex* 5' S1 probe was excised as a *Pst*1-*Ban*1 fragment (Fig. 19).

The labeled probe was denatured in solution and allowed to hybridise with *C. fimi* total RNA. The resulting RNA-DNA hybrids were treated with S1 nuclease to degrade the unhybridised segments of the DNA probe. The length of the RNA protected 5' portion of the DNA probe was determined by polyacrylamide gel electrophoresis under denaturing conditions. As seen in Fig. 20, autoradiography revealed four adjacent species of *cex* specific probe which had been protected at their 5' termini from nuclease S1 digestion by RNA from CMC-grown *C. fimi* (lane 6). The corresponding species of protected probe DNA were not observed in studies with yeast tRNA (Fig. 20, lane 7) or RNA from either glucose-grown (Fig. 20, lane 5) or glycerol-grown *C. fimi* (not shown).

FIGURE 18. Subcloning 5' flanking and terminal *cex* DNA. To facilitate the isolation of a *cex* mRNA 5' specific hybridisation probe, a segment of *C. fimi* DNA was subcloned from pUC12A25 to pUC18. The pUC12A25 and pUC18 DNAs were digested with *Nar*1 and *Pst*1 and fractionated in a 2.0% agarose gel. The 136 bp *Nar*1-*Pst*1 pUC12A25 fragment and the 2498 bp pUC18 fragment were recovered, ligated together (Maniatis, *et al.*, 1982) and used to transform JM101. Ampicillin resistant (Ap^r), Lac^- colonies were then screened by mini- lysate plasmid isolation and subsequent restriction enzyme analysis. A plasmid carrying the 136 bp *Nar*1-*Pst*1 fragment of *cex* within the unique *Nar*1 and *Pst*1 sites of pUC18 was isolated and designated pNG200. While the hybrid *Nar*1 site of pNG200 appeared to be a poor substrate for endonuclease cleavage (unpublished observations), the *Nar*1 isoschizomer *Ban*1 appeared to be better at cleaving the hybrid site (unpublished observations). To facilitate the recovery of the probe, two extraneous *Ban*1 sites (*Ban*1 recognizing the hexanucleotide sequence $5'G\uparrow GPyPuCC3'$) were eliminated following digestion of pNG200 with *Pvu*II and *Sma*1 and dilute ligation. This created pNG201 which, with only two *Ban*1 sites, was the source of *cex* 5'-specific S1 DNA probe for the mapping experiments with unlabeled RNA.

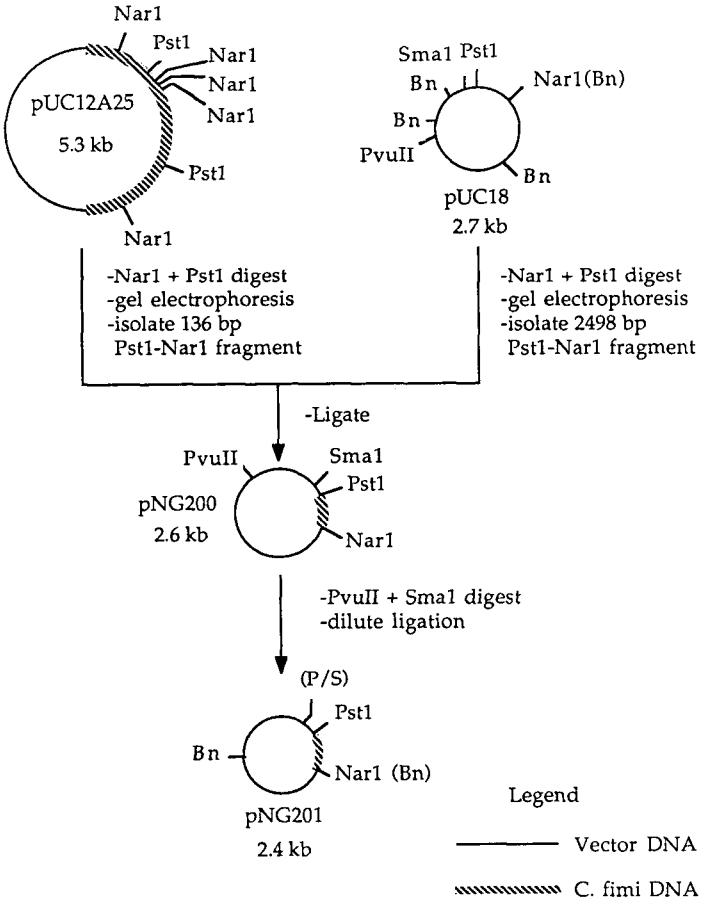

Nar1
Pst1
Nar1
Nar1
Nar1
pUC12A25
5.3 kb
Pst1
Nar1
Sma1 Pst1
Bn
Nar1(Bn)
Bn
PvuII
Bn
pUC18
2.7 kb
-Nar1 + Pst1 digest
-gel electrophoresis
-isolate 136 bp
Pst1-Nar1 fragment
-Nar1 + Pst1 digest
-gel electrophoresis
-isolate 2498 bp
Pst1-Nar1 fragment
-Ligate
PvuII
Sma1
Pst1
pNG200
2.6 kb
Nar1
-PvuII + Sma1 digest
-dilute ligation
(P/S)
Pst1
Bn
Nar1 (Bn)
pNG201
2.4 kb
Legend
Vector DNA
C. fimi DNA

FIGURE 19. Isolation of a *cex* 5' mRNA-specific S1 probe. The plasmid pNG201 was digested with *Ban*1, treated with CIAP and 5' labeled with [γ-^{32}P]-ATP and T4 PNK. Linear 5' labeled plasmid DNA was then digested with *Pst*1 and fractionated in a 5% polyacrylamide gel. After autoradiography, a gel fragment with the 136 bp *Pst*1-*Ban*1 probe was excised. The probe was electroeluted into TAE buffer, precipitated with ethanol and recovered by centrifugation.

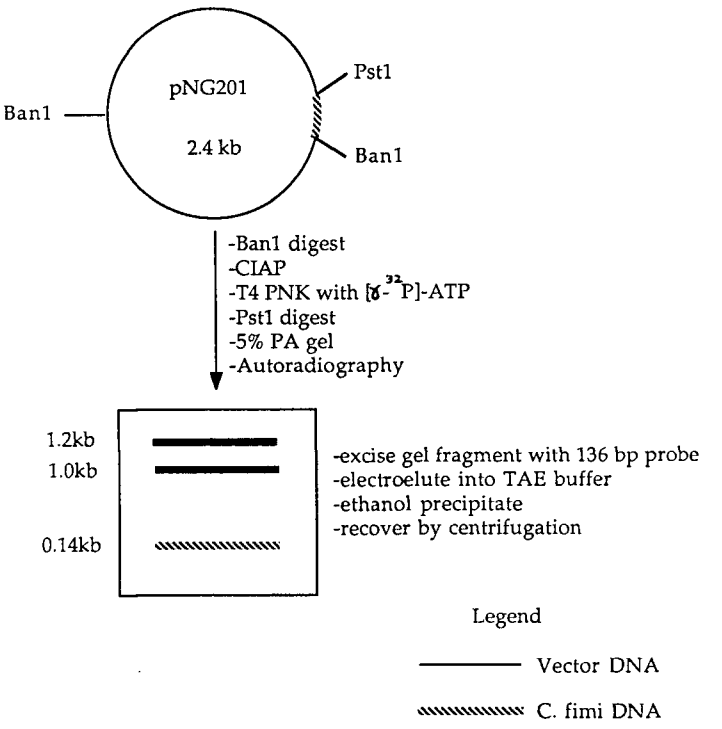
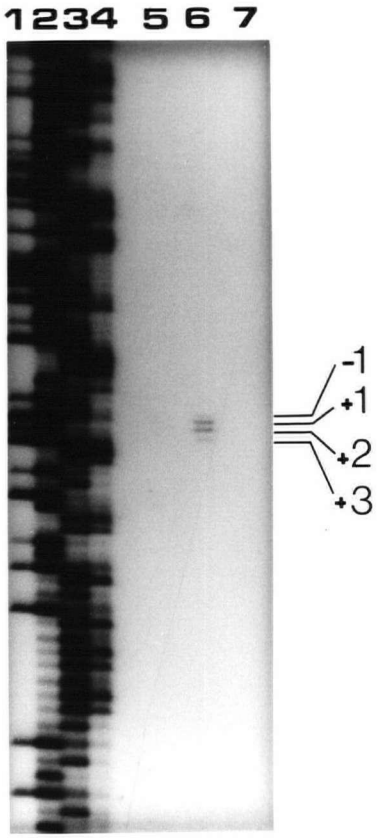


FIGURE 20. Mapping the 5' end of *cex* mRNA. After hybridisation with RNA and treatment with S1 nuclease the remaining *Pst*1-*Ban*1 probe (labeled at the 5' *Ban*1 site) was analysed in an 8% polyacrylamide-7M urea sequencing gel alongside probe sequenced by the base-specific chemical cleavage method (Maxam and Gilbert, 1980). Protection of the *Pst*1-*Ban*1 probe by: RNA from glucose-grown *C. fimi* (lane 5), RNA from CMC-grown *C. fimi* (lane 6), and yeast tRNA (lane 7). Lanes 1 through 4 contained the chemical sequencing ladders of the probe: G>A, G+A, T+C, and C>T, respectively. Numbers on the right identify species of the protected probe.



The most prominent protected species, +1, mapped to a C residue, 28 bases upstream of the translation initiation codon for the unprocessed *cex* gene product (Fig. 21; O'Neill *et al.*, 1986) and 72 bases from the ^{32}P-labeled 5' end of the probe.

3.2.2.2. Mapping the *cex* mRNA 5' terminus with RNA labeled *in vitro* with guanylyltransferase and [α-^{32}P]-GTP.

The hybrid protection study with the *Pst*1-*Ban*1 DNA probe labeled at the 5' *Ban*1 site (3.2.2.1. above) could have identified the 5' ends of *cex* mRNAs which were either intact or partially degraded at their 5' termini. Therefore, in a second hybrid protection analysis, analogous to that presented in section 3.1.2.2. above, total RNA from CMC-grown *C. fimi* was labeled *in vitro* with vaccinia virus guanylyltransferase enzyme and [α-^{32}P]-GTP and used in a hybrid protection analysis with an unlabeled, *cex* 5'-specific DNA probe.

To map the *cex* mRNA 5' terminus in this fashion, the 136 bp *Pst*1-*Ban*1 probe was tested for its ability to protect about 70 bases of the 5' end of a capped RNA species from nuclease digestion. However, the analysis did not resolve a specific signal from the background of nuclease-treated unhybridised capped RNA in this size range (not shown). A *Sau*3A1 DNA restriction fragment spanning a larger 5' flanking

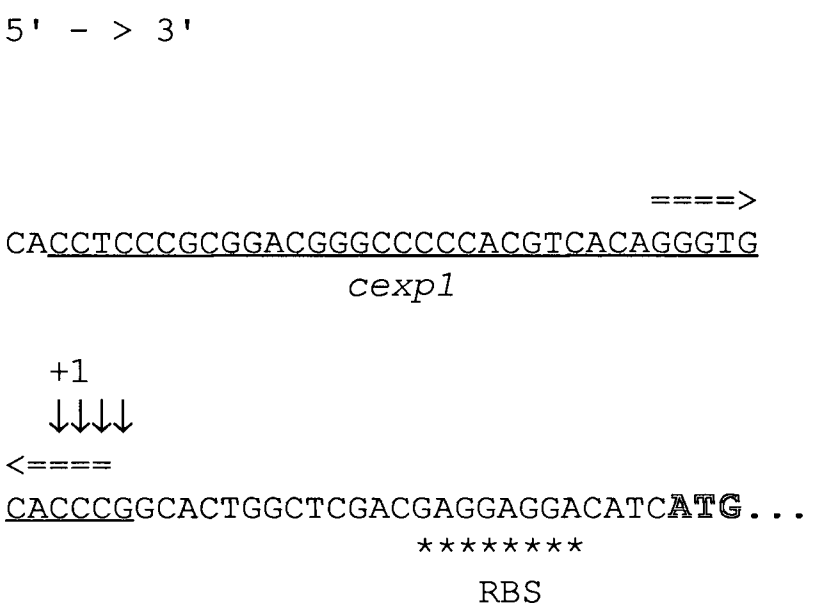


FIGURE 21. DNA sequence corresponding to the 5'-terminal region of *cex* mRNA. Only the sense strand is shown. The vertical arrows denote the 3' nucleotides of the protected fragments of the 5' S1 probe. The ATG codon is outlined. The putative ribosome binding site (RBS) is underlined with asterisks. The putative *cexp1* promoter region is underlined. An inverted repeat is overscored with horizontal arrows.

and terminal portion of the *cex* structural gene (Fig. 14, B') was therefore tested for its ability to protect about 246 bases of a capped RNA species from nuclease digestion. This *Sau*3A1 fragment was isolated from the plasmid pUC13Bam31 (Fig. 22), kindly provided by G. O'Neill. A species of protected RNA about 246 bases long was resolved by electrophoretic analysis following the nuclease-treatment of the hybrids that formed between the capped RNA and the *Sau*3A1 probe (Fig. 23, lane 2) which was not visible in the control experiment (Fig. 23, lane 1, RNA without probe).

3.2.3. Mapping the 3' ends of *cex* mRNA.

To identify the 3' ends of *cex* mRNA, transcripts synthesized *in vivo* were analysed by high resolution S1 nuclease mapping with a 254 bp *Sau*3A1-*Sal*1 DNA probe (Fig. 14, C). The probe, isolated from the plasmid pNG202 (Fig. 24) was labeled at the 3' *Sau*3A1 site with ^{32}P (Fig. 25). The probe was denatured in solution and hybridised with total *C. fimi* RNA and S1 nuclease was used to degrade the portions of the probe which were not protected by RNA. The lengths of the protected probe species still bearing the 3' ^{32}P-labeled-*Sau*3A1 terminus were then determined by gel electrophoretic analysis under denaturing conditions. As seen in Fig. 26, an autoradiograph of the analytical gel revealed a single species of protected probe (lane 6). This species mapped to a position 1564 bases from the +1 site of *cex* mRNA,

FIGURE 22. Representation of the plasmid pUC13Bam31. The 556 bp *Sau*3A1-*Sau*3A1 fragment of pUC12A25 was sub-cloned into the unique *Bam*H1 site of pUC13 to generate pUC13Bam31 (G. O'Neill, personal communication). The inserted fragment was liberated as a 596 bp *Eco*R1-*Hind*III fragment for the hybrid protection analysis with capped RNA.

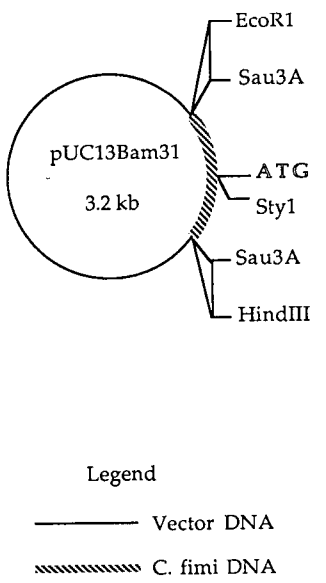


FIGURE 23. Mapping *cex* mRNA 5' terminus with capped RNA. *C. fimi* RNA labeled with guanylyltransferase and [α-^{32}P]GTP, and the *cex Sau*3A1 probe (Fig.14, B') were hybridised in solution and treated with S1 nuclease and RNaseA. The hybrids were then analysed in a 5% polyacrylamide-7M urea gel. The numbers on the left indicate the size and migration of 5' radiolabeled M13mp11 ssDNA *Hae*III fragments. The arrow on the right indicates the specific probe-protected RNA species (lane 2, RNA + probe). The results of the parallel negative control experiment without probe added to the labeled RNA are also shown (lane 1, RNA - probe).

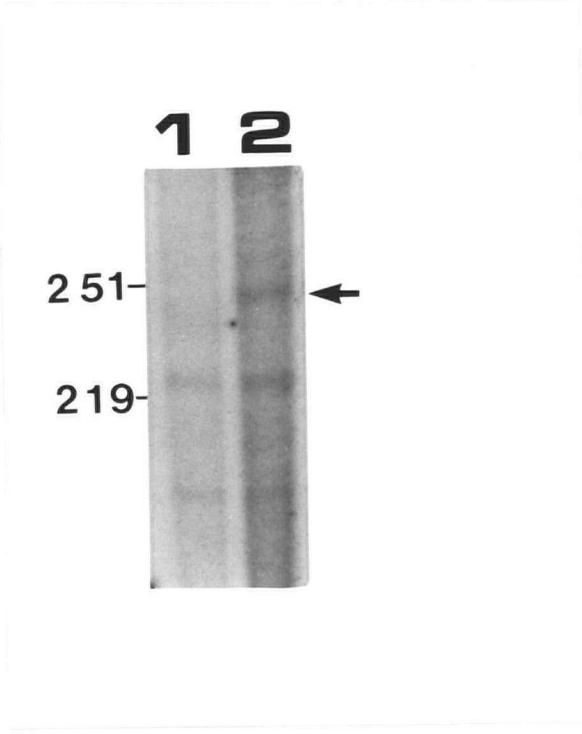


FIGURE 24. Subcloning the 3' terminal and flanking DNA of the *cex* structural gene. The *cex* 3'-specific hybridisation probe (Fig. 14, C) was isolated from a segment of *C. fimi* DNA which had been subcloned from pUC12A25 to pUC18. The pUC12A25 DNA was digested with *Sal*1, fractionated by electrophoresis through a 1.5% agarose gel and the 375 bp *Sal*1 fragment recovered (Maniatis *et al.*, 1982). This was then digested with *Sau*3A1 and ligated with pUC18 DNA which had been digested with *Bam*H1 and *Sal*1. Ampicillin resistant (Ap^r), Lac^- JM101 transformants were screened by mini-lysate plasmid isolation and subsequent restriction enzyme analysis for a plasmid bearing the 0.25 kb *Sau*3A1-*Sal*1 fragment of *cex* in the MCS of pUC18. By virtue of a C residue 3' flanking the *Sau*3a1 site of the *cex* fragment, the *Bam*H1 site of the vector was restored by ligation. The plasmid isolated was designated pNG202.

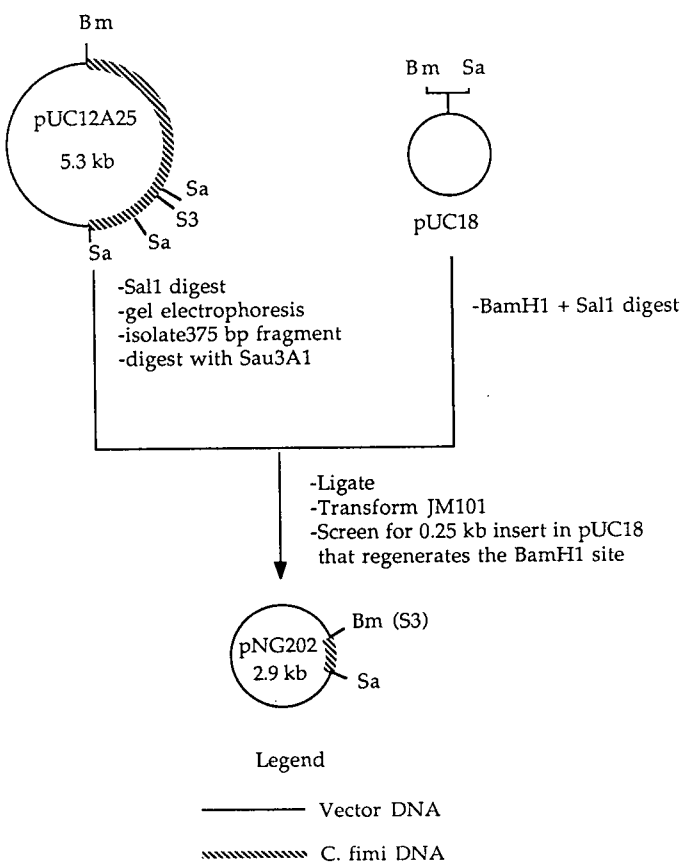


FIGURE 25. Isolation of a *cex* mRNA 3'-specific S1 probe. The plasmid pNG202 was digested with *Bam*H1 and 3'-end labeled with [α-^{32}P]-dGTP and the Klenow fragment of *E. coli* DNA polymerase I. Linear 3'-end labeled plasmid DNA was then digested with *Sal*1 and fractionated by electrophoresis in a 5% polyacrylamide gel. After autoradiography, a gel slice with the 254 bp *Sau*3A1-*Bam*H1 probe was excised. The probe was then electroeluted into TAE buffer, precipitated with ethanol and recovered by centrifugation.

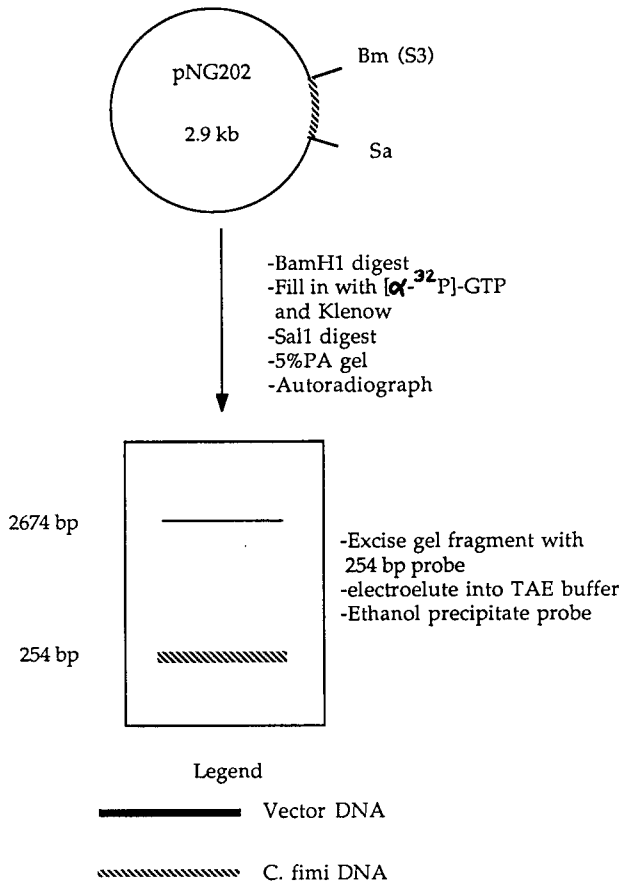
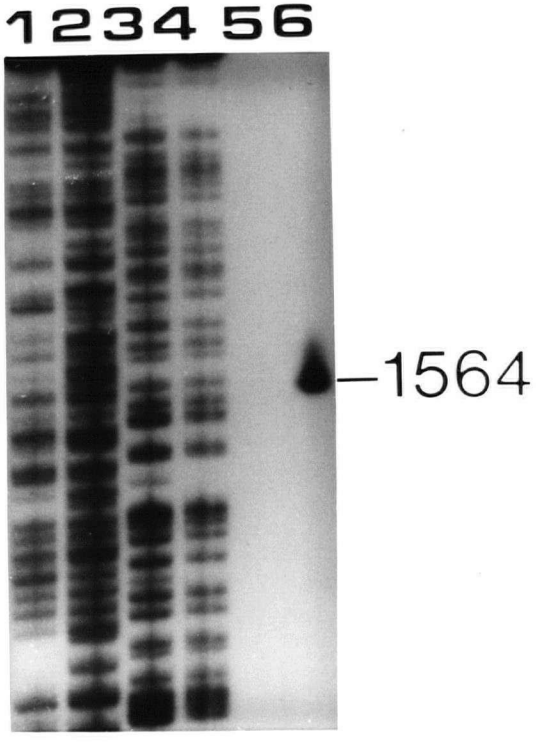


FIGURE 26. Mapping the 3' end of *cex* mRNA. After hybridisation with *C. fimi* RNA and treatment with S1 nuclease, the *cex*-specific DNA probe was analysed in a 5% polyacrylamide-8.3M urea sequencing gel alongside probe sequenced by the chemical cleavage method (Maxam and Gilbert, 1980). The *cex* *Bam*H1-*Sau*3A1 (site of the 3'-end labeled with ^{32}P) probe (Fig. 14, C) was hybridised with RNA from glucose-grown cells (lane 5) and CMC-grown cells (lane 6). Lanes 1 through 4 contained chemical sequencing ladders G>A, G+A, T+C, and C>T, respectively. Numbers on the right denote distance (in bases) from the +1 site (Fig. 20).



downstream of the *cex* translation termination codon (Fig. 27; O'Neill *et al.*, 1986). This species was not observed in mapping studies with RNA from glucose-grown *C. fimi* (Fig. 26, lane 5) or with yeast tRNA (not shown). The *cex* mRNA termination site was found to map to a region of inverted repeats (Fig. 27).

3.2.4. Steady state levels of *cex* mRNA.

The effects on *cex* expression by the carbon sources provided to *C. fimi* were further characterized by hybridisation analysis to determine the steady state levels of *cex* mRNA. In these experiments, RNA was isolated from glycerol-, glucose-, or CMC-grown cultures of *C. fimi* and analysed *in vitro* by quantitative filter hybridisations. Synthetic oligodeoxyribonucleotides, 30 bases in length and complementary in sequence to the first 10 codons of the *cex* structural gene, were 5' labeled with ^{32}P and used as hybridisation probes. A series of 2-fold dilutions of pNG201 plasmid DNA (Fig. 19) served as the standards for quantitative determinations. The results of these determinations are summarized in Table V. As expected, the steady state levels of *cex* mRNA in exponentially growing cells were affected by the carbon source, with the RNA from CMC-grown cells having about 2-fold more *cex* mRNA than RNA from glycerol- or glucose-grown cells. These results, together with the data obtained through northern blot and S1

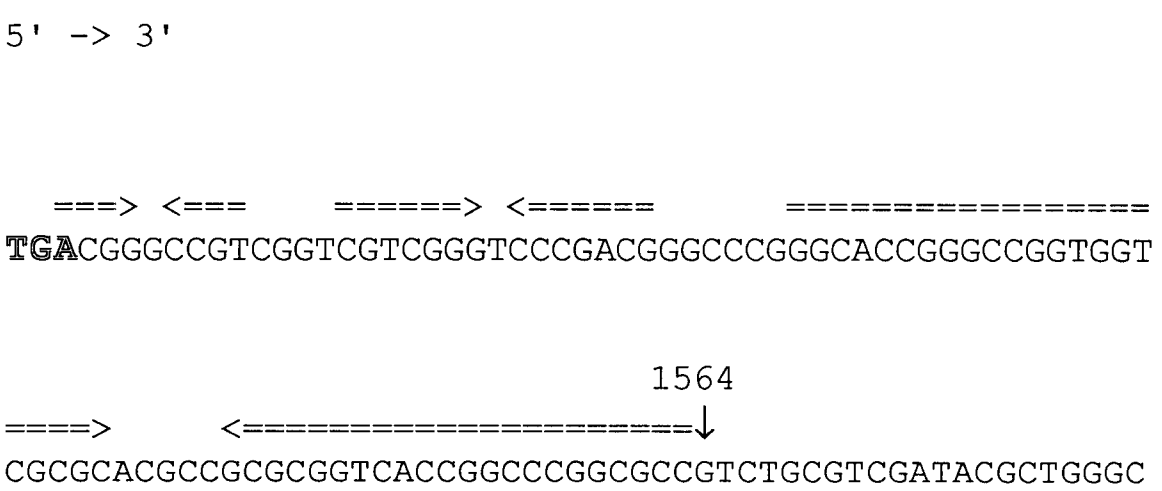


FIGURE 27. DNA sequence corresponding to the 3'- terminal region of *cex* mRNA. Only the sense strand is shown. The vertical arrow denotes the 5' nucleotide of the partially protected 3' S1 probe. The TGA stop codon is outlined. The sequences which may form stem loop structures are overlined with horizontal arrows. The numbering corresponds to the number of bases from the +1 site (Fig. 21).

Carbon source	*cex* mRNA[†]
Glycerol	18±3
Glucose	10±2
CMC	77±9

[†] amol *cex* mRNA per μg total *C. fimi* RNA

TABLE V. Steady state levels of *C. fimi cex* mRNA. Total RNA was prepared from cultures of *C. fimi* grown exponentially in basal medium supplemented with glycerol (0.2%), glucose (0.2%) or CMC (1.0%) as carbon source. Samples of RNA were immobilized on nylon membranes and filters were hybridised with 5' ^{32}P-labeled oligodeoxyribonucleotide probes 30 bases in length and complementary to the first 10 codons of the *cen*A structural gene (specific activity 5 X 10^7 cpm μg^{-1}). Following hybridisations at 52°C (10 ng probe per ml; 1 ml per 100 cm^2 membrane), the filters were washed at 52°C with three changes of 2X SSC / 0.1 % SDS and one change of 1X SSC / 0.1 % SDS. The filters were then washed at 60°C with 0.1X SSC / 0.1 % SDS and exposed to X- ray film at -70°C with intensifying screens. Autoradiographic signals were quantitated with a Helena QuickScan densitometer equipped with an integrator. Titrated amounts of the plasmids pNG101, pNG202 and pNG301 specific for the *cen*A, *cex* and *cen*B probes, respectively, served as the internal standards. Alkali- treated (0.2N NaOH for 2 h at 37°C) RNA samples served as the negative controls. All samples applied to the membranes contained 4 μg of total nucleic acid. The results are expressed as the mean (± SD) of two determinations in atom moles (amol) of *cex* mRNA per μg of total *C. fimi* RNA.

fine-mapping studies suggest that *cex* expression could be regulated by *C. fimi* at the transcriptional level by the carbon source provided during growth and that *cex* expression is induced by growth on soluble cellulosic substrate and is minimal with either glucose or glycerol provided as the carbon source.. However, since the rates of transcript initiation were not measured under the three growth conditions employed, the possibility exists that *cex* expression could also be regulated at the post-transcriptional level.

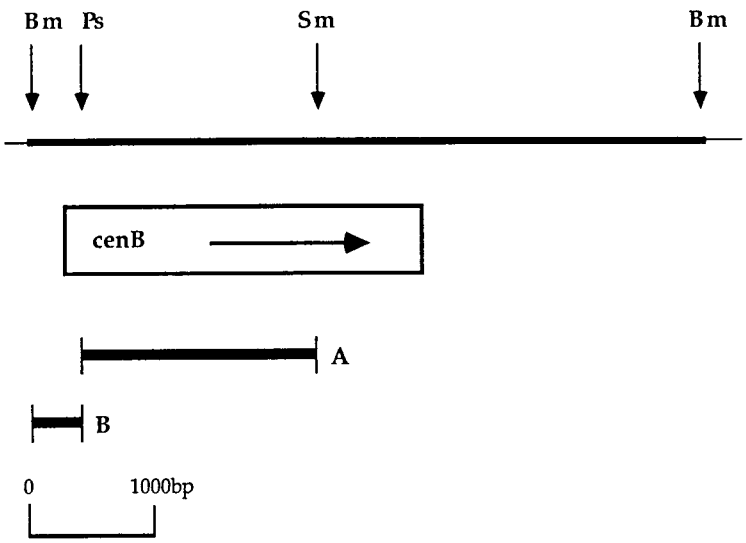
3.3. Characterization of the *cenB* transcripts of *C. fimi*.

3.3.1. Regulation by carbon source and approximate length of *cenB* mRNA.

A qualitative northern blot analysis was used to determine the approximate length of *cenB* mRNA and the influence on *cenB* transcription of the carbon source provided for *C. fimi* growth. For these experiments, *C. fimi* RNA was prepared from cultures grown in basal medium supplemented with either 0.2%(w/v) glycerol, 0.2%(w/v) glucose or 1.0%(w/v) CMC. The intragenic *Pst*1-*Sma*1 fragment of *cenB* (Fig. 28, A) was used as the hybridisation probe. The probe, carried on the plasmid pUC19C3PS (Fig. 29) was labeled by nick-translation.

An intense signal was observed in hybridisations between *C. fimi* RNA extracted from CMC-grown cells and the labeled

FIGURE 28. Partial restriction map of the *cen*B gene. Representation of the cloned 5.6-kilobase *Bam*H1-*Bam*H1 segment of *C. fimi* DNA containing the *cenB* gene. The structural gene is shown as a boxed region with the 3' end approximated from the northern blot data (this section). Translation is from left to right (5' -> 3'). A, *Pst*1-*Sma*1 Northern blot probe; B, *Bam*H1-*Pst*1 5' S1 probe. The restriction endonucleases are abbreviated as follows: Bm, *Bam*H1; Ps, *Pst*1; Sm, *Sma*1.



probe. The species of RNA detected by the probe was approximately 3200 bases in length (Fig. 30, lane 3). Less intense signals also corresponding to RNA species of about 3200 bases in length were detected in hybridisations between the probe and RNA from glycerol- and glucose-grown cells (Fig. 30, lanes 1 and 2, respectively).

3.3.2. Mapping the 5' ends of *cenB* mRNA.

A hybrid protection study was used to confirm the direction of *cenB* transcription and to identify the 5' ends of *cenB* mRNA. Transcripts synthesized *in vivo* were analysed by high resolution S1 nuclease mapping. The 400 bp *Bam*H1-*Pst*1 DNA restriction fragment (Fig. 28, A) isolated from the plasmid pNG301 (Fig. 31) was used as the hybridisation probe. The probe, labeled at the 5' *Pst*1 site, was denatured in solution and hybridised with total *C. fimi* RNA and S1 nuclease was used to degrade the unprotected portions of the DNA probe in any of the resulting RNA-DNA hybrids. The lengths of the protected probe species, still bearing the ^{32}P labeled 5'-*Pst*1 terminus, were then determined by gel electrophoresis under denaturing conditions. As seen in an autoradiograph of the analytical gel (Figure 32), the analysis revealed four distinct species of protected probe. These species all mapped upstream of the *cenB* translation initiation codon. Three of these species were closely spaced (Fig. 34, +1, +2, +3) and the fourth, migrated further down the gel

FIGURE 29. Representation of the plasmid pUC19C3PS. The 2.0 kb *Pst*1-*Sma*1 fragment of *cenB* is carried on the vector pUC19. This plasmid was nick- translated with [α-^{32}P]- dCTP and [α-^{32}P]-dGTP (specific activity 5 X 10^7 cpm μg^{-1}) and used as an intragenic probe for *cenB* mRNA in northern blot experiments. The restriction endonucleases are abbreviated as follows: R1, *Eco*R1; H3, *Hind*III; Ps, *Pst*1; Sm, *Sma*1.

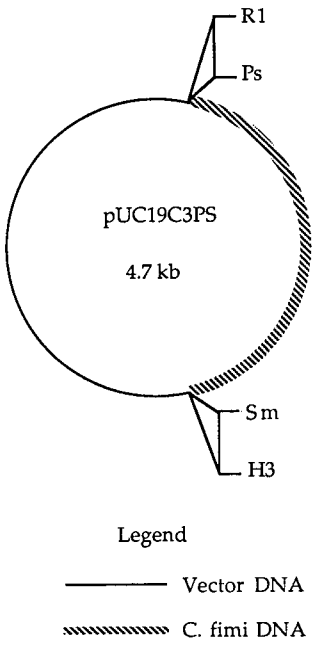


FIGURE 30. Northern blot analysis of *cenB*-specific transcripts. RNA was extracted from *C. fimi* cultures grown in basal medium supplemented with glycerol, or glucose, or CMC, fractionated in a formaldehyde gel containing 1.0% agarose, and transferred to a Biotrans membrane. The blot was hybridised with the nick-translated plasmid pUC19C3PS carrying the *cen*B intragenic *Pst*1-*Sma*1 fragment (Fig. 28 and 29). Lanes: M, labeled *Hind*III restriction fragments of lambda DNA (sizes in basepairs are indicated on the left); 1, RNA from glycerol-grown cells; 2, RNA from glucose-grown cells; 3, RNA from CMC-grown cells. Arrow indicates the major hybrid. the gel (Fig. 32, +52).

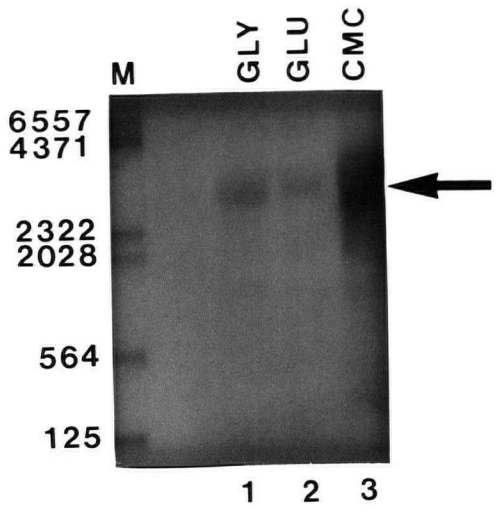


FIGURE 31. Isolation of a *cen*B 5' mRNA-specific S1 probe. The plasmid pUC19C3(ori2) carries a 5.2 kb *C. fimi* DNA insert with the *cen*B structural gene in the MCS of pUC19. The *cen*B gene is translated in the R1 -> H3 orientation. The ATG start codon is outlined. To isolate a plasmid with a unique *Pst*1 site, pUC19C3(ori2) was digested to completion with *Pst*1 and religated under dilute conditions. The Lac^- Ap^r JM101 transformants were screened by mini-lysate plasmid isolation and subsequent restriction enzyme analysis. A plasmid carrying only the 0.4 kb *Bam*H1- *Pst*1 fragment of pUC19C3 (ori2) in the MCS of pUC19 was isolated and designated pNG301. To isolate a *cen*B mRNA 5'-specific S1 probe, pNG301 was digested with *Pst*1, treated with CIAP and 5'-end labeled with [γ-^{32}P]-ATP and T4 PNK. Linear, end-labeled plasmid DNA was digested with *Bam*H1 and fractionated in a 5% polyacrylamide gel. After autoradiography, a gel slice with the 0.4 kb *Pst*1-*Bam*H1 probe was excised. The probe was then electroeluted into TAE buffer, precipitated with ethanol and recovered by centrifugation.

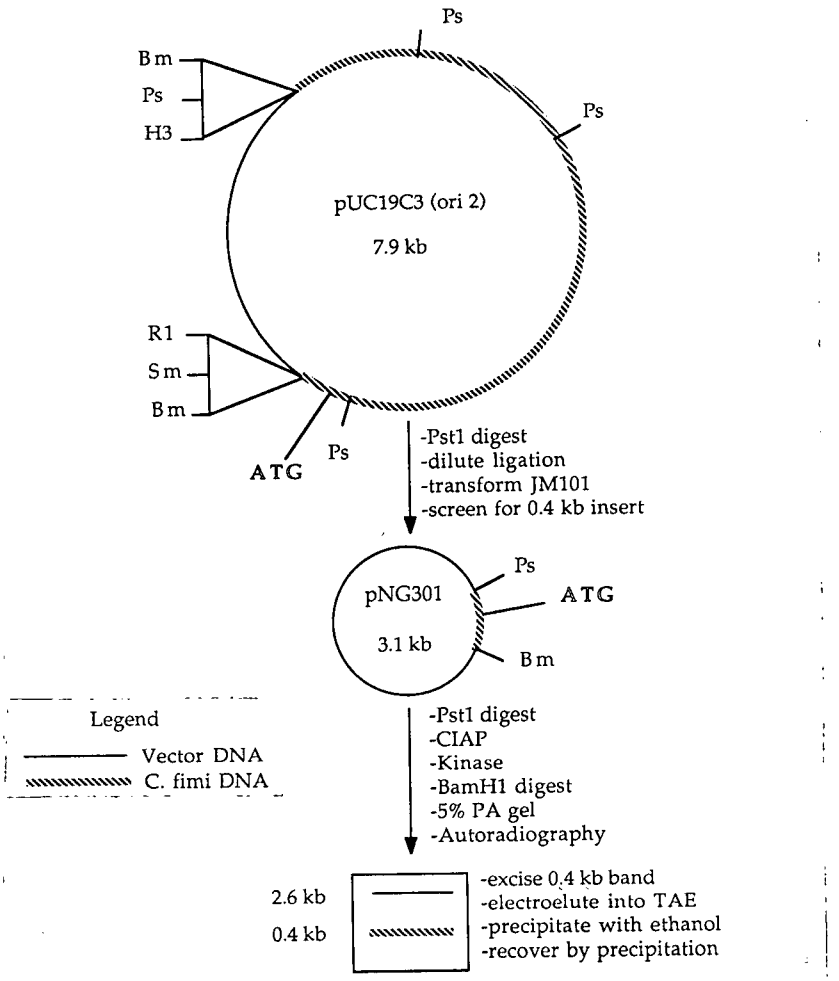
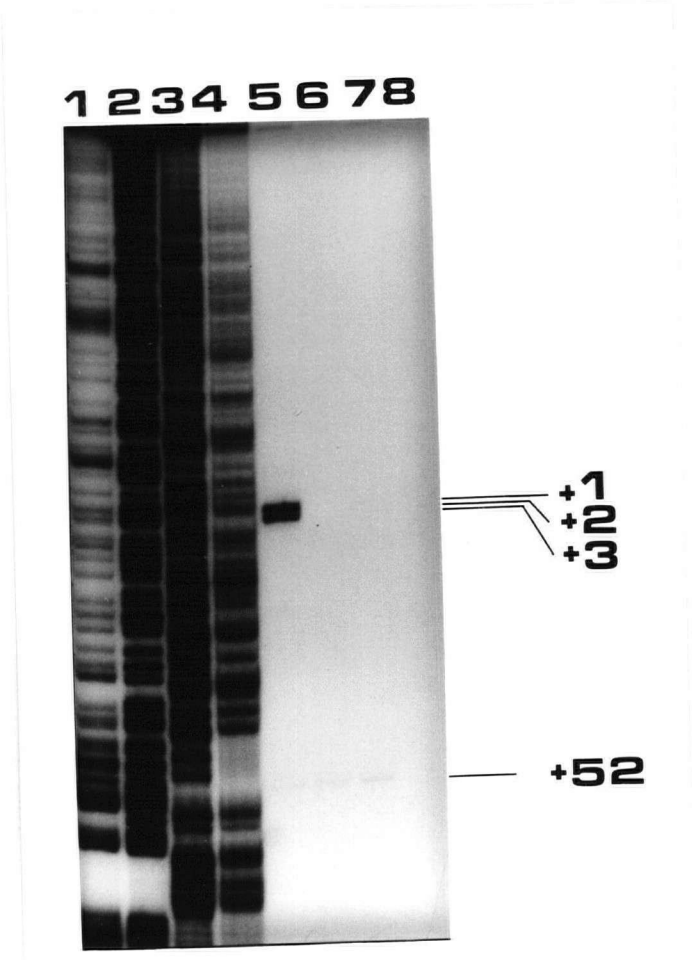


FIGURE 32. Mapping the 5' end of *cenB* mRNA. After hybridisation with RNA and treatment with S1 nuclease, the *cenB*-specific labeled DNA probe was analysed in an 8% polyacrylamide-7M urea sequencing gel alongside probe sequenced by the base-specific chemical cleavage method (Maxam and Gilbert, 1980). S1 protection of *cenB* *Bam*H1-*Pst*1 (site of 5' end labeled with ^{32}P) fragment by RNA from CMC-grown C.fimi (lane 5), RNA from glucose-grown *C. fimi* (lane 6), RNA from glycerol-grown *C. fimi* (lane 7), and yeast tRNA (lane 8). Lanes 1 through 4 contained chemical sequencing ladders G>A, G+A, T+C, and C>T, respectively. Numbers on the right identify species of the protected probe.



(Fig.32, +52). The most prominent species resolved in the analysis, +1, (Fig. 32, lane 5) corresponded to a G residue 201 bases from the ^{32}P-labeled 5'-end of the probe and 75 bases upstream of the initiation codon of the unprocessed *cenB* gene product (Fig. 33; Owalabi *et al.*, 1988).

The +1, +2 and +3 species were clearly visible in mapping experiments with RNA from CMC-grown *C. fimi* (Fig. 32, lane 5), while the fourth, weaker species was seen upon prolonged autoradiographic exposure of the dried gel to X-ray film (not shown). When RNA isolated from glucose-grown cultures was used in mapping experiments, only the +52 species was detected (Fig. 32, lane 6). In mapping studies with RNA isolated from glycerol-grown cells, the +52 species was detected as the major species (Fig. 32, lane 7), while the +1, +2, and +3 species were detected only after prolonged autoradiographic exposure to X-ray film (not shown). No hybrids were detected in control experiments with yeast tRNA (Fig. 32, lane 8).

3.3.3. Steady state levels of *cen*B mRNA.

The effects of the carbon sources provided to *C. fimi* on *cen*B expression were further characterized by hybridisation analysis to determine the steady state levels of *cen*B mRNA. In these experiments, RNA was isolated from glycerol-, glucose-, or CMC-grown cultures of *C. fimi* and analysed *in vitro* by quantitative filter hybridisations. Synthetic

oligodeoxyribonucleotides, 30 bases in length and complimentary in sequence to the first 10 codons of the *cen*B structural gene, were 5'-labeled with ^{32}P and used as hybridisation probes. A series of 2-fold dilutions of pNG301 plasmid DNA (Fig. 31) served as internal standards for quantitative determinations. The results of these determinations are summarized in Table VI. As expected, the steady state levels of *cen*B mRNA in exponentially growing cells were affected by the carbon source provided during growth, with the RNA from CMC-grown cells having about 10 to 20-fold more *cen*B mRNA than RNA from glycerol- or glucose-grown cells. That only low amounts of *cen*B mRNA were detected in the RNA prepared from glycerol- or glucose-grown cells was in keeping with the findings of the 5' S1 mapping studies. Together with the data obtained through northern blot and S1 fine- mapping studies, these results showed that the level of *cen*B expression was a function of the carbon source provided to *C. fimi* during growth. It also appears from the data in Fig. 32 that *cen*B transcription is directed from two tandem promoters, *cenBp1* and *cenBp2*, of which the more distal *cenBp1* is inducible as a function of growth on cellulosic substrate.

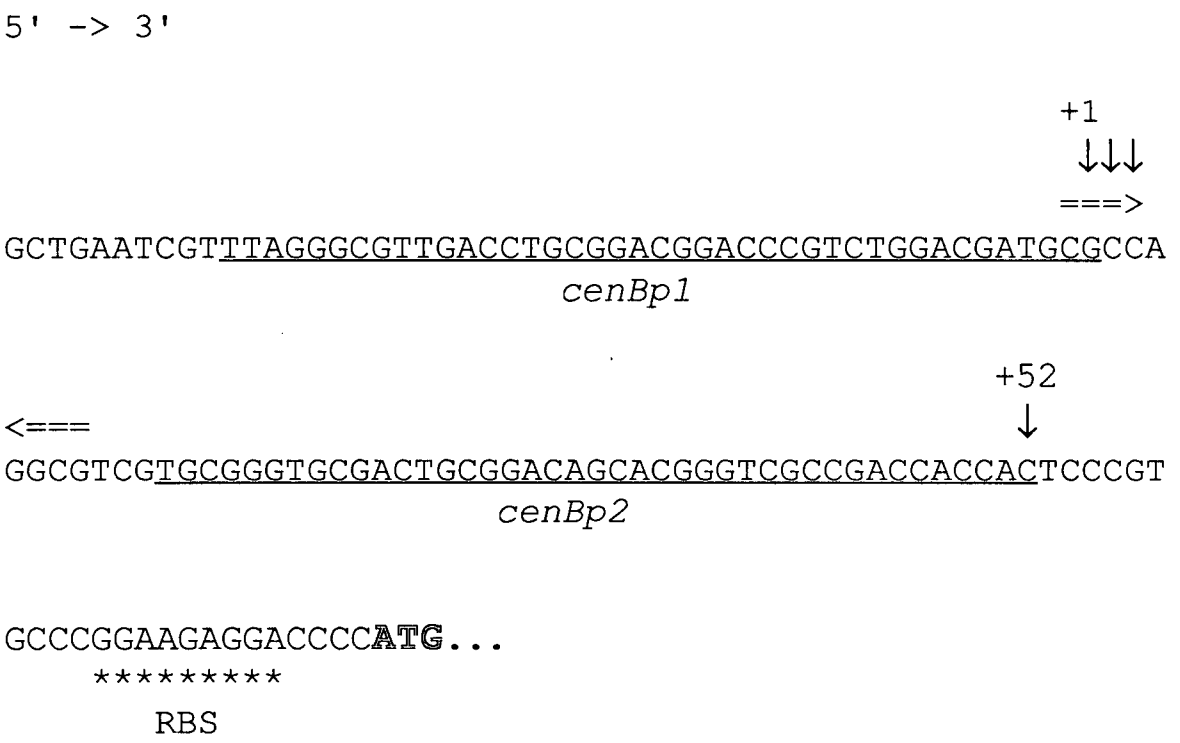


FIGURE 33. DNA sequence corresponding to the 5'-terminal region of *cen*B mRNA. Only the sense strand is shown. The 3' nucleotides of the partially protected 5' S1 probe are shown as vertical arrows. The ATG codon is outlined. The putative ribosome binding site (RBS) is underlined with asterisks. The putative *cenBp1* and *cenBp2* promoter regions are underlined. An inverted repeat is shown overlined with bold arrows.

3.4. Identification of a *C. fimi cex*-linked gene.

3.4.1. Sequence inspection to identify putative *C. fimi* genes.

C. fimi DNA sequences flanking the *cen*A, *cex* and *cen*B genes were inspected to identify putative open reading frames beginning with an initiation ATG codon, preceded by a ribosome-binding site (RBS) and putative *C. fimi* promoter sequences (as found preceding the *cen*A, *cex* and *cen*B genes). One such stretch of DNA was found about 700 bases upstream of, and in the opposite orientation to the *cex* structural gene (Fig. 34). Although the distal DNA sequence contiguous to the *Bam*H1 site had not been determined, the presence of a *C. fimi* promoter-like sequence, RBS sequence and an ATG in appropriate configuration prompted further investigation to identify a transcript and, if found, to map the 5' terminus within the known flanking sequence.

3.4.2. Hybrid protection analysis to confirm the presence of a *cex*-linked gene.

C. fimi RNA synthesized *in vivo* was analysed by high-resolution S1 nuclease mapping. The 644 bp *Bam*H1-*Pst*1 DNA restriction fragment isolated from the plasmid pUC12A25 (Fig. 35) was used as the hybridisation probe. The probe, labeled at the 5' *Bam*H1 site, was denatured and hybridised in solution with total *C. fimi* RNA and S1 nuclease was used to

Carbon source	*cen*B mRNA[†]
Glycerol	32±5
Glucose	24±5
CMC	261±5

[†] amol *cen*B mRNA per µg total *C. fimi* RNA

TABLE VI. Steady state levels of *C. fimi cen*B mRNA. Total RNA was prepared from cultures of *C. fimi* growing exponentially in basal medium supplemented with glycerol (0.2%), glucose (0.2%) or CMC (1.0%) as carbon source. Samples of RNA were immobilized on nylon membranes and filters were hybridised with 5' ^{32}P labeled oligodeoxyribonucleotide probes 30 bases in length and complimentary to the first 10 codons of the *cen*B structural gene (specific activity 5 X 10^7 cpm μg^{-1}). Following hybridisations at 52°C (10 ng probe ml^{-1}; 1 ml per 100 cm^2 membrane), the filters were washed at 52°C with three changes of 2X SSC / 0.1 % SDS and one change of 1X SSC / 0.1 % SDS. The filters were then washed at 60°C with 0.1X SSC / 0.1 % SDS and exposed to X- ray film at -70°C with intensifying screens. Autoradiographic signals were quantitated with a Helena QuickScan densitometer equipped with an integrator. Titrated amounts of the plasmids pNG101, pNG202 and pNG301 specific for the *cen*A, *cex* and *cen*B probes, respectively, served as the internal standards. Alkali- treated (0.2N NaOH for 2 h at 37°C) RNA samples served as the negative controls. All samples applied to the membranes contained 4 µg of total nucleic acid. The results are expressed as the mean (± SD) of two determinations in atom moles (amol) of *cen*B mRNA per µg of total *C. fimi* RNA.

degrade the unprotected portions of the DNA probe in any of the resulting RNA-DNA hybrids. The lengths of the protected probe species still bearing the ^{32}P-labeled 5' *Bam*H1 terminus were determined by gel electrophoretic analysis under denaturing conditions. As seen in Figure 36, an autoradiograph of the analytical gel revealed distinct species of protected probe. These species which migrated closely together all mapped upstream of the RBS- like sequence. No hybrids were seen in control experiments with yeast tRNA (Fig. 36, lane 8). The most prominent species, +1, mapped to an A residue, 48 bases upstream of the ^{32}P-labeled 5' end of the probe, and 39 bases from the ATG proposed as the translation initiation codon for this previously unidentified gene (Fig. 37). The gene has been named *clg* for '*cex*-linked gene'.

The most prominent species of protected probe were observed in mapping experiments with RNA from glycerol-, glucose-, and CMC-grown *C. fimi*. However, they appeared to be less abundant when *C. fimi* was grown on glycerol. Quite in contrast to the data obtained with *cen*A, *cex*, and *cen*B (see section 3.3), expression of *clg* did not appear to be reduced dramatically by growth on glucose as there were abundant *clg* transcripts present in RNA prepared from glucose-grown cells. A comparison of the putative *clg* promoter sequence with other putative *C. fimi* promoters is presented in the Discussion (section 4).

FIGURE 34. Representation of cloned *C. fimi* DNA containing the 5' terminal portion of the *clg* gene and its spatial relationship to the *cex* gene. The restriction endonuclease recognition sites are abbreviated as follows: Bm, *Bam*H1; Ps, *Pst*1; St, *Sty*1. The *C. fimi* DNA sequence proximal to the *Bam*H1 site of pUC12A25 is shown with the ribosome binding site (RBS) and *clg* translation start codon (underlined with an arrow).

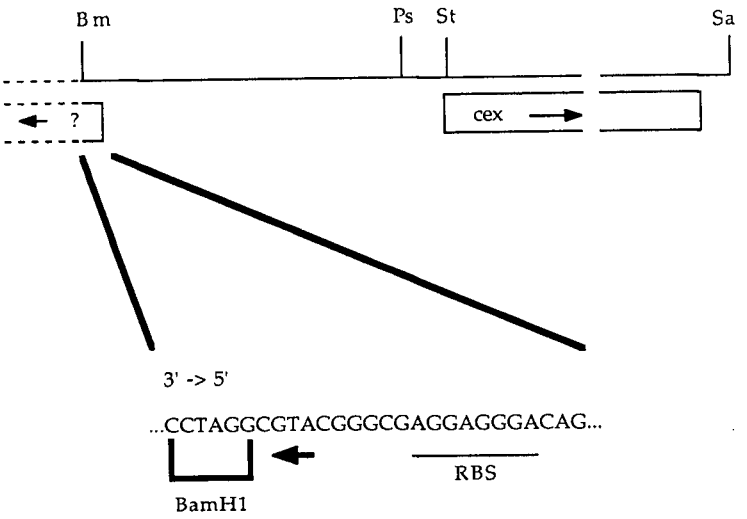


FIGURE 35. Isolation of a hybridisation probe to map the 5' terminus of a putative *C. fimi* gene transcript. The plasmid pUC12A25 was digested with *Bam*H1, treated with CIAP and 5'-labeled with [γ-^{32}P]-ATP. and T4 PNK. Linear 5'-labeled plasmid DNA was then digested with *Pst*1 and fractionated in a 5% polyacrylamide gel. After autoradiography, a gel fragment with the 644 bp *Bam*H1-*Pst*1 probe was excised. The probe was electroeluted into TAE buffer, precipitated with ethanol and recovered by centrifugation.

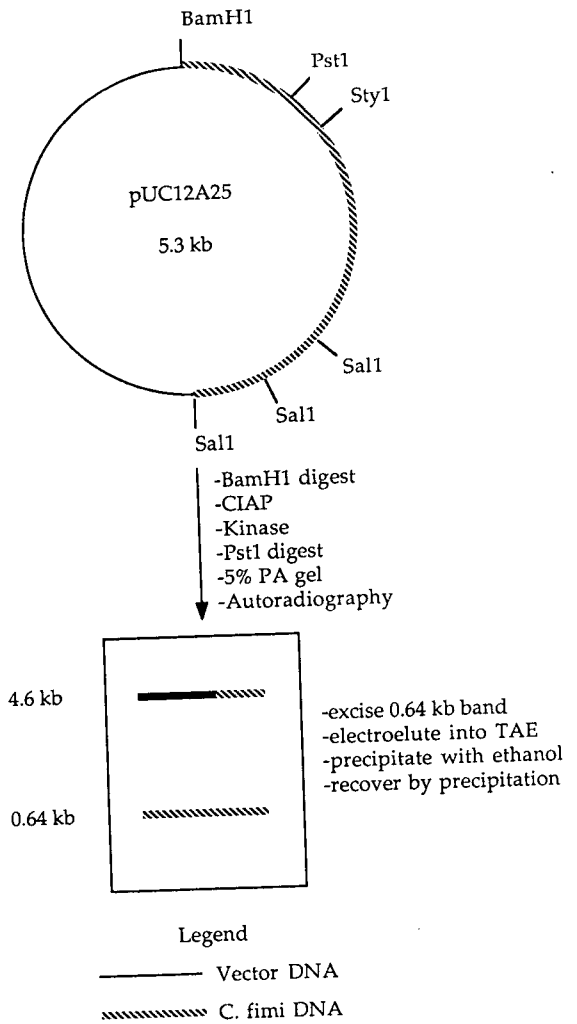
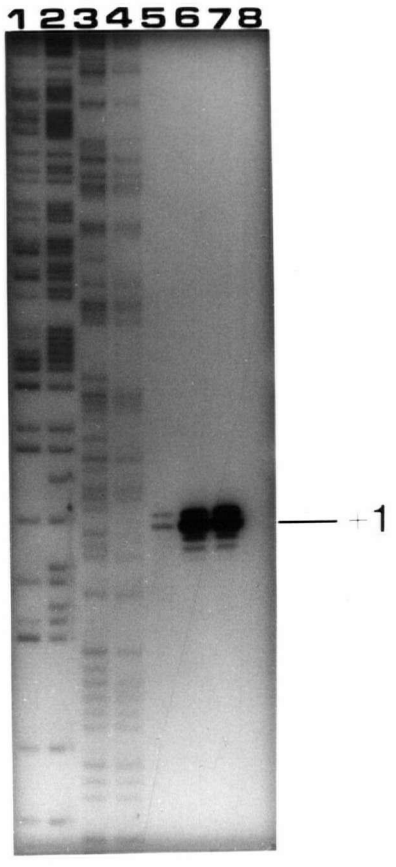


FIGURE 36. Mapping the 5' end of a *cex*-linked gene transcript. After hybridisation with RNA and treatment with S1 nuclease the remaining *Bam*H1-*Pst*1 probe (labeled at the 5' *Bam*H1 site) was analysed in an 8% polyacrylamide -7M urea sequencing gel alongside probe sequenced by the base-specific chemical cleavage method (Maxam and Gilbert, 1980). Protection of the probe by: RNA from glycerol-grown *C. fimi* (lane 5), RNA from glucose-grown *C. fimi* (lane 6), RNA from CMC-grown *C. fimi* (lane 7) and yeast tRNA (lane 8). Lanes 1 through 4 contained the chemical sequencing ladders of the probe: G>A, G+A, T+C, C>T, respectively. Numbers on the right identify the species of protected probe.



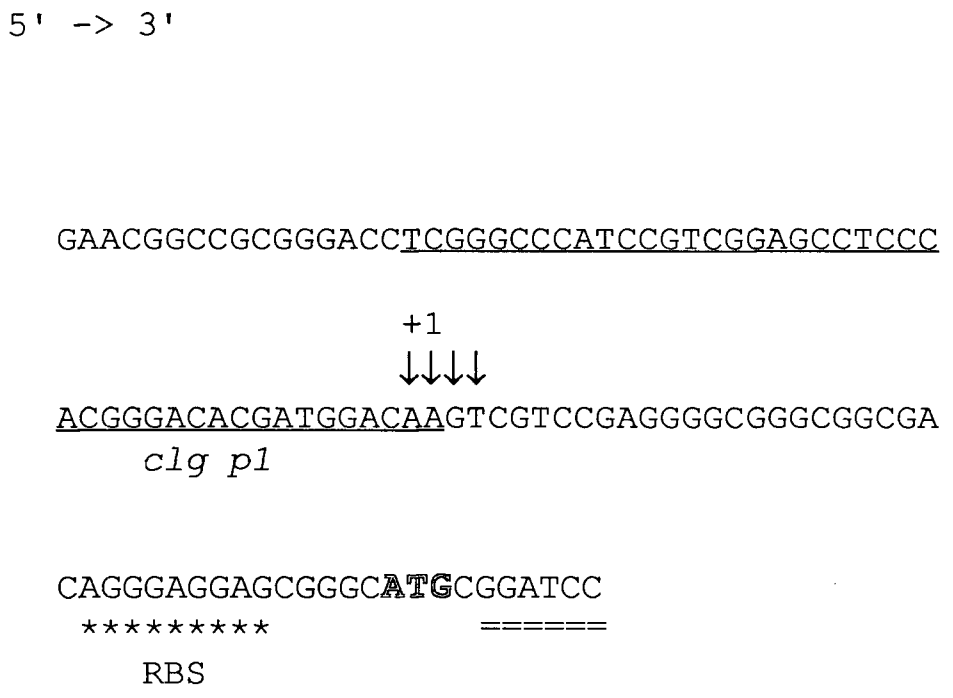


FIGURE 37. DNA sequence corresponding to the 5' terminal portion of *clg*. Only the sense strand of the *Bam*H1 proximal portion of the cloned *C. fimi* DNA carried on pUC12A25 containing *clg* is shown. The arrows denote the 3' nucleotides of the protected fragments of the 5' S1 probe. The ATG codon is outlined. The putative ribosome binding site (RBS) is underlined with asterisks. The putative *clg* promoter region is underlined. The *Bam*H1 site is underlined with double lines.

3.5. Characterization of the *abg* transcripts of *Agrobacterium sp.* Strain ATCC 21400.

3.5.1. Approximate length of *abg* mRNA.

A qualitative northern blot analysis was used to determine the approximate length of *abg* mRNA isolated from *Agrobacterium* sp. Strain ATCC 21400. For this experiment, *Agrobacterium* RNA was prepared from cultures grown in low-salt LB medium. The intragenic 241 bp *Hind*III fragment of *abg* (Fig. 38, A) was used as the hybridisation probe. The probe fragment, carried on the plasmid pTZ19-B (Fig. 39; provided by W.W. Wakarchuk), was labeled by nick-translation. The species of the *in vivo abg* transcript detected by the probe was approximately 1500 bases in length (Fig. 40). This transcript is of sufficient length to encode the *abg* structural gene (see Table I; Wakarchuk et al., 1988).

3.5.2. Mapping the 5' ends of *abg* mRNA.

A hybrid protection study was used to confirm the direction of *abg* transcription and to localise the *abg* transcript 5' ends. Transcripts synthesized *in vivo* were analysed by high resolution S1 nuclease mapping. The 0.4 kb *Eco*R1-*Sty*1 DNA restriction fragment (Fig.38, B) isolated from the plasmid pUC13::Δ9R5 (Fig. 41) was used as the hybridisation probe. The probe, labeled at the 5' *Sty*1 site, was denatured in solution and hybridised with total *Agrobacterium* RNA and

FIGURE 38. Partial restriction map of the *abg* gene. Representation of cloned *Agrobacterium* sp. Strain ATCC 21400 DNA containing the *abg* gene. The structural gene is shown as a boxed region and is translated from left to right (5' -> 3'). A, *Hind*III-*Hind*III intragenic northern blot probe; B, *Eco*R1-*Sty*1 5' S1 probe; C, *Nco*1-*Sal*1 3' S1 probe. The restriction endonucleases are abbreviated as follows: R1, *Eco*R1; H3, *Hind*III; Nc,*Nco*1; RV, *Eco*RV; Sa, *Sal*1; St, *Sty*1.

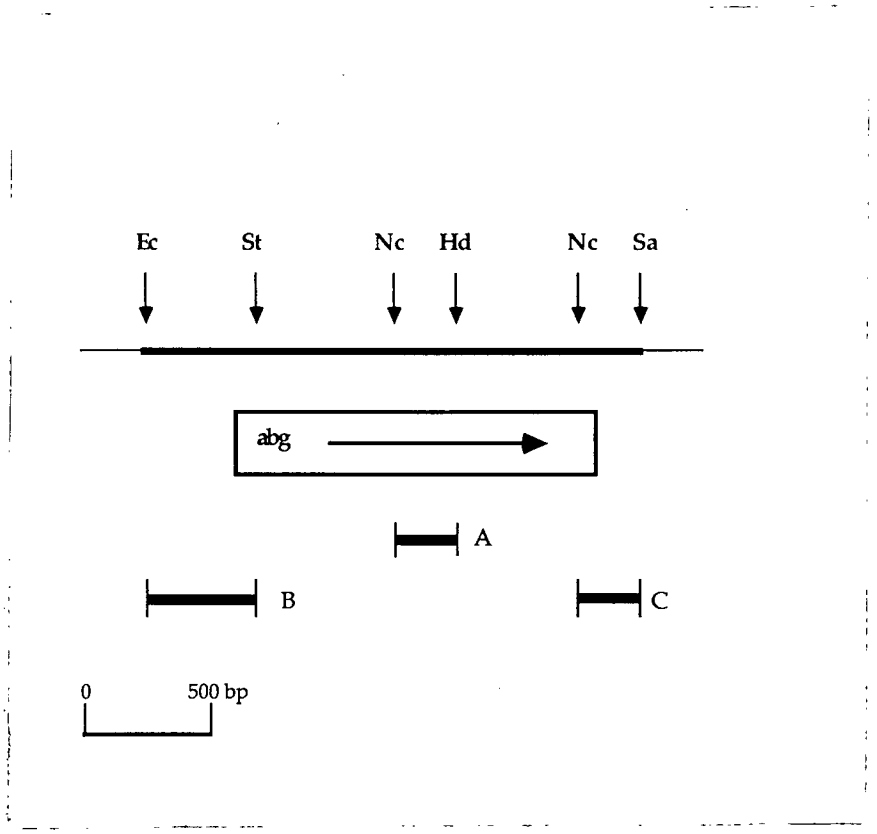


FIGURE 39. Representation of the plasmid pTZ19-B. The 0.24 kb *Hind*III fragment of *abg* is carried on the vector pTZ19R (Vieira, J., U.S. Biochemical company technical literature accompanying the vectors pTZ18/19, 1985). Plasmid DNA was purified by CsCl -EtBr gradient centrifugation (Maniatis *et al.*, 1982). Using [α-^{32}P]-dTTP, dATP, dGTP , dCTP and *E. coli* DNA polymerase I, the plasmid was nick- translated to a specific activity of 5 X 10^6 cpm μg^{-1} DNA and used as the intragenic *abg* hybridisation probe in the northern blot experiments. The restriction endonuclease recognition sites in the vector MCS are abbreviated as follows: Bm, *Bam*H1; H3, *Hind*III; Kp, *Kpn*1; Ps, *Pst*1; R1, *Eco*R1; Sa, *Sal*1; Sc, *Sac*1; Sm, *Sma*1; and Xb, *Xba*1.

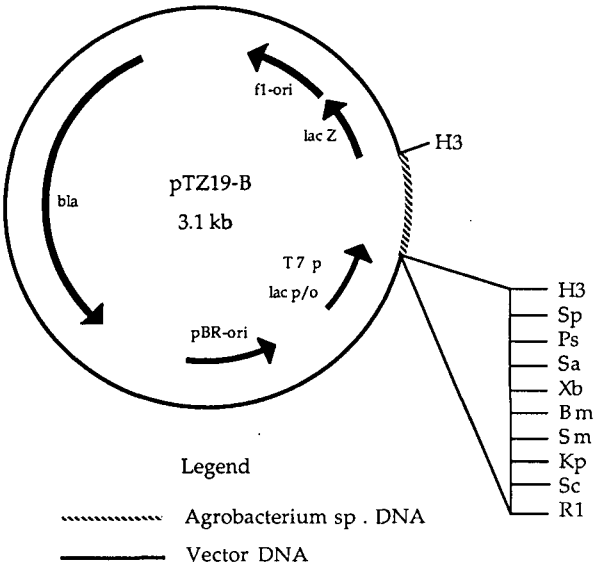


FIGURE 40. Northern blot analysis of *abg* transcripts. RNA was prepared from *Agrobacterium* sp. Strain ATCC 21400 grown in low-salt LB medium, fractionated in a formaldehyde gel containing 1.0% agarose and transferred to a Biotrans membrane. The blot was hybridised with nick-translated plasmid pTZ19-B DNA. The arrow indicates the major hybrid. M, labeled *Hind*III restriction fragments of lambda DNA (sizes in bases are indicated on the left).

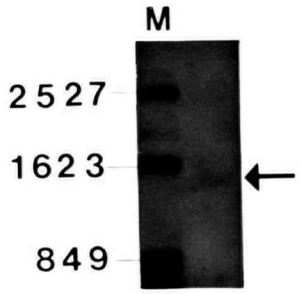
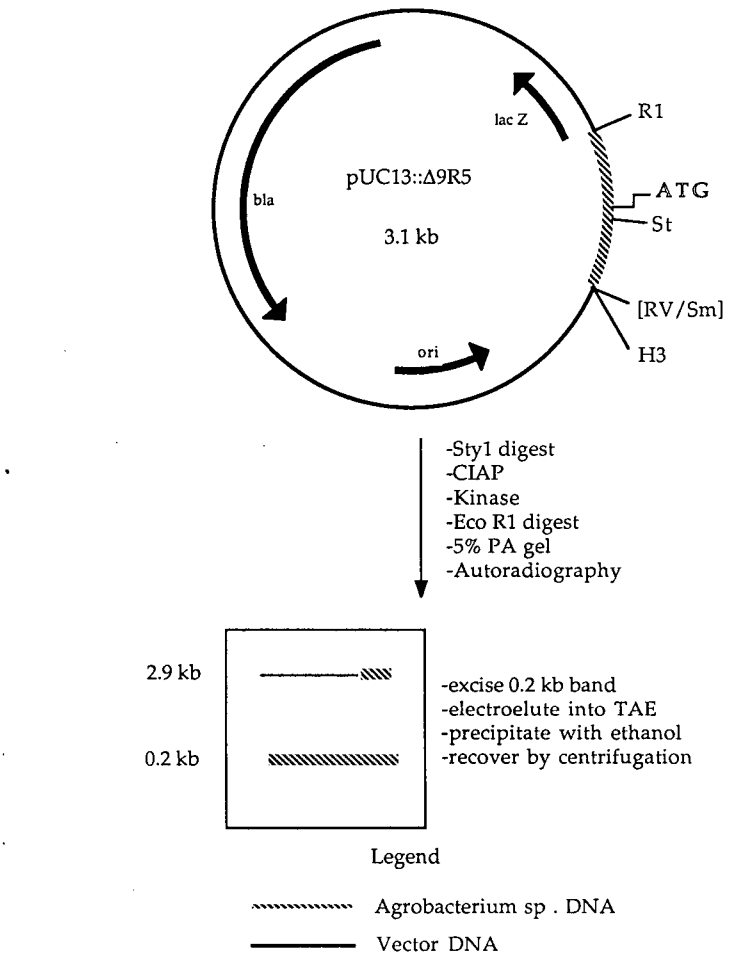


FIGURE 41. Representation of the plasmid pUC13::Δ9R5. An *Eco*R1-*Eco*RV 0.4 kb DNA fragment from *Agrobacterium* sp. Strain ATCC 21400 is carried within the MCS of the plasmid pUC13 (Wakarchuk, 1987). This plasmid was the source of the DNA probe used to detect the 5' terminus of *abg* mRNA in S1 mapping experiments. The plasmid was digested with *Sty*1, treated with CIAP, and 5'-labeled with [γ-^{32}P]-ATP and T4 PNK. The labeled plasmid DNA was then digested with *Eco*R1 and fractionated by electrophoresis in a 5.0% polyacrylamide gel. After autoradiography, a gel slice with the 232 bp *Eco*R1-*Sty*1 fragment labeled at the 5' *Sty*1 site was excised. The probe was electroeluted into TAE buffer, precipitated with ethanol and recovered by centrifugation. Restriction endonucleases are abbreviated as follows: H3, *Hind*III; R1, *Eco*R1; St, *Sty*1;. RV/Sm denotes the hybrid *Sma*1-*Eco*Rv site created by ligation during the construction of the plasmid. The position of the translation initiation (ATG) codon is also shown. The translation of the *abg* gene is in the R1 -> H3 orientation.



S1 nuclease was used to degrade the unpaired portions of the DNA-RNA hybrids. The lengths of the protected probe species still bearing the ^{32}P-labeled 5'-*Sty*1 terminus were then determined by gel electrophoresis under denaturing conditions. As seen in Figure 42, autoradiography of the analytical gel revealed several closely spaced species of probe (lane 5) which may have resulted from the S1 treatment. The major species corresponded to *abg* mRNA 5' end which was 115 bases from the ^{32}P-labeled 5' end of the probe and 22 bases upstream of the translation initiation codon (ATG) determined for the *abg* gene product (Fig. 43; Wakarchuck et al., 1988). These species were not detected in control experiments with yeast tRNA (lane 6).

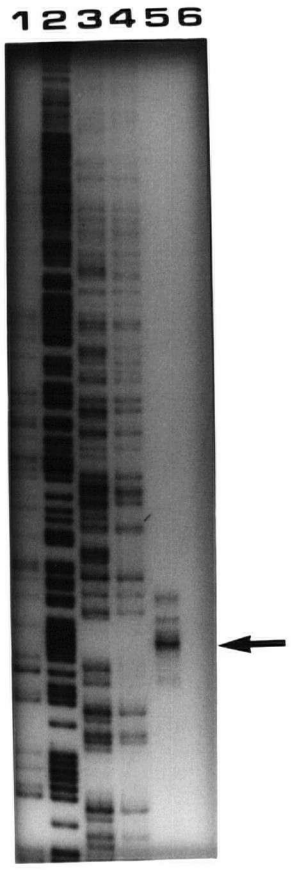
3.5.3. Mapping the 3' end of *abg* mRNA.

To identify the 3' end of *abg* mRNA, transcripts synthesized *in vivo* were analysed by S1 nuclease mapping using the 184 bp *Nco*I-*Sal*1 DNA restriction fragment as the hybridisation probe (Fig. 38, C). The probe, isolated from the plasmid pABG5, was labeled at the 3' *Nco*I site with ^{32}P (Fig. 44). The probe was denatured in solution and hybridised with total *Agrobacterium* RNA and S1 nuclease was used to degrade the unprotected portions of the resulting DNA-RNA hybrids. The length of the protected probe species still bearing the ^{32}P-labeled 3'-*Nco*I terminus was determined by gel electrophoresis under denaturing conditions. As seen

in Figure 45, autoradiography of the analytical gel revealed a distinct species of protected probe (lane 5). This species corresponded to *abg* mRNA 3' end which was 1474 bases from the +1 site (Fig. 49). This is in good agreement with the northern blot data (Fig. 43) which indicated that the transcript was approximately 1500 bp long.

The *abg* mRNA 3' end was 71 bases downstream of the translational stop codon (TGA) of the *abg* structural gene and was also 91 bases from the ^{32}P-labeled end of the probe DNA. This species of probe was not observed in the control experiments with yeast tRNA (Fig. 48, lane 6). From the results of the northern blot and S1 mapping experiments, and since no open reading frame was found immediately downstream of the *abg* stop codon, the *abg* gene appears to be monocistronic.

FIGURE 42. Mapping the 5' end of *abg* mRNA . Following hybridisation of the *Eco*R1-*Sty*1 DNA with RNA (lane 5, *Agrobacterium* sp. Strain ATCC 21400 RNA; lane 6, yeast tRNA) and digestion with S1 nuclease, probe DNA was analysed in an 8.0% polyacrylamide-7M urea sequencing gel alongside probe sequenced by the chemical cleavage method (Maxam and Gilbert, 1980). Lanes 1 through 4 contain the sequence ladders G>A, G+A, T+C, C>T, respectively. The arrow indicates the migration of the major protected probe species.



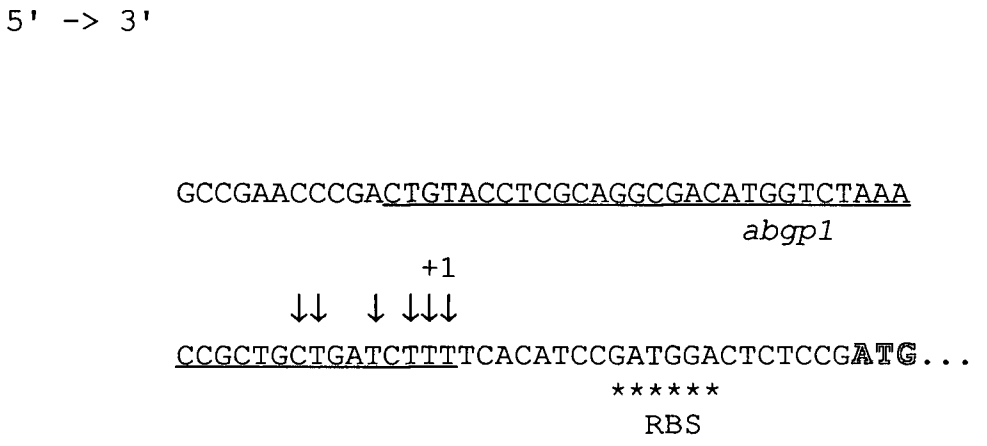


FIGURE 43. DNA sequence corresponding to the 5'-terminal region of *abg* mRNA. Only the sense strand is shown. The arrows denote the 3' nucleotides of the protected fragments of the 5' S1 probe. The ATG codon is outlined. The putative ribosome binding site (RBS) is underscored with asterisks. The putative *abgp1* promoter sequence is underlined.

FIGURE 44. Representation of the plasmid pABG5. An *Eco*R1-*Hind*III 1.6 kb DNA fragment of *Agrobacterium* sp. Strain ATCC 21400 is carried on the plasmid pUC18 (Wakarchuk, 1987). This plasmid was the source of the DNA probe used to detect the *abg* mRNA 3'-terminus in S1 mapping experiments. The plasmid was digested with *Nco*1, and 3'-labeled with [α-^{32}P]-dATP, dCTP and the Klenow fragment of DNA polymerase I to a specific activity of 5 X 10^6 cpm μg DNA^{-1}. The labeled plasmid DNA was then digested with *Sal*1 and fractionated by electrophoresis in a 5.0% polyacrylamide gel. After autoradiography, a gel slice with the 183 bp *Nco*1-*Sal*1 fragment labeled at the 3' *Nco*1 site was excised. The probe was electroeluted into TAE buffer, precipitated with ethanol and recovered by centrifugation. The restriction endonuclease recognition sites are abbreviated as follows: Nc, *Nco*1; Sa, *Sal*1; and St, *Sty*1. Sm/Exo denotes the hybrid *Sma*1 site generated from the ligation of a blunt-ended exonuclease III deletion fragment into the MCS *Sma*1 site of pUC18 (Wakarchuk, 1987).

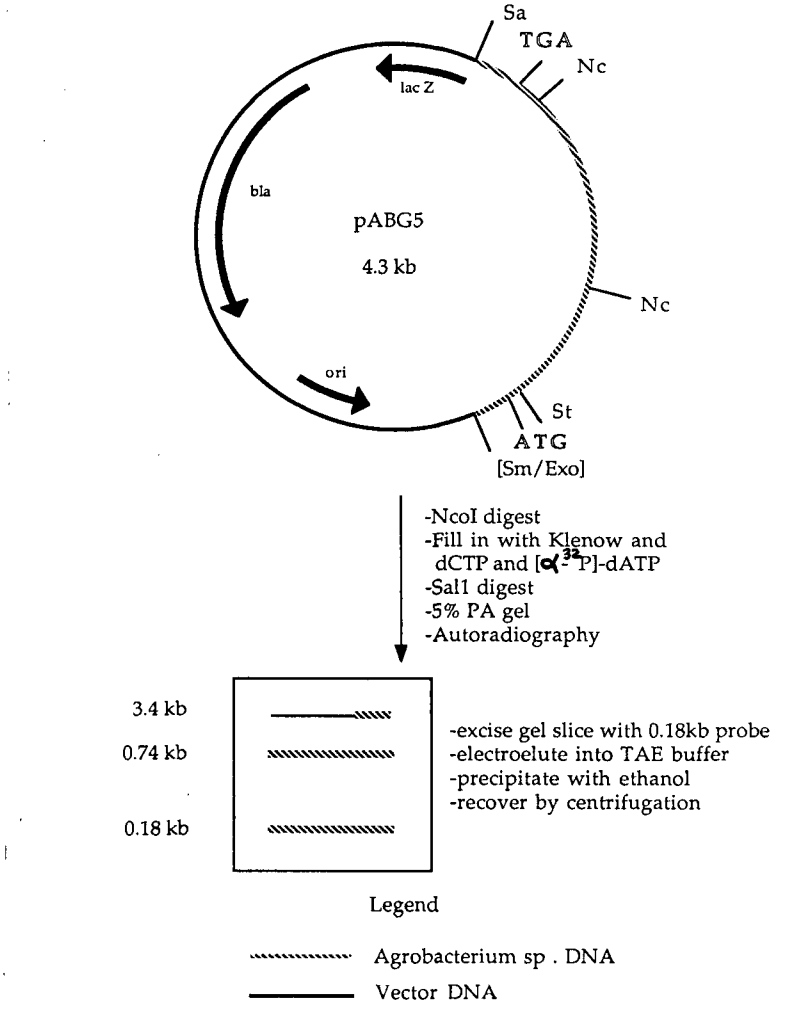
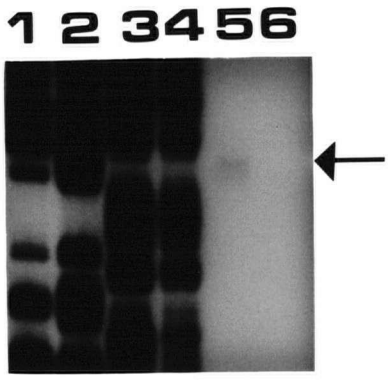


FIGURE 45. Mapping the 3' end of *abg* mRNA. Following the hybridisation of the *Nco*1-*Sal*1 DNA with RNA (lane 5, *Agrobacterium* sp. Strain ATCC 21400 RNA; lane 6, yeast tRNA) and digestion with S1 nuclease, probe DNA was analysed in an 8.0% polyacrylamide-7M urea sequencing gel alongside probe sequenced by the chemical cleavage method (Maxam and Gilbert, 1980). Lanes 1 through 4 contain the sequence ladders G>A, G+A, T+C, C>T, respectively. The arrow indicates the migration of the protected probe.



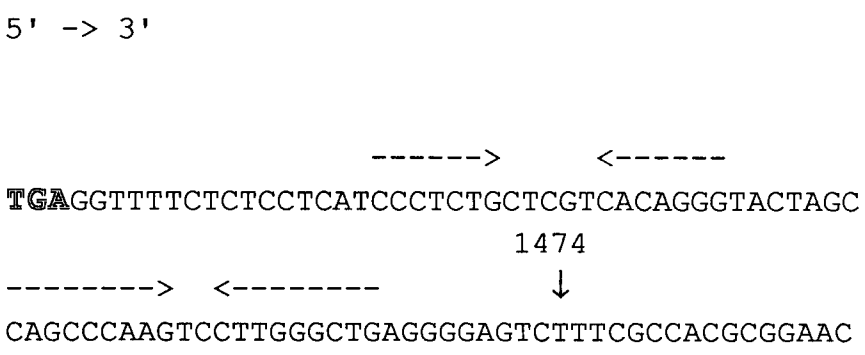


FIGURE 46. DNA sequence corresponding to the 3'- terminal region of *abg* mRNA. Only the sense strand is shown. The vertical arrow denotes the 5' nucleotide of the protected fragment of the 3' S1 DNA probe. The TGA stop codon is outlined. The nucleotide sequences which may form stem-loop structures are shown as opposing horizontal arrows.

4. Discussion

4.1. The *cen*A, *cex* and *cen*B transcripts of *C. fimi*

The *in vivo* transcripts of the *cen*A, *cex*, and *cen*B genes encoding the extracellular enzymes EngA, Exg and EngB, respectively, of *C. fimi* strain ATCC 484 have been investigated. By northern (RNA) blot analysis of *C. fimi* RNA prepared from cells grown on the soluble cellulosic substrate CMC, the *cen*A, *cex* and *cen*B genes appeared to be transcribed as mRNAs 1,400, 1,500 and 3,200 bases long, respectively. While the data suggests that these mRNA are monocistronic, transcription and RNA processing could also result in apparent monocistronic mRNA which were actually generated from longer transcripts. This is also discussed below with the S1 mapping data.

Both *cen*A and *cen*B mRNAs were detected by northern blot analysis in RNA prepared from cells grown on glycerol, a carbon source not known to induce, but which may repress cellulase expression in *Cellulomonas spp.* (Beguin *et al.*, 1977). No *cex* mRNA was detected by northern blot analysis in these preparations. Although not quantitative, the northern blot data did indicate that under non-inducing growth conditions the relative levels for *cen*A and *cen*B mRNAs (both encoding endoglucanases) appeared to be greater than those observed for *cex* mRNA (encoding an exoglucanase). These findings suggested that the substrate provided during *C. fimi*

growth was somehow affecting the expression of *cen*A, *cex* and *cen*B.

The *in vivo cen*A, *cex* and *cen*B mRNA 5' termini were localised by nuclease S1 hybrid-protection studies using 5'-^{32}P-labeled DNA probes. It should be noted that this type of analysis only defines the ends of the transcripts, and these ends do not necessarily correspond to true transcript start and stop sites (i.e. they could map mRNA termini which result from the processing of longer 'read-through' transcripts). Therefore, the S1 mapping studies were used to localise the regions of *C. fimi* DNA potentially responsible for directing the initiation and termination of transcription since no other system was yet available to define these control elements. It should also be mentioned that these were the first experiments to localise the *C. fimi* transcriptional regulatory regions and no other *C. fimi* promoter sequences were available for direct comparisons. Further genetic analysis is therefore required to unequivocally identify the functional *C. fimi* promoters and terminators of these genes.

With the *cen*A mRNA 5'-specific probe, four *cen*A mRNA 5' termini were found between 62 and 50 bases upstream from the ATG codon, three closely spaced and the fourth about 11 bases further upstream (Fig. 6). Since heterogeneity due to S1 artifacts are often associated with the analysis of G+C rich DNA, this data at best suggested *cen*A transcription was being directed from two promoters (Fig. 7): promoter *cenAp1*

directing transcription from position +1 (which appeared to be the strongest autoradiographic signal in Fig. 6), and the promoter *cenAp2* which was directing transcription from position -11 (a weaker signal relative to +1 in Fig. 6,).

With the *cex* mRNA 5'-specific S1 DNA probe, four possible *cex* mRNA 5' termini were identified between 29 and 26 bases upstream from the ATG codon. Taking into account the 'end-heterogeneity' inherent in S1 analysis, the transcription of the *cex* gene appeared to be directed from a single promoter, *cexp1* (see Fig. 21).

To demonstrate that true 5' termini had been localised for both the *cen*A and *cex* genes, the vaccinia virus capping enzyme was employed to *in vitro* label the 5' termini of *C. fimi* transcripts synthesized *in vivo*. These capped transcripts were used in hybrid protection studies with unlabeled DNA restriction fragment probes. Since primary transcripts were the only suitable substrates for the labeling reaction, then only such transcripts, *a priori*, could be protected by *cen*A or *cex* specific DNA probes in the hybrid protection studies. As a result of these procedures it was confirmed that true 5' termini had been mapped for both the *cen*A and *cex* genes (Fig. 9 and 23). However, it was not possible to demonstrate the presence of the *cen*A distal (*cenAp2* directed) start site by this approach, possibly owing to the rather low specific-activity of the capped RNA (about 1/10 that of the ^{32}P-labeled DNA probe)

coupled with the relatively low abundance of the mRNA species (as determined from the autoradiograms of S1 analytical gels). Therefore, identification of a *cenAp2* promoter was based solely on the results of the S1 mapping data using the labeled DNA probe.

Hybrid protection studies were also used to localise the *in vivo cen*B mRNA 5' termini. Three 5' termini were resolved in the S1 experiments. These were clustered upstream of the *cen*B ATG translation initiation codon. Similar clusterings were also observed in the *cen*A and *cex* mapping studies, possibly reflecting some flexibility in the selection of a transcript initiation site by *C. fimi* RNA polymerase or the heterogeneity inherent in S1 analyses as already mentioned. It should be mentioned that similar clusterings of protected species have been observed in the mapping experiments conducted by another investigator in our lab who was mapping the 5' ends of *C. fimi cen*C mRNA encoding endoglucanase EngC (B. Moser, personal communication).

A less abundant *cen*B transcript 5' end was found 52 bases closer (+52) to the *cen*B ATG codon and was detected in mapping experiments with RNA from *C. fimi* grown on any one of the three substrates. The +52 species was a very minor product of the analysis and the only control which suggested the authenticity of this species was the control experiment with yeast tRNA. Mapping experiments with capped RNA were not performed for *cen*B as the costs of these analysis had

become prohibitive.

At best, the *cen*B mapping data can be interpreted to suggest two tandem functioning promoters (see Fig. 33): the distal *cenBp1* directing transcription predominantly from the +1 position (the strongest signal on the autoradiograms) and the proximal (and very tentative) *cenBp2* directing transcription from position +52. Transcription from *cenBp1* was evidently induced as a function of *C. fimi* growth on CMC while *cenBp2* appeared to direct low-level (and most likely constitutive) transcription that was not induced as a function of CMC-supported growth. Why fewer transcripts might initiate at site +52 in cells grown in CMC-medium than in cells grown in either glycerol- or glucose-medium may result from initiations at the *cenBp1* promoter interfering with the frequency of initiations at the *cenBp2* promoter. This tandem arrangement of a regulated, (inducible) promoter and a constitutive promoter closely resembles that reported for the promoters of the *Streptomyces lividans* galactose operon (Fornwald *et al.*, 1987).

The finding that predominantly G and C residues corresponded to the 5' ends mapped by hybrid protection analysis for all but the *clg* transcripts (which initiated with A) was not unexpected considering the high G+C content of *C. fimi* DNA. However, this observation is of interest since the majority of prokaryotic gene transcripts so far mapped initiate with pppG or pppA (Hawley and McClure, 1983;

Rosenberg and Court, 1979; Moran *et al.*, 1982). Another interesting observation was that although *C. fimi* DNA is very G+C rich, the 5' termini mapped for the *cen*A, *cex* and *cen*B transcripts fell within regions that were relatively A+T rich, characteristic of the transcriptional regulatory regions observed for most prokaryotic genes (Hawley and McClure, 1983; Rosenberg and Court, 1979; Moran *et al.*, 1982).

To identify regions which could represent the DNA sequences recognized by *C. fimi* RNA polymerase, comparisons were made between the sequences immediately preceding the mapped 5' termini themselves to see if any common DNA sequences could be found. As well, comparisons were made to the characterized prokaryotic promoter '-10' and '-35' regions of other organisms to see whether the *C. fimi* DNA sequences in these regions matched with known promoter sequences. These comparisons were also used in the identification of the *clg* gene by sequence inspection. The best -35 and -10 region similarities are summarized in Figures 47A and 47B.

FIGURE 47. Promoter region similarities. A. The DNA sequences (sense strand) flanking the mapped *cen*A, *cex*, *cen*B, *clg*, and *abg* transcriptional start sites are compared for similarities with the -35 and -10 regions of characterized promoters pBR322 *tet* (Hawley and McClure, 1983); *E. coli* σ^{70} consensus promoter (Hawley and McClure, 1983); *tsrp1* (Hopwood *et al.*, 1986); *ermp1* and *ermp2* (Bibb, Janssen and Ward., 1985). B. The promoter regions of the *C. fimi* genes are compared to each other to identify their -35 and -10 region similarities. The mapped 5' ends are underlined.

A.

PROMOTER '-35' AND '-10' REGION SIMILARITIES

5' -> 3'

		-35	-10
1.	*cenAp1*:	TCCTCATCCGCTCGCGCCGTGGGGCATTCGTCGGGTTTCCTCGTCGGG	
	pBR322 *tet*	TTGACA----------TTTAAT	
	ermp1	TGGACA----------TAGGAT	
2.	*cenAp2*:	CGATTAGGAAATCCTCATCCGCTCGCGCCGTGGGGCATTCGTCGGGTT	
	E. coli σ^{70}	TTGACA----------TATAAT	
3.	*cenBp2*:	GCCAGGCGTCGTGCGGGTGCGACTGCGGACAGCACGGGTCGCCGACCACCACTC	
	tsrp2	AGGGCA----------TAGGGT	
4.	*cexp1* :	TTCAGCACCTCCCGCGGACGGGCCCCCACGTCACAGGGTGCACCCG	
	tsrp2	AGGGCA----------TAGGGT	
5.	*cenBp1*:	TTTAGGGCGTTGACCTGCGGACGGACCCGTCTGGACGATGCGCC	
	ermp2	TTGACG----------GAGGAT	
6.	*clgp1* :	CCTCGGGCCCATCCGTCGGAGCCTCCCACGGGACACGATGGACAAGT	
	ermp2	TTGACG----------GAGGAT	
7.	*abgp1*	GACTGTACCTCGCAGGCGACATGGTCTAAACCGCTGCTGATCTTTTC	
	E. coli σ^{70}	TTGACA----------TATAAT	
	nif	GGTAT-------TGCT	
		C C A	

B.

C. fimi PROMOTER REGION SIMILARITIES

5' -> 3'

	'-35'	'-10'	5'->
cenAp1:	TCCTCATCC **GCTCGCG**	CCGTGGGGCATTCGTC **GGGTTT**	CCTCGTCGGG
cenAp2:	CGATTAGGA **AATCCTC**	ATCCGCTCGCGCCGTGG **GGCATT**	CGTCGGGTT
cenBp2:	GCCAGGCGT **CGTGCGG**	GTGCGACTGCGGACAGC **ACGGGT**	CGCCGACCACCACTC
cexp1 :	TTCAGCACC **TCCCGCG**	GACGGGCCCCCACGTCA **CAGGGT**	GCACCCG
cenBp1:	TTTAGGGCG **TTGACCT**	GCGGACGGACCCGTCTG **GACGAT**	GCGCC
clgp1 :	CCTCGGGCC **CATCCGT**	CGGAGCCTCCCACGGGA **CACGAT**	GGACAAGT

After comparison with known promoter sequences, some matches (for -10 and -35 region similarities) were observed between the *Streptomyces spp.* promoters and the *C. fimi* sequences immediately upstream of the mapped *cen*A, *cex* and *cen*B mRNA 5' termini. In the absence of functional data on *C. fimi* promoters, not much can be drawn in the way of conclusions from these comparisons. However, that the best matches were to the *Streptomyces sp.* promoters was not surprising, since *Streptomyces sp.*, like *C. fimi*, are gram-positive bacteria with high G+C DNA. Therefore, one might expect that the sequences which signal the promotion of transcription in these bacteria might be similar. For example, immediately upstream of the *cen*B +1 site (putative *cenBp1*) the sequence TTGACC is very similar to the *Streptomyces erythraeus ermp2* -35 sequence TTGACG (Bibb, Janssen and Ward, 1985) and the *C. fimi* sequence GACGAT is very similar to the *Streptomyces* -10 sequence GAGGAT with a reasonable 18 bp spacing between the two stretches of *C. fimi* sequence. The putative *cexp1* sequence CAGGGT showed a similarity with *Streptomyces tsrp1* -10 region sequence TAGGGT (Bibb *et al*, 1985). As well, the putative *clgp1* promoter sequence CACGAT showed homology with the *ermp2* promoter -10 sequence GAGGAT. It is curious that the best promoter matches were -10 region matches between the *C. fimi* cellulase encoding genes and the antibiotic resistance genes of *Streptomyces*, with fewer -35 region similarities. Possibly,

the *C. fimi* RNA polymerase recognizes different binding site sequences ('-35') than the *Streptomyces* polymerase, yet they may recognise similar sequences ('-10') which signal for the initiation of transcription.

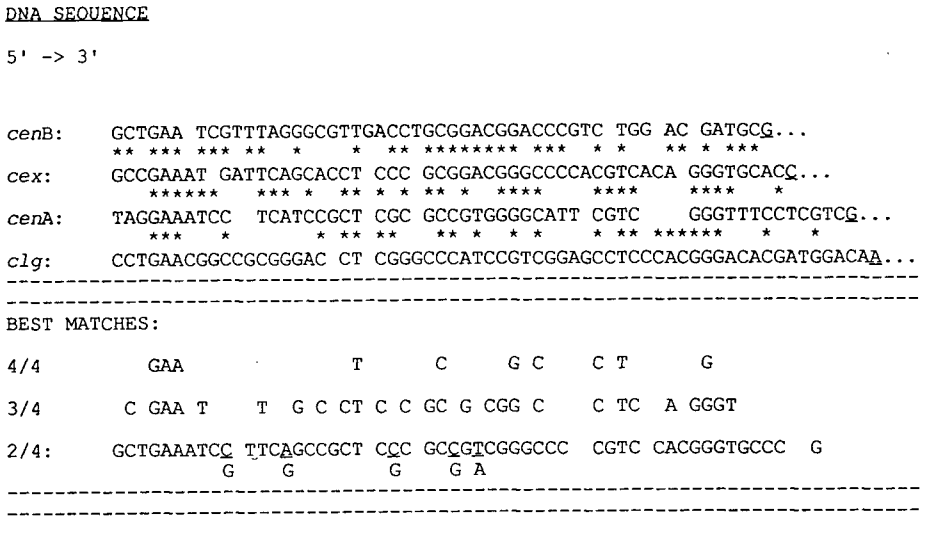
Mapping studies have been performed to determine whether putative *C. fimi* promoters were functional in heterologous hosts. With RNA isolated from *E. coli* cells carrying recombinant plasmids with putative *C. fimi* promoter DNA, transcripts were not found which initiated within the inserts (Greenberg, *et al*, 1987b). The *cen*A and *cex* genes from *C. fimi* have also been placed on a plasmid shuttle-vector to facilitate their engineering in *E. coli* and subsequent introduction into the non-cellulolytic gram-positive bacteria *Brevibacterium lactofermentum* (Paradis *et al.*, 1988). Although the *cen*A and *cex* genes were functionally expressed in the *Brevibacterium*, S1 mapping experiments have since demonstrated the sequences which signal the initiation of transcription of these genes in *Brevibacterium* were probably within the vector DNA and not within the flanking *C. fimi* DNA inserts (F. Paradis, personal communication). There are at least three possibilities as to why the appropriately initiating transcripts from the recombinant cellulase genes may not have been detected in the heterologous hosts. Firstly, the *E. coli* and *Brevibacterium* RNA polymerases may not have been able to recognize the *C. fimi* promoters. Secondly, the RNA polymerases may have recognized the

promoters but were incapable of initiating or elongating the transcripts. Thirdly, the resulting 'hybrid' transcripts may have been intrinsically unstable.

Previously determined DNA sequences flanking the *cen*A, *cex* and *cen*B genes were inspected for regions fitting a 'promoter- ribosome binding site- ATG codon' paradigm. This lead to the identification of a previously uncharacterized gene, upstream of, and divergent to, the *cex* gene. This 'found gene' has been designated *clg* for '*cex*- linked gene'. Hybrid protection analysis has confirmed the existence of a *clg* transcript and has localised the *clg* mRNA 5' terminus. This approach proved that the putative *cen*A, *cex* and *cen*B promoter sequences already localised through hybrid-protection analysis could be used to identify other transcriptional units by inspection of *C. fimi* DNA. While the structure of *clg* is not yet known (the entire gene has yet to be cloned and sequenced) and the function of the *clg* gene-product remains uncharacterized, it was interesting to find *clg* RNA in abundance from both glucose- and CMC-grown cells.

A comparative analysis of the 5' flanking DNA sequences of the *cen*A, *cex*, *cen*B, and *clg* genes has revealed a region of 50 bases in which these four sequences display at least 64% homology (Fig. 48). The most important feature from these comparisons was that in each case, a 'GAA-box' could be

FIGURE 48. Conserved DNA sequences flanking mapped 5' mRNA ends of *C. fimi* genes. Gaps have been introduced into the sequences to allow for best matches. Matches are denoted by asterisks. The +1 sites for *cenBp1*, *cexp1*, *cenAp1*, and *clg* are underlined, and transcripts are indicated by dots. Shown are the nucleotides which are 100% (4/4), 75% (3/4) and 50% (2/4) conserved.



found upstream of the mapped mRNA 5' termini. Such a 'GAA-box' has been shown to be present upstream of the *cen*C gene (B. Moser, unpublished observations). This conserved 5' flanking sequence may represent a cis-regulatory element involved in the expression of the *C.fimi* genes. It should be interesting to see, once a suitable gene-transfer system is developed for re-introduction of cloned DNA into *C. fimi*, which structural features of the 5' flanking regions are in fact recognized either *in vivo* or *in vitro* by the *C. fimi* transcription machinery.

To see whether the regions localised by S1 mapping do represent promoters or recognition sites for regulatory proteins, gel retardation studies could be performed with fractionated *C. fimi* protein extracts and the end-labeled DNA fragments used as 5' probes in the mapping experiments. As well, methylation protection (DNA footprinting) studies could also be performed towards the same end. These experiments are designed to give a high resolution picture of those regions of flanking DNA which serve as contact points for DNA binding proteins or RNA polymerase.

Transcript 3' termini were mapped for the *cen*A and *cex* mRNAs by hybrid protection analysis with end-labeled DNA probes. These experiments identified 3' ends which may have arisen either from true transcript termination or via processing of longer transcripts. Inverted repeats were found immediately downstream of the *cen*A and *cex*

translational stop codons (a TGA for both genes), and the 3' termini mapped closely to these regions. Since these structures resembled rho-independent termination signals as found in *E. coli* (Rosenberg and Court, 1979), although none of these *C. fimi* structures precede a run of consecutive T residues as found in *E. coli*, they may represent *C. fimi* transcription terminator signals. It should be interesting to see whether *cen*B and *cen*C mRNA 3' termini are identified downstream of similar sequences.

The steady state levels for *cen*A, *cex* and *cen*B mRNAs were determined by hybridisations with synthetic oligonucleotide probes labeled to high specific activities. Interestingly, the transcripts of the two endoglucanase genes, *cen*A and *cen*B were found to be present at similar levels (about 250 amol $\mu g\ RNA^{-1}$) under inducing (CMC) growth conditions while the steady state *cex* transcripts were found to be less abundant (77 amol $\mu g\ RNA^{-1}$; or 31% of the *cen*A and *cex* levels) under similar growth conditions. The level of *cen*A mRNA under glycerol-growth ($\approx$50 amol $\mu g\ RNA^{-1}$) was found to be slightly greater than for either *cen*B ($\approx$30 amol $\mu g\ RNA^{-1}$) or *cex* ($\approx$20 amol $\mu g\ RNA^{-1}$) mRNAs. That under non-inducing growth conditions the northern blots had detected levels of both *cen*A and *cen*B mRNAs but not *cex* mRNA may reflect the relatively lower specific-activities of the northern blot probes relative to the slot-blot oligo-probes.

The transcript levels were clearly lowest in the RNA

samples from glucose-grown cells. The steady state level of *cen*A mRNA was reduced some 20-fold as a function of glucose-growth when compared to CMC-growth. The *cex* and *cen*B mRNA levels were similarily reduced 5- to 6-fold as a function of glucose-growth. Glycerol may repress *cen*A expression, although the strength of this repression may be less than observed with glucose. The higher level of *cen*A mRNA compared to either *cex* or *cen*B mRNA in the presence of glycerol could be explained by either *cen*A being transcribed at higher levels (more initiations) or that *cen*A mRNA is more stable under glycerol growth.

It is tempting to speculate on the molecular mechanisms governing cellulase gene expression in *C. fimi*. The patterns of gene expression observed at the transcriptional level as functions of the substrate provided during *C. fimi* growth are consistent with repressible and inducible regulatory mechanisms, although neither a repressor protein nor true inducer moeity has ever been demonstrated. However, the elucidation of inducible and constitutive transcription, does imply that this organism can discriminate between various carbon sources and respond in a manner controlled, at least in part, at the transcriptional level. This is in contrast to *Clostridium thermocellum* which produces a number of constitutive endoglucanase activities throughout its vegatative growth phase (Hammerstrom *et al.*, 1955; Garcia-Martinez *et al.*, 1980).

In latter experiments (not presented here), the steady state levels of *cen*A, *cex* and *cen*B mRNAs in RNA prepared from cells grown in the presence of CMC and glucose could not be demonstrated to be significantly greater than the levels observed for similar cultures grown on glucose alone. However it was noted that *C. fimi* grown on glucose + CMC supplemented plate agar did produce exoglucanase activity as determined by the hydrolysis of the synthetic substrate 4-methylumbelliferyl-β-D-cellobioside (MUC). No MUCase activity was detected from *C. fimi* grown on plate agar supplemented with glucose alone. An explanation for this disparity is not yet known, but may involve as yet uncharacterized components of the *C. fimi* cellulase system.

A low level of constitutive expression of extracellular cellulases, most notably endoglucanases, has been observed in many cellulolytic organisms (Beguin *et al.*, 1987; Coughlan, 1985; Canevascini *et al.*, 1979; Enari and Niku-Paavola, 1987). *C. fimi* grown on glucose- or glycerol-supplemented minimal media agar still expresses some endoglucanase activity (Greenberg *et al.*, unpublished observation) although which endoglucanases are elaborated is not known. A constitutive low-level expression of one or two endoglucanase activities by *C. fimi* seems preferable to the production of a full complement of cellulases in the absence of an appropriate substrate. It is easily envisaged that once cellulose were encountered, it could be hydrolyzed and or modified by the extracellular endoglucanase activities to

produce an (as of yet uncharacterized) inducer moeity. The inducer, upon entering the cell would then be able to effect cellulase gene expression. The drop in the intracellular level of the global inducer (either through metabolism or repressor titration) would probably coincide with exhaustion of cellulosic substrate, and then *C. fimi* would return to basal-level cellulase production.

That the induction of cellulase gene expression in *C. fimi* should occur in a trans-acting (i.e. global) manner is suggested both by the apparent monocistronic nature of the cellulase genes so far characterized (no polycistronic mRNA encoding cellulases have yet been identified; see also Beguin *et al.*, 1986) and that these genes do not appear to be tightly linked at the DNA level as determined from linkage studies with a *C. fimi* lambda- library (N.M. Greenberg, unpublished observations). These observations seem to suggest both cis- and trans-acting regulatory elements facilitating coordinate cellulase expression/repression in response to changing environmental conditions.

Catabolite repression (Jacob and Monod, 1961) has been suggested as one possible form of regulatory control of cellulase biosynthesis in *C. fimi* (Beguin *et al.*, 1977; Greenberg *et al.*, 1987a,b) and other prokaryotic and eucaryotic cellulolytic organisms (Coughlan, 1985). However the molecular mechanism of this repression in *C. fimi* has remained uncharacterized. The results which demonstrated the

low-levels of cellulase gene expression in glucose-grown cells could be explained by catabolite repression, although the addition of cyclic AMP (cAMP) to growing cultures did not appear to overcome this repression (Beguin *et al.*, 1977, and N.M. Greenberg, unpublished observations) as has been observed in *E. coli*.

4.2. The *abg* transcripts of *Agrobacterium* sp. Strain ATCC 21400.

The *in vivo* transcripts of the *abg* gene encoding the β-glucosidase Abg of *Agrobacterium* sp. were also investigated in these studies. The β-glucosidases are essential in cellulolysis, as they are responsible for the hydrolysis of cellobiose to glucose, which the cell can use as a source of carbon and energy. By northern blot analysis of *Agrobacterium* sp. RNA, the *abg* gene was found to be transcribed as a mRNA of approximately 1,500 bases in length. The *in vivo abg* mRNA 5' termini were localised by nuclease S1 hybrid protection analysis with a 5'-labeled DNA probe. The major 5' end mapped 22 bases upstream of the *abg* translation initiation codon. To identify regions which could represent the DNA sequences recognized by *Agrobacterium* sp. ATCC 21400 RNA polymerase, comparisons were made between the sequence immediately upstream of the mapped 5' end and characterized prokaryotic promoter -10 and -35 regions. The best matches are summarized in Figure 47. It was not unexpected that

there were hexanucleotide sequences resembling the *E. coli* σ^{70} consensus promoter sequences (Hawley and McClure, 1983) at appropriate distances upstream of the *abg* mRNA 5' terminus, since the recombinant *abg* gene was found to be expressed in *E. coli* regardless of its orientation with respect to vector promoters (Wakarchuk, 1987; Wakarchuk *et al.*, 1988). The putative *abg* promoter region was also found to resemble the *nif* promoter regions as identified in *Klebsiella spp* (Kim *et al.*, 1986).

The transcript 3' end was determined for *abg* mRNA by hybrid protection analysis with a 3'-labeled DNA probe. The inverted repeats immediately downstream of the Abg stop (TGA) codon resembled a rho-independent termination signal as found for *E. coli* genes (Rosenberg and Court, 1979; Yanofsky, 1981) and unlike the *C. fimi* genes, preceded a run of consecutive T residues. The mapping of the *abg* mRNA 3' terminus in this region suggests that the inverted repeat sequences may function as termination signals in this *Agrobacterium* sp. Therefore, the initiation and termination sites for *abg* mRNA were found to be about 1,480 bases apart, a finding in good agreement with the size observed for *abg* mRNA in the northern blot experiment.

5. REFERENCES

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

2. Bates, N. 1987. Characterization of *cbg*: a cloned gene encoding an extracellular β-glucosidase from *Cellulomonas fimi*. M.Sc. Thesis, University of British Columbia, Vancouver, Canada.

3. Beguin, P., H. Eisen, and A W. Roupas. 1977. Free and cellulase-bound cellulases in a *Cellulomonas* species. J. Gen. Microbiol. **101**:191-196.

4. Beguin, P., M. Rocancourt, M.-C. Chebrou, and J.-P. Aubert. 1986. Mapping of mRNA encoding endoglucanaseA from *Clostridium thermocellum*. Mol. Gen. Gent. **202**:251-254.

5. Beguin, P., N.R. Gilkes, D.G. Kilburn, R.C. Miller, Jr., G.P. O'Neill, R.A.J. Warren. 1987. Cloning of cellulase genes. CRC Critical Rev. Biotechnology. **6**:129-162. CRC Press, Inc.

6. Bergey, D.H., R.S. Breed, R.W. Hammer, F.-C. Harrison, and F.M. Huntoon. 1974. 629-631. In: R.E. Buchanan and N.E. Gibbons (ed.), Bergey's manual of determinative bacteriology, 8th ed. The Williams & Wilkins Co., Baltimore.

7. Berk, A.J., and P.A. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**:721-732.

8. Berk A.J., and P.A. Sharp. 1978. Structure of the adenovirus 2 early mRNAs. Cell **14**:695-711.

9. Bibb, M.J., M.J. Bibb, J.M. Ward, and S.N. Cohen. 1985.

Nucleotide sequences encoding and promoting expression of three antibiotic resistance genes indigenous to *Streptomyces*. Mol. Gen. Genet. **199**:26-36.

10. Bibb,, M.J., G.R. Janssen, and J.M. Ward. 1985. Cloning and analysis of the erythromycin resistance gene (*erm*E) of *Streptomyces erythraeus*. Gene **38**:215-226.

11. Birnboim,H.C., and J.Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513-1523.

12. Bolivar,F.R., R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heyneker, A.W. Boyer, J.H. Crosa, and S. Falkow. 1977. Construction and characterization of new cloning vehicles.II. A multipurpose cloning system. Gene **2**:95-113.

13. Coughlan, M.P. 1985. The properties of fungal and bacterial cellulases with comment on their production and application. Biotech. Genetic Engineering Rev. **3**:39-109.

14. Day,A.G., and S.G. Withers. 1986. The purification and characterization of a β-glucosidase from *Alcaligenes faecalis*. Can. J. Biochem. Cell Biol. **64**:914-922.

15. Duckworth, H.E. and E.A. Thompson. 1982. Eds., Proc. Intl. Symp. on Ethanol from Biomass, The Royal Society of Canada, Winnepeg, Canada.

16. Duong, T.-V.C., E.A. Johnson, and A.L. Demain. 1983. Thermophilic, anaerobic and cellulolytic bacteria. Top. Enzyme Ferment. Biotechnol., **7**:156.

17. Ehrenberg, L., I. Fedorcsak, and F. Solymosy. 1976. Diethylpyrocarbonate in nucleic acids research. J. Mol. Biol. **16**:189-262.

18. Favoloro,J., R. Triesman, and R. Kamen. 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718-749.

19. Enari, T-M., and M-L. Niku-Paavola. 1987. Enzymatic hydrolysis of cellulose: Is the current theory of the mechanisms of hydrolysis valid? CRC Critical Rev. Biotechnol. CRC Press, Inc.

20. Eveleigh, D.E. 1983. The fermentation of biomass. p.365-391 *In*: Biomass utilization. W.A. Cote (ed.) Plenum Press, N.Y.

21. Fan, L.T., Y.H. Lee and D.H. Beardmore. 1980. Major chemical and physical features of cellulosic materials as substrates for enzymatic hydrolysis. Adv. Biochem. Eng. **14**:101-117.

22. Fornwald, J.A., F.J. Schmidt, C.W. Adams, M. Rosenberg, and M.E. Brawner. 1987. Two promoters, one inducible and one constitutive, control transcription of the *Streptomyces lividans* galactose operon. Proc. Natl. Acad. Sci. USA **84**:2130-2134.

23. Garcia-Martinez, D.V., A. Shinmyo, A. Madia, A.L. Demain. 1980. Studies on cellulase production by *Clostridium thermocellum*. Eur. J. Appl. Microbiol. Biotechnol. **9**:189.

24. Gardner, K.H. and J. Blackwell. 1974. The structure of native cellulose. Biopolymers **13**:1975-2001.

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

26. Gilkes, N.R., M.L. Langsford, D.G. Kilburn, R.C. Miller,

Jr., and R.A.J. Warren. 1984b. Mode of action and substrate specificitys of cloned bacterial genes. J. Biol. Chem. **259**:10455-10459.

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

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

29. Hall, D.O. 1983. Biomass for energy-fuels now and in the future. p. 1-22, *In*: Biomass utillization, W.A. Cote., ed. Plenum Press, N.Y.

30. Hammerstrom, R.A., K.D. Claus, J.W. Coghlan, and R.H. McBee. 1955. The constitutive nature of bacterial cellulases. Arch. Biochem. Biophys. **56**:123.

31. Hawley D.K., and W.R. McClure. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237-2255.

32. Han, Y.W., and V.R. Srinivasan. 1968. Isolation and characterization of a cellulose utilising bacterium. Appl. Microbiol. **16**:1140-1145.

33. Han, Y.W., and V.R. Srinivasan. 1969. The purification and characterization of β-glucosidase of *Alcaligenes faecalis*. J. Bacteriol. **100**:1355-1363.

34. Hopwood, D.A., M.J. Bibb, K.F. Chater, G.R. Janssen, F. Malpartida, and C.P. Smith. 1986. Regulastion of gene expression in antibiotic producing *Streptomyces*. Symp. Soc. Gen. Microbiol. **39**:251-276.

35. Jacob, F., and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. **3**:318-356.

36. Keddie, R.M. 1974. GenusIII. *Cellulomonas* Bergey et al. 1923, 154, emend. mut. char. Clark 1952, 50, p. 629-631. In: R.E. Buchanan and N.E. Gibbons (ed.), Bergey's manual of determinative bacteriology, 8th ed. The Williams & Wilkins Co., Baltimore.

37. Kennell, D., and I. Bicknell. 1973. Decay of messenger ribonucleic acid from the lactose operon of *Escherichia coli* as a function of growth temperature. J. Mol. Biol. **74**:21-31.

38. Kim, Y.-M., K.-J. Ahn, T. Beppu, and T. Vozumi. 1986. Nucleotide sequence of the *nifLA* operon of *Klebsiella oxytoca* NG13 and characterization of the gene products. Mol. Gen. Genet. **205**:253-259.

39. Kohchi, C., and A. Toh-e. 1985. Nucleotide sequence of *Candida pelliculosa* β-glucosidase gene. Nucleic Acids Res. **13**:6273-6282.

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

41. Lehninger, A.L. Sugars, storage polysaccharides and cell walls. p. 249-277. *In*: Biochemistry, Second Ed., Worth Pub. Inc., New York.

42. Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratory. Cold Spring Harbour, N.Y.

43. Maxam A.M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods

Enzymol. **65**:499-560.

44. Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20-79.

45. Miller, J.H. 1972. *In*: Experiments in molecular genetics. Cold Spring Harbor Laboratory, New York.

46. Miller, R.C. Jr., E.T. Young II, R.H. Epstein, H.M. Krisch, T. Mattson, and A. Bolle. 1981. Regulation of synthesis of the T4 DNA polymerase (gene 43). Virology **110**:98-112.

47. Moran, C.P., N. Lang, S.F.C. LeGrice, G. Lee, M. Stephens, A.L. Sonenshein, J. Pero, and R. Losick. 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. **186**:339-346.

48. Moss, B. 1981. 5' end labeling of RNA with capping and methylating enzymes. p.253-266. *In* J.G. Chirikjian and T.S.Papas (ed.), Gene amplification and analysis, vol. 2. Elsevier/North Holland Publishing Co., Amsterdam.

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

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

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

52. Owolabi, J.B., P. Beguin, D.G. Kilburn, R.C. Miller, Jr., and R.A.J. Warren. 1988. The structural gene for endoglucanase B of *Cellulomonas fimi* and its expression

in *Escherichia coli*. J. Appl. Environ. Microbiol.(in press).

53. Paradis, F.W., R.A.J. Warren, D.G. Kilburn and R.C. Miller, Jr. 1988. The expression of *Cellulomonas fimi* cellulase genes in *Brevibacterium lactofermentum*. Gene (in press).

54. Postma, P.W. 1986. Catabolite repression and related processes. Symp. Soc. Gen. Microbiol. **39**:319-353.

55. Rees, D.A., Morris, E.R., Thom,D., and J.K. Madden. 1982. Shapes and interactions of carbohydrate chains. p.195-290 *In*: Polysaccharides, vol.1. Academic Press, N.Y.

56. Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. **13**:319-353.

57. Ryu, D.D.Y., and M. Mandels. 1980. Cellulases: biosynthesis and applications. Enzyme Microb. Technol. **2**:91-101.

58. Shewale, J.G. 1982. β-glucosidase: Its role in cellulase synthesis and hydrolysis of cellulose. Eur. J. Biochem. **14**:435-443.

59. Shine,J. and L. Dalgarno. 1974. The 3' terminal sequence of *Escherichia coli* 16S ribosomal RNA: complimentary to nonsense triplets and ribosome binding sites. Proc. Natl.Acad. Sci. USA **71**:1342-1346.

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

61. Stackebrandt, E., and O. Kandler. 1979. Taxonomy of the genus *Cellulomonas*, based on phenotypic characters and

deoxyribonucleic acid -deoxyribonucleic acid homology, and proposal of seven neotype strains. Int. J. Syst. Bacteriol. **29**:273-282.

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

63. Thayer, D.W., S.V. Lowther, and J.G. Phillips. 1984. Cellulolytic activities of strains of the genus *Cellulomonas*. Int.J. Syst. Bacteriol. **34**:432-438.

64. Thomas, R.J. 1983. Wood anatomy and permeability. p. 1-32. *In*: Wood and Agricultural Residues: Research on Uses for Feed, Fuels and Chemicals, J. Slotes (ed.), Academic press, New York.

65. van Wezenbeek, P.M.G.F., T.J.M. Hulsebos, and J.G.G. Shoenmakers. 1980. Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. Gene **11**:129-148.

66. Viera, J., and J. Messing. 1982. The pUC plasmids and M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259-268.

67. von Gabain, A., J.G. Belasco, J.L. Schottel, C.Y. Chang, and S.N. Cohen. 1983. Decay of mRNA in *Escherichia coli*: investigation of the fate of specific segments of transcripts. Proc. Natl. Acad. Sci USA **80**:653-657.

68. Wakarchuk, W.W. 1987. The molecular cloning and characterization of a β-glucosidase gene from an *Agrobacterium*. Ph.D. Thesis, University of British Columbia, Vancouver, Canada.

69. Wakarchuk, W.W., N.M. Greenberg, D.G. Kilburn, R.C. Miller, Jr., and R.A.J. Warren. 1988. Structure and transcription analysis of the gene encoding a cellobiase

from *Agrobacterium* sp. strain ATCC 21400. J. Bacteriol. **170**:000-000.

70. Wakarchuk, W.W., D.G. Kilburn, R.C. Miller, Jr., and R.A.J. Warren. 1984. The preliminary characteriszation of the β-glucosidases of *Cellulomonas fimi*. J. Gen. Microbiol. **130**:1385-1389.

71. Wakarchuk, W.W., D.G. Kilburn, R.C. Miller, Jr., and R.A.J. Warren. 1986. The molecular cloning and expression of a cellobiase gene from *Agrobacterium* in *Escherichia coli*. Mol. Gen. Genet. **205**:146-152.

72. Weaver, R.F., and C. Weissman. 1979. Mapping of RNA by a modification of the Berk and Sharp procedure: the 5' termini of 15S β-globin mRNA precurssor and mature 10S β-globin mRNA have identical map units. Nucleic Acids Res. **7**:1175-1193.

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

74. Wich, G., H. Hummel, M. Jarsch, U. Bar, and A. Bock. 1986. Transcription signals for stable RNA genes in *Methanococcus*. Nucleic Acids Res. **14**:2459-2479.

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

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

77. Yanisch-Perron, C., J. Viera, and J. Messing. 1985.

Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene **33**:103-119.

78. Yanofsky, C. 1981. Attenuation in the control of expression of bacterial operons. Nature (london) **289**:751-758.