MOLECULAR CLONING, CHARACTERIZATION AND EXPRESSION OF THE ENDOGLUCANASE C GENE OF CELLULOMONAS FIMI AND PROPERTIES OF THE NATIVE AND RECOMBINANT GENE PRODUCTS

by

BERNHARD MOSER

Diploma, Federal Institute of Technology, Zurich, Switzerland, 1981

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

October 1988

© Bernhard Moser, 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
Vancouver, Canada

Date **OCT. 14, 1988**
ABSTRACT

In addition to substrate-associated cellulases, *Cellulomonas fimi* secretes endoglucanases (endo-1,4-β-D-glucan glucanohydrolases, EC 3.2.1.4.) which are recovered from the cellulose-free culture supernatant of cells grown on microcrystalline cellulose. Two such enzymes, C3.1 and C3.2 with M_r's of 130'000 and 120'000, respectively, were purified to homogeneity. The two endoglucanases were shown to share the same N-terminal amino acid sequence and to hydrolyze carboxymethylcellulose (CMC) with similar efficiencies (236u/mg protein for C3.1 and 367u/mg protein for C3.2).

The recombinant lambda vector L47.1-169 was identified from a C.fimi DNA-lambda library on the basis of hybridization with C3.1/2-specific oligonucleotide probes. The subclone pTZ18R-8 only moderately expressed CMCase activity. The 5'-terminus of cenC (the gene coding for C3.1/2) was localized in the insert by Southern transfer experiments and nucleotide sequence analysis. Results from total C.fimi RNA-DNA hybrid protection analyses defined the boundaries of cenC in pTZ18R-8 and led to the tentative identification of -10 and -35 promoter sequences.

To improve the expression of cenC, its entire coding sequence, except for the start codon GTG, was fused in frame to the ATG codon of a synthetic ribosomal binding site (PTIS) and placed under the transcriptional control of the lac p/o system. Induction of the resulting clone (JM101[pTZP-cenC]) led to impaired growth in liquid cultures because overproduction of CenC inhibited cell division and eventually led to cell death. Analysis of cell fractions by SDS-PAGE revealed a dominant (>10% of total cell extract proteins), clone-specific protein with a M_r of approximately 140'000 which was found exclusively in the
cytoplasmic fraction. Conversely, 60% of the total CMC-hydrolyzing activity was localized in the periplasmic fraction indicating that the export of CenC is required for maximal expression of endoglucanase activity.

Isolation of the cellulolytic activities from an osmotic shock led to the purification to homogeneity of two recombinant cellulases, CenC1 and CenC2, with \( M_r \) of 130'000 and 120'000, respectively. Both enzymes hydrolyzed CMC with similar efficiencies (278u/mg protein for CenC1 and 390u/mg protein for CenC2). In addition, amino acid sequence analyses showed the two enzymes to have the same N-termini as the native enzymes and proved furthermore that the CenC leader peptide was functional in *Escherichia coli*. 
TABLE OF CONTENTS

ABSTRACT
TABLE OF CONTENTS
LIST OF TABLES
LIST OF FIGURES
LIST OF ABBREVIATIONS
ACKNOWLEDGEMENTS

I. INTRODUCTION
   1. Background
   2. The cellulolytic system secreted by Cellulomonas fimi
      2.1. Biochemical characterization of the cellulolytic components of C.fimi
      2.2. Characterization of cloned C.fimi cellulase genes and their recombinant products
      2.2.1. Molecular cloning of C.fimi cellulase genes
      2.2.2. Subcloning and expression of cex
      2.2.3. Subcloning and expression of cenA
      2.2.4. Comparative biochemical analyses of Cex and CenA, and of their genes
      2.2.5. Subcloning and expression of cenB
      2.2.6. Identification of an E.coli clone expressing β-glucosidase activity
   3. Objectives for the research of this thesis

II. MATERIALS AND METHODS
   1. Bacterial strains, plasmids and phage vectors
   2. Biochemicals
   3. Preparation of rabbit serum specific for C3.2
   4. Protein biochemistry
| 4.1. | Preparation of Affigel-10-RαC3.2 affinity column | 22 |
| 4.2. | Determination of protein concentrations | 22 |
| 4.3. | Polyacrylamide gel electrophoresis (SDS-PAGE) | 24 |
| 4.4. | Western blot analysis | 24 |
| 4.5. | Amino acid sequence analysis and amino acid composition determination | 24 |
| 4.6. | Fractionation of bacterial cells | 25 |
| 5. | Enzyme assays | 25 |
| 5.1. | DNS-CMCase assay | 25 |
| 5.2. | pNPCase assay | 26 |
| 5.3. | β-lactamase assay | 26 |
| 5.4. | Glucose-6-P-DHase assay | 27 |
| 5.5. | Congo red plate assay | 27 |
| 6. | Electron microscopy | 28 |
| 7. | DNA and RNA methodology | 28 |
| 7.1. | Isolation of C.fimi DNA | 28 |
| 7.2. | Isolation of plasmid and bacteriophage DNA | 29 |
| 7.3. | Preparation of template DNA for sequencing | 29 |
| 7.4. | Construction of plasmid deletions for sequencing | 30 |
| 7.5. | DNA sequencing | 35 |
| 7.6. | Construction of the C.fimi DNA-lambda library | 35 |
| 7.7. | Synthesis and purification of oligonucleotides | 36 |
| 7.8. | Labeling of DNA with $^{32}$P | 36 |
| 7.9. | Screening methods for recombinant bacteriophage and plasmid vectors | 37 |
| 7.10. | DNA and RNA dot-blot analysis | 38 |
| 7.11. | Southern transfer analysis | 38 |
| 7.12. | cDNA synthesis by primer extension | 39 |
| 7.13. | Isolation of C.fimi RNA | 39 |
| 7.14. | RNA-DNA hybrid protection analysis | 40 |
III. RESULTS AND DISCUSSION

1. Isolation and characterization of endoglucanases from *C.fimi* culture supernatant
   1.1. Purification of C3.1 and C3.2
   1.2. Characterization of C3.1 and C3.2
2. Molecular cloning of cenC
   2.1. Preliminary experiments
   2.2. Cloning strategy
   2.3. Subcloning and initial characterization of cenC
   2.4. Sequence analysis of the 5'-end of cenC
3. Analyses of *in vivo* cenC transcripts of *C.fimi*
   3.1. Mapping of the 5'-ends of cenC transcripts
   3.2. Mapping of the 3'-ends of cenC transcripts
4. Overproduction of CenC in *E.coli*
   4.1. Construction of pTZP-cenC
   4.2. Characterization of the expression of cenC in JM101[pTZP-cenC]
5. Initial characterization of purified, recombinant CenC
   5.1. Purification of recombinant CenC1 and CenC2
   5.2. Amino acid sequence analysis of CenC1 and CenC2
6. Final remarks

IV. REFERENCES

V. APPENDIX

Flow-chart protocol for the molecular cloning of cenC and the construction of the cenC high-expression vector, pTZP-cenC
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Comparison of values for the kinetic parameters of <em>C.fimi</em> cellulases</td>
<td>6</td>
</tr>
<tr>
<td>II. Cloned cellulase genes of <em>C.fimi</em> and their products</td>
<td>8</td>
</tr>
<tr>
<td>III. Flow-chart of the purification of C3.1 and C3.2</td>
<td>44</td>
</tr>
<tr>
<td>IV. Amino-terminal amino acid sequence data of C3.1, C3.2 and the C3.2-internal tryptic peptide T-115</td>
<td>49</td>
</tr>
<tr>
<td>V. Partial amino acid composition analysis of C3.2</td>
<td>50</td>
</tr>
<tr>
<td>VI. Cross-screen of recombinant L47.1 clones</td>
<td>59</td>
</tr>
<tr>
<td>VII. Endoglucanase activities of various cenC clones</td>
<td>96</td>
</tr>
<tr>
<td>VIII. Effect of IPTG on growth and viability of Jm101[pTZP-cenC]</td>
<td>97</td>
</tr>
<tr>
<td>IX. Localization of CenC activity in Jm101[pTZP-cenC]</td>
<td>102</td>
</tr>
<tr>
<td>X. Flow-chart of the purification of CenC1 and CenC2</td>
<td>105</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Proposed bifunctional structure of CenA and Cex</td>
<td>16</td>
</tr>
<tr>
<td>2.</td>
<td>ELISA of rabbit B3 anti C3.2 serum</td>
<td>23</td>
</tr>
<tr>
<td>3.</td>
<td>Purification of template DNA by alkaline sucrose gradient centrifugation</td>
<td>31</td>
</tr>
<tr>
<td>4.</td>
<td>Generation of deletions following a modification of the Dale procedure</td>
<td>33</td>
</tr>
<tr>
<td>5.</td>
<td>Purification scheme for C3.1 and C3.2</td>
<td>42</td>
</tr>
<tr>
<td>6.</td>
<td>SDS-PAGE analysis of C3.1 and C3.2</td>
<td>46</td>
</tr>
<tr>
<td>7.</td>
<td>Western blot analysis of purified C3.1 and C3.2</td>
<td>48</td>
</tr>
<tr>
<td>8.</td>
<td>Partial nucleotide sequence of pUC13-1/43</td>
<td>53</td>
</tr>
<tr>
<td>9.</td>
<td>Oligonucleotide screening probes for the lambda L47.1-C.fimi DNA library</td>
<td>56</td>
</tr>
<tr>
<td>10.</td>
<td>Autoradiogram of a plaque-filter lift hybridization experiment</td>
<td>58</td>
</tr>
<tr>
<td>11.</td>
<td>Diagram of the recombinant clone lambda L47.1-I69</td>
<td>60</td>
</tr>
<tr>
<td>12.</td>
<td>DNA dot-blot hybridization experiment</td>
<td>62</td>
</tr>
<tr>
<td>13.</td>
<td>Nucleotide sequence analysis of the plasmids M13mp19-B4 and M13mp19-B5</td>
<td>63</td>
</tr>
<tr>
<td>14.</td>
<td>Diagram of the Genescribe vector system pTZ18/19-R/U</td>
<td>65</td>
</tr>
</tbody>
</table>
Figure | PAGE
--- | ---
15. Restriction enzyme map of the recombinant DNA in pTZ18R-8 | 67
16. Southern transfer analyses of pTZ18R-8 DNA | 69
17. Sequencing strategy for the insert in pTZ18R-8/5-5 | 70
18. Effect of using 7d-dGTP during sequencing C.fimi DNA | 72
19. Nucleotide sequence of the 5'-region of cenC | 73
20. Dot-blot analysis of total C.fimi RNA | 75
21. Preparation of the probe for the 5'-cenC transcript mapping | 76
22. S1 nuclease protection analysis of 5'-ends of cenC transcripts | 77
23. Fine-mapping of 5'-ends of cenC transcripts | 79
24. 5'-region of cenC showing the proposed transcription start sites | 80
25. Scheme used for the preparation of the probe for the 3'-cenC transcript mapping | 82
26. Mapping of 3'-ends of C.fimi transcripts | 85
27. Nucleotide sequence at the 3'-end of cenC | 86
28. Construction of pTZP-cenC | 89
29. Construction of PTIS-cenC fusion by primer extension | 91
30. Congo red plate assay | 94
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>Growth and activity profiles of JM101[pTZP-cenC]</td>
<td>98</td>
</tr>
<tr>
<td>32</td>
<td>Electron micrographs of JM101[pTZP-cenC] cells</td>
<td>101</td>
</tr>
<tr>
<td>33</td>
<td>SDS-PAGE analysis of protein samples from different compartments of JM101[pTZP-cenC]</td>
<td>104</td>
</tr>
<tr>
<td>34</td>
<td>SDS-PAGE analysis of CenC1 and CenC2</td>
<td>107</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

a.a. amino acid(s)
Amp ampicillin
bp base pair(s)
cbg C.fimi gene encoding Cbg
Cbg C.fimi β-glucosidase
cenA(B,C) C.fimi gene encoding CenA(B,C)
CenA(B,C) C.fimi endoglucanase A(B,C)
cex C.fimi gene encoding Cex
Cex C.fimi exoglucanase
CMC carboxymethylcellulose
Con A concanavalin A
DNS dinitrosalicylic acid
dNTP deoxynucleotide triphosphate
IPTG isopropyl-β-D-thiogalactoside
kb kilo base pair(s)
kDa kilo dalton(s)
lac lactose operon
lac p/o lac promoter-operator
LB Luria broth
λ bacteriophage lambda
PAGE polyacrylamide gel electrophoresis
pfu plaque forming unit(s)
PL leftward promoter of lambda
pNPC p-nitrophenolcellobioside
PTIS portable translation initiation site
RBS ribosomal binding site
SDS sodium dodecyl sulfate
ACKNOWLEDGEMENTS

I wish to thank Drs. D.G. Kilburn, R.C. Miller, Jr. and R.A.J. Warren for their guidance and support during discussions, benchwork and completion of this thesis. I am indebted to Dr. K. Humphrey for his advice and helpful suggestions during advisory committee meetings and proof reading of this thesis. I gratefully acknowledge Dr. N.R. Gilkes for the constructive criticism and readiness to help whenever I knocked at his door. I wish to thank Ms. Kielland, Dr. R. Olefson and Dr. D. McKay for performing amino acid sequence analyses and amino acid composition determination, Drs. R. Bradley and D.G. Scraba for the electron microscopy work, and Dr. T. Atkinson for synthesizing the oligonucleotides. I am grateful to E. Kwan for lending me her skillful hand in purifying the native endoglucanases. I especially wish to thank the ladies and gentlemen as well as my fellow graduate students (don't give up, pretty soon you'll belong to the former class!) of the cellulase group for their support and all the good times we shared. To F. Rolli I wish to express my gratitude for her understanding and sacrifices during the predawn period of this work. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

I dedicate this thesis to Franziska and to my parents, Hanni and Hans Moser.
I. INTRODUCTION

I.1. Background

As a main photosynthetic product the β-glucan, cellulose, is the most abundant, renewable polycarbon on earth. Not lastly due to its chemical and physical properties, cellulose is only slowly degraded by a wide variety of microorganisms. Among the most extensively studied organisms producing cellulolytic enzymes are the fungal species *Trichoderma* and *Sporotrichum*, actinomycetes of the species *Streptomyces*, and the bacterial species *Cellulomonas*, *Erwinia*, *Clostridium* and *Bacillus*. It is not the purpose of this introduction to elaborate on the properties of their cellulolytic systems and to carry out a detailed comparative analysis. Instead, I will refer the reader to numerous excellent review articles (Beguin *et al.*, 1987, Coughlan, 1985, Bisaria and Ghose, 1981, Mandels, 1983, and Eriksson and Wood, 1985).

A general model for the degradation of cellulose by cellulolytic enzymes emerged mainly from studies of the fungi *Trichoderma* sp. and *Sporotrichum pulverulentum*. In a first step, the β-1,4-glycosidic bonds in amorphous regions of cellulose are hydrolyzed by endoglucanases (endo-1,4-β-D-glucan glucanohydrolase, EC 3.2.1.4). There seems to be some controversy as to whether substrate bound endoglucanases are capable of disrupting the crystalline structure of cellulose to form microamorphous regions prior to their enzymatic action (Klyosov *et al.*, 1986, Ryu *et al.*, 1984) or whether additional factors such as hydrogen bondase (Enari and Niku-Paavola, 1987) or other protein components like C1 (Montenecourt, 1983) are the sole requirements for the "amorphogenesis" of cellulose. The newly created ends of cellulose macromolecules are subsequently attacked at their nonreducing ends by cellobiohydrolases (1,4-β-D-glucan cellobiohydrolase, EC
3.2.1.91, also referred to as exoglucanases) to produce cellobiose. Finally, small soluble celldextrins including cellobiose are degraded to glucose by cellobiases or generally β-glucosidases (EC 3.2.1.21). A consensus is emerging with respect to the cellular location of cellobiases: fungi secrete the enzyme to hydrolyze cellobiose extracellularly whereas bacteria convert the substrate to glucose by cell associated cellobiases.

The biosynthesis of cellulases in virtually all microorganisms examined to date is subject to regulation by the carbon source. Cellulose in the growth medium has an inducing effect whereas glucose or readily metabolized sugars repress the expression of cellulases (Coughlan, 1985). However, the structure of the actual inducer which is generally effective in a variety of cellulolytic systems is still elusive to date. It is widely accepted that in the presence of cellulose, low levels of constitutively expressed endo- and exoglucanases produce soluble oligosaccharides which act as inducers or are subsequently converted to inducer molecules perhaps by β-glucosidases or glycosyltransferases. In the fungal species *Trichoderma reesei* (Sternberg and Mandels, 1979) and *Trichoderma viride* (Eriksson and Hamp, 1978) such an inducer molecule has been identified as sophorose (2-O-β-glucopyranosyl-D-glucose), which exhibited a dual role in the regulation of cellulose expression (Sternberg and Mandels, 1980). At low concentrations (0.5μM) the production of β-glucosidases is repressed, thereby decreasing the rate of sophorose hydrolysis. This leads to the accumulation of the inducer (approx. 300-fold increase in concentration) to a concentration high enough for the induction of endo- and exoglucanase expression. Cellobiose (and lactose) is not considered to be the natural inducer since its inducing activity
is only observed at unphysiologically high concentrations at which exoglucanases are inhibited (Coughlan, 1985).

It is a common characteristic of most cellulolytic systems to express multiple forms of individual cellulase activities. We are only beginning to understand the reason for the complex nature of such cellulose degrading apparatus. Certainly, the stereochemical and physical properties of cellulose require an array of cellulases with different activities and fine-specificities. It is not surprising then that a variety of genes contribute to the diversity of cellulolytic activities. However, electrophoretic and amino acid sequencing analysis as well as immuno-crossreactivity determinations produced a picture far more complex than expected (Coughlan, 1985). It is certain that glycosylation of eukaryotic cellulases contributes to the variety of chromatographically distinct cellulolytic components (Teeri, 1987). Yet, the physiological function of this post-translational modification is still speculated upon. It has been suggested that glycosylation is required for secretion of fungal cellulases (Sheir-Neiss and Montenecourt, 1984). In addition, it may confer thermal and pH stability as well as limited protection against the action of secreted proteases. It has been observed in *Trichoderma reesei* that the concentration and diversity of cellulase components change with increasing age of the culture (Gong and Tsao, 1979). The release of smaller, active fragments from purified cellulases by protease preparations from the same organism emphasizes the role of proteolysis in generating the multiplicity of cellulolytic activities in a given system (Langsford et al., 1987, Arfman et al., 1987). Proteases may play an important role *in vivo* by allowing reutilization of extracellular enzymes by an organism growing under nitrogen limitation.
I.2. The cellulolytic system secreted by *Cellulomonas fimi*.

I.2.1. Biochemical characterization of the cellulolytic components of *C.fimi*

The cellulase system of *C.fimi* appears to be complex (Langsford et al., 1984). Supernatants of cultures of this Gram-positive, coryneform, mesophilic bacterium grown on a low concentration of Avicel (microcrystalline cellulose; 0.1%) contain up to 10 distinguishable cellulase activities as determined by their ability to hydrolyze the soluble substrate carboxymethylcellulose (CMC). As in fungal systems (Coughlan, 1985), and a few cellulolytic bacterial species such as *Clostridium* (Ng and Zeikus, 1981, Creuzet and Frixon, 1983), *Pseudomonas* (Yamane et al., 1970), and *Thermomonospora* (Calza et al., 1985), some of the cellulolytic components secreted by *C.fimi* are glycosylated (Langsford et al., 1984).

Initial studies focused on the biochemical characterization of cellulases which were tightly bound to the insoluble substrate Avicel. The two main components were recovered from Avicel by elution with 6M guanidine hydrochloride and further purified to homogeneity by Concanavalin A-sepharose and MonoQ anion-exchange column chromatography. One enzyme was an exoglucanase (gCex; the prefix g indicates that the protein is glycosylated) with a Mr of 55'000. Its isoelectric point (pI) was shown to be 5.8, and the enzyme was glycosylated exclusively with mannose (8% by weight; Langsford, 1988). The specific activity of purified gCex on the substrate p-nitrophenolcellobioside (pNPC) was 9.3 units per mg of protein (1unit = 1μmole p-nitrophenol released per minute), and on CMC it was 85 units per mg of protein (1unit = 1μmole reducing sugar produced per minute). The values for the kinetic
parameters, Km and Vmax, using the substrates CMC and pNPC for the exoglucanase (as well as for other native and recombinant C.fimi cellulases, see below) are listed in Table I.

The second enzyme was an endoglucanase (gCenA) with the Mr of 57'000, a pI of 8.2 and containing approx. 10% mannose by weight (Langsford, 1988). As is the case for true endoglucanases, the activity of gCenA on aryl-celllobioside substrates was marginal whereas CMC was efficiently hydrolyzed, as the specific activity of 370 units per mg of protein indicates.

Not all of the cellulolytic components secreted by C.fimi are associated with the insoluble substrate Avicel. Quite an array of cellulase activities arises during the growth of the cells on low concentration of Avicel, which remain free in the culture supernatant. The range of molecular weights of these enzymes suggests that not all of them are the result of differential glycosylation or proteolytic cleavage of a restricted repertoire of enzymes but instead are encoded by separate genes other than cex or cenA. One would expect that an efficient cellulolytic system contains, in addition to substrate-bound cellulases, soluble enzymes with high substrate specificities for soluble intermediates generated during cellulose hydrolysis. Such endoglucanases are currently being investigated (see this thesis).

C.fimi also synthesizes β-glucosidases. At least two are found in whole cell extracts as determined by non-denaturing PAGE followed by assaying aryl-glucosidase (pNPGase) and cellobiase activity (Wakarchuck et al., 1984). The expression of the pNPGase activity was not affected significantly by the carbon source used for growth of the cells. On the other hand, the faster migrating cellobiase was markedly growth substrate dependent, being maximally expressed in the presence of cellobiose or Avicel. These two distinct β-glucosidases have not been isolated and
TABLE I. Comparison of values for the kinetic parameters of *C.fimi* cellulases. The values for the kinetic parameters are based on hydrolysis of CMC, pNPG and pNPC by glycosylated, native or non-glycosylated, recombinant cellulases. Values for Kms are listed as mg/ml for CMC and mM for pNPG and pNPC; Vmaxs are expressed as μmoles products released per minute, per mg protein.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Enzyme</th>
<th>Km</th>
<th>Vmax</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMC</td>
<td>gCex</td>
<td>3.18±/-0.21</td>
<td>42.9±/-1.5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ngCex</td>
<td>3.04±/-0.23</td>
<td>35.8±/-1.9</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>gCenA</td>
<td>0.19±/-0.1</td>
<td>62.5±/-1.1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ngCenA</td>
<td>0.17±/-0.1</td>
<td>56.6±/-1.0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ngCenB</td>
<td>0.51±/-0.05</td>
<td>26.1±/-1.1</td>
<td>2</td>
</tr>
<tr>
<td>pNPC</td>
<td>gCex</td>
<td>0.64±/-0.03</td>
<td>9.3±/-0.2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ngCex</td>
<td>0.70±/-0.02</td>
<td>11.4±/-0.1</td>
<td>1</td>
</tr>
<tr>
<td>pNPG</td>
<td>ngCbg</td>
<td>~ 0.13</td>
<td>~ 1063</td>
<td>3</td>
</tr>
</tbody>
</table>

a The prefix *g* indicates the glycosylated form of the enzyme obtained from *C.fimi* cultures; *ng* stands for the non-glycosylated, recombinant form of the enzyme.

b The numbers refer to the references cited: 1, Langdford et al., 1987; 2, Owolabi, 1988; 3, Paradis, personal communication.
purified yet.

Cellobiose phosphorylases (EC 2.4.1.20) of organisms which do not secrete β-glucosidases are thought to play an important role in the metabolism of cellobiose. Preliminary studies of a cellobiose phosphorylase activity in Cellulomonas sp. DSM 20108 indicate that this enzyme is located in the cytosol (Schimz and Decker, 1985). Maximal activity was recovered when the cells were grown with cellobiose as sole carbon source. However, no biochemical data are available to date for this enzyme. C.fimi has not been assayed for this activity. In fact, virtually nothing is known about the metabolism of cellobiose in C.fimi.

In addition to the expression of cellulases, C.fimi secretes proteases. The major proteolytic activity is inhibited by phenylmethylsulfonyl fluoride (PMSF), suggesting it to be a serine protease (Langsford et al., 1984, Langsford, 1988). Quantitative measurements of protease production in supernatants of cultures grown with different carbon sources show that protease synthesis is regulated independently of cellulose synthesis by catabolite repression during growth on glucose. The C.fimi protease preparations cleave purified gCenA and gCex into smaller fragments without affecting the total cellulolytic activity (Langsford et al., 1987, Langsford, 1988), suggesting strongly that some of the small molecular weight cellulolytic activities observed in older C.fimi cultures are derived by proteolysis of bigger cellulases.
TABLE II. Cloned cellulase genes of \textit{C.fimi} and their products.

Letters in superscript: \(a\), specific activities in units per sample volume or mg proteins; \(b\), estimated values for the molecular weights of purified enzymes based on SDS-PAGE.

Letters in superscript behind numbers for the specific activities indicate what the particular values are referred to: \(\text{ve}\), ml of cell extract; \(\text{vs}\), ml of culture supernatant; \(\text{vl}\), ml of cell lysate; \(\text{p}\), mg of cell extract proteins; \(\text{ps}\), mg of culture supernatant proteins. One unit equals one \(\mu\)mole of products (reducing sugar, p-nitrophenol) released per minute.

Arrows (\(\leftrightarrow\)) indicate the constructs from which the respective recombinant enzymes were purified.

Numbers in brackets refer to references cited: (1), Whittle et al., 1982; (2), Gilkes et al., 1984a; (3), O'Neill et al., 1986b; (4), Curry et al., 1988; (5), O'Neill, 1986; (6), Langsford, 1988; (7), Wong et al., 1986a; (8), Wong, 1986; (9), Skipper et al., 1985; (10), Johnson et al., 1986; (11), Guo et al., 1988; (12), Owolabi et al., 1988a; (13), Owolabi, 1988; (14), Moser, this thesis; (15), Paradis, personal communication; (16), Gilkes et al., 1988.
TABLE II. Cloned cellulase genes of *C. fimi* and their products

<table>
<thead>
<tr>
<th>gene</th>
<th>construct (vector)</th>
<th>host</th>
<th>promoter</th>
<th>substrate</th>
<th>spec. act$^a$</th>
<th>M$_r$$^b$</th>
<th>spec. act$^c$</th>
<th>M$_r$</th>
<th>spec. act$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>cex</em></td>
<td>pDW1 (pBR322)</td>
<td><em>E. coli</em> C600</td>
<td>tet ?</td>
<td>CMC</td>
<td>0.027$^{ve}$</td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pEC1 (pBR322)</td>
<td>&quot;</td>
<td>tet ?</td>
<td>CMC</td>
<td>0.147$^P$</td>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pUC12-1.1-PTIS</td>
<td>&quot;</td>
<td>lac</td>
<td>pNPC</td>
<td>0.146$^P$</td>
<td>(3)</td>
<td>&lt;-&gt; <em>ngCex</em> (5,16)</td>
<td>47'300</td>
<td>11.4 (pNPC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pCP3cex (pcI857)</td>
<td>&quot;</td>
<td>$\lambda$ P$_L$</td>
<td>pNPC</td>
<td>0.042$^P$</td>
<td>(3)</td>
<td>&lt;-&gt; 59'000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pMV3.cex</td>
<td><em>S. cerevisiae</em></td>
<td>MEL1</td>
<td>pNPC</td>
<td>3.0$^P$</td>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>cenA</em></td>
<td>pEC2 (pBR322)</td>
<td><em>E. coli</em> C600</td>
<td>tet</td>
<td>CMC</td>
<td>0.217$^P$</td>
<td>(7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pcEC2 (pUC)</td>
<td><em>E. coli</em> JM101 (F')</td>
<td>lac ?</td>
<td>CMC</td>
<td>0.015$^P$</td>
<td>(7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ speact = specific activity; $^b$ M$_r$ = molecular weight; $^c$ speact% = specific activity%
TABLE II. continued

<table>
<thead>
<tr>
<th>clone (vector)</th>
<th>host</th>
<th>promoter</th>
<th>substrate</th>
<th>spec. act(^a)</th>
<th>(M_r) (^b)</th>
<th>spec. act(^\gamma)</th>
<th>(M_r)</th>
<th>spec. act(^\gamma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pK2.4 (pOP)</td>
<td>S. cerevisiae</td>
<td>ADC1</td>
<td>CMC</td>
<td>1.60(^{vs}) (9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pREC2.2 (pJAJ103)</td>
<td>R. capsulatus</td>
<td>rxcA</td>
<td>CMC</td>
<td>8.75(^{P}) (10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUC18-1.6cenA</td>
<td>E.coli</td>
<td>lac</td>
<td>CMC</td>
<td>1.10(^{ve}) (11) &lt;-&gt; ngCenA (8,16)</td>
<td>48'700</td>
<td>4.13 (CMC)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**cenB**

<table>
<thead>
<tr>
<th>clone (vector)</th>
<th>host</th>
<th>promoter</th>
<th>substrate</th>
<th>spec. act(^a)</th>
<th>(M_r) (^b)</th>
<th>spec. act(^\gamma)</th>
<th>(M_r)</th>
<th>spec. act(^\gamma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pEC3 (pBR322)</td>
<td>E.coli</td>
<td>tet ?</td>
<td>CMC</td>
<td>0.007(^{P}) (12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C600</td>
<td></td>
<td></td>
<td>0.027(^{P}) (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pJB301 (pUC)</td>
<td>E.coli</td>
<td>lac</td>
<td>CMC</td>
<td>0.147(^{P}) (12) &lt;-&gt; ngCenB (13)</td>
<td>110'000</td>
<td>13.57 (CMC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pJB303 (pUC)</td>
<td>&quot;</td>
<td>lac</td>
<td>CMC</td>
<td>0.157(^{P}) (12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**cenC**

<table>
<thead>
<tr>
<th>clone (vector)</th>
<th>host</th>
<th>promoter</th>
<th>substrate</th>
<th>spec. act(^a)</th>
<th>(M_r) (^b)</th>
<th>spec. act(^\gamma)</th>
<th>(M_r)</th>
<th>spec. act(^\gamma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L47.1-I69 (Lambda L47.1) NM538</td>
<td>E.coli</td>
<td>(\lambda P_L) ?</td>
<td>CMC</td>
<td>&lt;0.01(^{\downarrow})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) spec. act: specific activity
\(^b\) \(M_r\): molecular weight
\(^\gamma\) spec. act\(^\gamma\): specific activity corrected for impurities

\(^{vs}\): vs.
\(^{P}\): P value
\(^{ve}\): ve
<table>
<thead>
<tr>
<th>construct (vector)</th>
<th>host</th>
<th>promoter</th>
<th>substrate</th>
<th>spec. act. (^a)</th>
<th>(M_r^b)</th>
<th>spec. act. (^\gamma)</th>
<th>(M_r)</th>
<th>spec. act. (^\gamma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTZ18R-8</td>
<td>E.coli</td>
<td>lac ?</td>
<td>CMC</td>
<td>0.008P (14)</td>
<td></td>
<td>C3.2 (14)</td>
<td></td>
<td>120'000 367 (CMC)</td>
</tr>
<tr>
<td></td>
<td>JM101 (F')</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTZP-cenC</td>
<td>&quot;</td>
<td>lac</td>
<td>CMC</td>
<td>0.350P (14)&lt;&gt;CenCl(14)</td>
<td>130'000 278 (CMC)</td>
<td>CenCl (14)</td>
<td>120'000 390 (CMC)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(pBR322)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E.coli</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C600</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUC13-62Δ31</td>
<td>E.coli</td>
<td>lac ?</td>
<td>pNPG</td>
<td>120P (15)&lt;&gt;Cbg (15)</td>
<td>180'000 12'950P</td>
<td>Cbg (15)</td>
<td>120'000 390 (CMC)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>JM83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
I.2.2. Characterization of cloned *C. fimi* cellulase genes and their recombinant products

I.2.2.1. Molecular cloning of *C. fimi* cellulase genes

Whittle *et al.* were the first group to report the molecular cloning of a cellulase gene (Whittle *et al.*, 1982). Total genomic *C. fimi* DNA was cloned into pBR322, and the resulting *C. fimi* DNA library was subsequently screened with antibodies against proteins in *C. fimi* culture supernatants. The recombinant clone pDW1 expressed CMCase activity as determined by the DNS-CMCase assay (Table II). A subsequent screening by the immunoscreening procedure of a new *C. fimi* DNA library of DNA digested to completion with BamHI and cloned into pBR322 yielded 61 immunopositives, of which 38 exhibited cellulase activity (Gilkes *et al.*, 1984b). Based on antigenicity and cellulase activity, further characterization of some of these positives revealed clones determining three distinct cellulases. The plasmids in these clones were referred to as pEC1, pEC2 and pEC3 (Table II). pEC1 carried a subfragment of the insert in pDW1. Viscometric and colorimetric analyses of CMC hydrolysis showed that the enzymes encoded by pEC2 and pEC3 behaved as endoglucanases, whereas the cellulase encoded by pEC1 showed exoglucanase activity (Gilkes *et al.*, 1984a, Gilkes *et al.*, 1984b). Sequencing of the genes in pEC1 and pEC2 demonstrated that pEC1 encoded the recombinant counterpart (ngCex; prefix ng stands for the non-glycosylated form of Cex) of the native exoglucanase, gCex (O'Neill *et al.*, 1986a), and that the gene encoding the recombinant analogue (ngCenA) of the native endoglucanase, gCenA, was contained partially on pEC2 (Gilkes *et al.*, 1984a, Wong *et al.*, 1986a).
1.2.2.2. Subcloning and expression of cex

The gene cex was localized on a 2.58kb DNA fragment in pEC1 with a coding region of 1452bp (484 codons). The coding region for the N-terminal amino acid sequence of the mature Cex is preceded by a sequence encoding a leader peptide of 41 amino acids (O'Neill et al., 1986a). The amino acid composition predicts a Mr of 47'100 (Gilkes, et al., 1988) which agrees well with the estimated molecular mass of the purified exoglucanase on SDS-PAGE (47.3kDa). Maximal expression of cex in E.coli was achieved by replacing the natural C.fimi transcriptional and translational control signals with the leftward promoter of phage lambda and a synthetic ribosomal binding site (PTIS). The resulting construct, pCP3cex (Table II), led to the overproduction of ngCex exceeding 20% of total cellular proteins (O'Neill et al., 1986b), resulting in the formation of insoluble, cytosolic ngCex aggregates.

Optimal expression of cex with concomitant export into the periplasm of E.coli was achieved with the construct pUC12-1.1(PTIS). The processed, recombinant exoglucanase was found in the periplasmic fraction of the host cells, but was not excreted into the culture medium. After random mutagenesis of E.coli C600 09098 with nitrosoguanidine, "leaky" mutants were isolated which prompted the release of up to 42% of total cellulolytic acitivity into the culture supernatant after transformation with either pEC1 (Gilkes et al., 1984c) or pEC2 (see below). The use of "leaky" mutants as hosts for the production of cellulases greatly facilitates their isolation and purification.

An expression vector was constructed for the secretion of exoglucanase by Saccharomyces cerevisiae (Curry et al., 1988). The nucleotide sequence coding for the signal peptide of α-
missing the coding sequence for the Cex leader peptide, and its expression was placed under the control of the MEL1 promoter (Table II). Yeast transformed with this plasmid (pMV3.cex) produced active extracellular, glycosylated exoglucanase.

I.2.2.3. Subcloning and expression of cenA

The initial clone, pEC2, coding for the endoglucanase CenA, did not contain the entire cenA gene. It encoded a Tet-CenA fusion polypeptide in which the N-terminal sequence of the tetracycline resistance determinant was fused in-frame to the coding sequence of cenA (Wong et al., 1986a). An additional C.fimi DNA cloning experiment resulted in a clone (pcEC2) which contained a contiguous stretch of DNA encoding the missing N-terminal sequence and the rest of the mature endoglucanase (Table II). The coding sequence of cenA was determined by nucleotide sequence analysis. As in the case of cex, the G plus C content was relatively high (72.5%), manifesting itself in the biased codon usage. Over 98% of the codons contained either a G or C in the third position (Wong, 1986). The complete sequence of cenA was 1347bp long, encoding 449 amino acids. The sequence coding for the mature ngCenA was preceded by a sequence of 31 codons encoding a leader peptide. This leader sequence was shown to function in E.coli since part of the endoglucanase activity was recovered from the periplasm (approx. 50% of total CMCase activity). The estimated molecular weight for the purified ngCenA of 48'700 was not in agreement with the size given by the predicted amino acid sequence (43.8kDa; Gilkes et al., 1988).

Subcloning of cenA into plasmid pUC18 to give pUC12-1.6cenA increased the length of the CenA leader peptide by 9 amino acids, including the first 6 amino acids of the E.coli β-galactosidase gene, and led to an 800-fold increase in the level of expression.
of cenA in E.coli ( Guo et al., 1988 ). Overexpression of ngCenA by this construct also resulted in a temperature dependent leakage of periplasmic proteins, including CenA, into the culture medium.

In addition, the endoglucanase gene was fused to the leader sequence of the yeast preprotoxin K1 gene ( plasmid pK2.4 ) and introduced into Saccharomyces cerevisiae ( Skipper et al., 1985 ). The fusion protein was excreted into the culture medium and was shown to be glycosylated.

In Rhodobacter capsulatus, cenA was expressed by the plasmid pREC2.2 ( Table II ) as a fusion protein ( Johnson et al., 1986 ). The 5'-terminal portion of the B870b gene was linked to the coding sequence of cenA, and expression of the hybrid gene was controlled by the R.capsulatus rxcA promoter. CMCase activity was found predominantly in the cytosol.

I.2.2.4. Comparative biochemical analyses of cex and cenA, and of their genes

Like their native, glycosylated counterparts, both ngCenA and ngCex bind to Avicel. Based on their predicted amino acid sequences, each enzyme features three distinct regions ( Fig. 1 ). A short sequence of approx. 20 amino acids containing exclusively proline and threonine residues ( Pro-Thr box ) is conserved almost perfectly in the two enzymes. This sequence was postulated to function as a hinge region, dividing the enzymes into two functionally different domains: a) An irregular region, rich in hydroxyamino acids, of low charge density, and which was predicted to have little secondary structure, was 50% conserved in both enzymes ( Warren et al., 1986 ); biochemical analysis confirmed this region to be the cellulose binding domain ( Langsford et al.,
FIGURE 1. Proposed bifunctional structure of CenA and Cex. Numbers indicate the lengths of the cellulases in amino acids; PT denotes Pro-Thr box and AS stands for putative active sites.
1987, Gilkes et al., 1988). and b) An ordered region of higher charge density which was predicted to have secondary structure but which did not appear to be conserved in the cellulases. The putative active site sequences, Glu-Xaa7-Asn-Xaa6-Thr, were identified by comparison with lysozyme and localized within this particular region of CenA and Cex. The order of these two functionally distinct domains is reversed in the two enzymes (Warren et al., 1986). The order and structure of these features in the exo- and endoglucanase suggest that their genes have arisen by sequence shuffling and duplication.

Since *E.coli* does not glycosylate proteins, the cloning of the genes *cex* and *cenA* in *E.coli* allowed analysis of the function of the glycosyl groups of *C.fimi* cellulases. It was shown that glycosylation did not affect the stability of either enzyme to extremes of pH and temperature. In addition, no differences were observed in the values for the kinetic parameters of the two forms of Cex and CenA (Table I), (Arfman et al., 1987, Langsford et al., 1987). However, the glycosylated enzymes, when bound to insoluble cellulose, were protected from attack by the *C.fimi* serine protease, while the non-glycosylated, recombinant enzymes yielded active, truncated products with greatly reduced affinity for Avicel (Langsford et al., 1987). As a result, the binding domains of the recombinant cellulases remained bound to cellulose, whereas the catalytic domains still expressing intact enzymatic activity were released into the medium (Langsford et al. 1987, ). Amino acid sequence analysis of the truncated proteins indicated that the dominant protease cleavage sites in both cellulases were located at the C-terminal end of the hinge regions (Pro-Thr box; see Figure 1), (Gilkes et al., 1988). Presumably, glycosylation of threonine residues in the Pro-Thr box prevents the cellulose associated enzymes from being cleaved while they are still
actively involved in hydrolysis of insoluble cellulose. The effect of glycosylation on the stability of the binding of the cellulases to microcrystalline cellulose is currently being investigated.

I.2.2.5. Subcloning and expression of cenB

Plasmid pEC3 contains a 5.6kb C. fimi DNA insert coding for the endoglucanase CenB. Not surprising, the expression of CMCase activity was only moderate. Subcloning into pUC19 and replacement of the C. fimi transcriptional and translational regulatory signals with those of the E. coli lac operon ( pJB301; see Table II ) led to a 20-fold increase in specific activity ( Owolabi et al., 1988a ). Deletion mutants missing up to 35% of the cenB DNA at its 3'-end still expressed full CMCase activity, indicating that the active site resided in the N-terminal portion of CenB. Like Cex and CenA, endoglucanase B exhibited high affinity for crystalline cellulose and could be purified to homogeneity in one step by an Avicel affinity column ( 110kDa, based on SDS-PAGE ). In addition, it was shown that the C. fimi serine protease cleaved the enzyme into two polypeptides of 65kDa and 43kDa, one of which still retained the capacity to bind to cellulose ( 65kDa ), ( Owolabi, 1988 ). The sequence of cenB has not been determined yet, and biochemical data on the native C. fimi enzyme are still elusive.

I.2.2.6. Identification of an E. coli clone expressing β-glucosidase activity

The E. coli clone harbouring pEC62 was one of a few clones which were identified on the basis of their reactivity with an immune serum specific for C. fimi culture supernatant antigens ( Gilkes et al., 1984b ) but which did not hydrolyze CMC. Initial characterization of pEC62 showed that the activity was a β-
glucosidase, hydrolyzing aryl-cellobiosides, aryl-glucosides as well as celllobiose. A protein with the molecular weight of approx. 180'000, exhibiting pNPGase activity (p-nitrophenolglucoside), was isolated from the highest expressing clone, pUC13-62Δ31 (see Table II; Paradis, personal communication). Preliminary data on the kinetic parameters of this β-glucosidase are listed in Table I. It hydrolyzes not only celllobiose but is even more effective on cellotriose, cellotetrose and cellopentose as substrates, indicating that this β-glucosidase is not a true and exclusive celllobiase. Whether this recombinant enzyme corresponds with one of the two β-glucosidases identified in C.fimi cell extracts has to be seen.

1.3. Objectives for the research of this thesis

The long-term goal of our group is to understand the complex cellulolytic system secreted by C.fimi. The general approach to reach this end is to dissect the cellulase system, biochemically characterize its components and define their enzymatic properties. Since information regarding the genetics of C.fimi is rudimentary at best, we have chosen to apply molecular cloning techniques for the characterization of the genes involved in cellulose hydrolysis. Expression of these genes in E.coli or any other suitable host is providing us with a source of recombinant cellulases devoid of any contaminating cellulolytic activities. By applying genetic engineering methodology to these genes, we hope to be able to improve the efficiency of cellulose degradation by the components individually or in combination.

To date, we have cloned three C.fimi genes coding for cellulases with high affinities for crystalline cellulose. These
cellulases with high affinities for crystalline cellulose. These cellulases produce soluble cellulose intermediates which need to be further degraded by enzymes with high specific activities for the soluble substrates. In order to fill this gap in our understanding of the \textit{C.fimi} cellulase system, I concentrated on cellulases which accumulated in the substrate-free culture supernatant of cells grown on microcrystalline cellulose, using the general approach outlined above.

This thesis describes the purification of two \textit{C.fimi} endoglucanases with low affinity for insoluble cellulose. A gene was isolated from a lambda \textit{C.fimi} DNA library and was shown to encode the recombinant counterparts of the two purified, native enzymes. By transcript mapping experiments and DNA sequence analysis this gene was further defined. Finally, the construction of a high expression vector system resulted in the overproduction of the endoglucanases and allowed their preliminary characterization.

A summary flow-chart of the strategies for the cloning of the \textit{C.fimi} endoglucanase gene, \textit{cenC}, into \textit{E.coli} and for the construction of the \textit{cenC} high-expression vector, \textit{pTZP-cenC}, is appended (see Appendix).
II. MATERIALS AND METHODS

II.1. Bacterial strains, plasmids and phage vectors

_Cellulomonas fimi_ ATCC 484 was used throughout the work described in this thesis. The _E.coli_ strain JM101 has been described previously (Yanish-Perron et al., 1985). The properties of the plasmids pUC12 and pUC13, as well as M13mp19 are described by Messing (1983). For the maintenance and properties of the plasmids pTZ18/19-R/U, the reader is referred to the Genescribe-Z™ laboratory manual (UBS Genescribe-Z™ protocol). Properties of and protocols for the maintenance of bacteriophage lambda L47.1 have been described by Brammar (1982).

II.2. Biochemicals

The complementary strands of the portable translation initiation site (PTIS) were synthesized at the regional DNA synthesis laboratory at the University of Calgary. Nucleotides and primers for sequencing were obtained from Pharmacia and New England Biolabs, respectively. Radioactive nucleotides were from Amersham. Restriction endonucleases and DNA modifying enzymes were purchased from several sources (Bethesda Research Lab., Pharmacia, New England Biolabs, and Boehringer Mannheim). Lambda L47.1 DNA and the _in vitro_ packaging kit were purchased from Amersham. Chemicals for electrophoresis were supplied by Bio-Rad Labs. All HPLC grade solvents for FPLC were obtained from Fisher Scientific. Nitrocellulose membranes BA85 were from Schleicher & Schuell Inc., Biodyne membranes P/N BNRG3050 were obtained from Pall Ultrafine Filtration Corp., and HATF cellulose discs were purchased from Millipore Corp.
II.3. Preparation of rabbit serum specific for C3.2

100µg of purified C3.2 endoglucanase were mixed with incomplete and complete Freund's adjuvant at a 1:1 ratio and injected subcutaneously into a New Zealand white rabbit. After two additional boosts (50µg of C3.2 each), the rabbit was bled from an ear, and the serum was tested against purified cellulases C3.1 and C3.2, *C.fimi* culture supernatant proteins, and *E.coli* cell extract proteins in an enzyme linked immunosorbent assay (ELISA; see Fig.2) as described elsewhere (Voller *et al.*, 1976). Subsequently, large amounts of blood were collected by cardiac puncture, and the isolated immunoserum, supplemented with 0.02% Na⁺-azide, was stored at -70°C.

II.4. Protein biochemistry

II.4.1. Preparation of Affigel-10-RαC3.2 affinity column

The affinity column material was prepared according to the protocol of the supplier of the gel matrix (Bio-Rad Labs.). After coupling of total rabbit serum proteins to Affigel-10, the column material was washed with several changes of 50mM Tris-HCl, pH 7.2 plus 1M NaCl and 50mM glycine-HCl, pH 2.5, and finally stored in PBS (0.2g KH₂PO₄, 2.17g Na₂HPO₄ (7H₂O), 8g NaCl and 0.2g KCl per liter) plus 0.02% Na⁺-azide at -70°C.

II.4.2. Determination of protein concentrations

The protocol for assaying proteins with the Bio-Rad dye reagent is based on the colorimetric determination of soluble proteins as described by Bradford (1976). Bovine serum albumin was used as standard.
FIGURE 2. ELISA of rabbit B3 anti C3.2 serum. Test antigens were coupled in triplicate to microtitre plates using the following amounts of protein per well: 20ng of C3.1, 50ng of C3.2, 2 μg of C.fimi culture supernatant proteins and 10μg of E.coli cell extract proteins. The optical absorbances at 405nm reflect the mean values of the activity of alkaline phosphatase conjugated to goat IgG antibodies against rabbit IgG bound to the test antigens.
II.4.3. Polyacrylamide gel electrophoresis (SDS-PAGE)

Electrophoretic separation of proteins was accomplished by a modification of the method of Laemmli (1970). 7.5% polyacrylamide gels with 0.1% SDS were used for the analysis of the C3.1/2 and CenCl/2 enzymes. The high molecular weight protein standards were supplied by Sigma. The gels were stained with Coomassie blue and dried on cellophane paper.

II.4.4. Western blot analysis

The procedure for the transfer of proteins from polyacrylamide gels to nitrocellulose membranes was described previously (Towbin et al., 1979). Glycoproteins were visualized by incubating the blocked (1% BSA in TBS: 20mM Tris-HCl, pH 7.5, 500mM NaCl) and prewashed (TBS with and without 0.2% Tween) membranes with Concanavalin A-horse radish peroxidase conjugate (4μg/ml in TBS plus 0.2% BSA) for 1 to 3h at room temperature. Before the addition of the HRP color development reagent, the membranes were washed several times as outlined above. Color development was initiated and stopped as directed by the supplier of the HRP color development reagent (Bio-Rad Labs.).

II.4.5. Amino acid sequence analysis and amino acid composition determination

Protein samples for amino acid sequence analysis were desalted by Biogel P-6DG chromatography and lyophilized prior to their shipment to the sequencing laboratories. N-terminal amino acid sequences were determined either by Dr. D. McKay, Protein Sequence Facility, University of Calgary, or by S.L. Kielland,
Microsequencing Facility, University of Victoria, on an Applied Biosystems 470A gas-phase sequenator. The amino acid composition of C3.2 was determined by Dr. D. McKay with a Beckman 6300 amino acid analyzer.

II.4.6. Fractionation of bacterial cells

The cellular locations of $\beta$-lactamase, glucose-6-P-DHase and endoglucanase were determined as follows. Cells harboring the plasmid pTZP-cenC were grown at 30°C in Luria broth (LB; Miller, 1972) supplemented with 100$\mu$g of ampicillin/ml to an OD (600nm) of 1.0. Induction of expression was initiated by addition of isopropyl-$\beta$-D-thiogalactoside (IPTG) to 1mM, and incubation of the culture at 30°C was continued for another 1$^{1/2}$h. The culture was divided into two parts, and the cells were harvested by centrifugation. The cells of one part were subsequently passed several times through a French Pressure cell, and the cell debris were separated from the soluble lysate (total cell extract) by centrifugation. The second cell aliquot was subjected to osmotic shock treatment as described elsewhere (Nossal and Heppel, 1966). The periplasmic fraction was separated from the spheroplasts by centrifugation. Finally, the cytoplasmic fraction was prepared by passing the spheroplasts through the French Press and recovering the soluble components by centrifugation.

II.5. Enzyme assays

II.5.1. DNS-CMCCase assay

Total cell extract, cytosolic and periplasmic fractions as well as culture supernatant samples were examined for cellulolytic
activity by assaying CMC hydrolysis using dinitrosalicylic acid (DNS; Miller et al., 1960). The assay conditions were as follows: 500µl of CMC-solution (4% low viscosity CMC in phosphate buffer: 50mM potassium phosphate, pH 7.0, 0.02% Na⁺-azide) were incubated at 37°C with samples of enzyme preparations in a total reaction volume of 800µl. At intervals, 100µl aliquots were removed and the reactions stopped by addition of 800µl of DNS-solution (500µl DNS reagent, 275µl of phosphate buffer, plus 25 µl of glucose at 1mg/ml). After the final aliquot was taken, the samples were steamed for 15 min, and the optical absorbance at 550nm was measured. For the quantitative determination of CMCase activity (µmoles reducing sugars produced per min) a glucose standard assay (with glucose being titrated from 0 to 100µg) was run parallel to the experiment.

II. 5.2. pNPCase assay

Hydrolysis of p-nitrophenolcelllobioside (pNPC) was followed spectrophotometrically by measuring the release of p-nitrophenol. 500µl of an enzyme preparation was incubated at 37°C with 500µl of pNPC-solution (5mM pNPC in 100mM potassium phosphate, pH 7.0, 0.02% Na⁺-azide). Aliquots were removed and added to 500µl of 1M Na₂CO₃, and the optical absorbance was measured at 410nm. The micromolar extinction coefficient for pNP at 410nm in Na₂CO₃ at 25°C is 18.8 per cm (Stoppock et al., 1982).

II. 5.3. β-lactamase assay

The β-lactamase activity in cellular fractions was assayed by monitoring the increase in absorbance at 486nm as a result of hydrolysis of nitrocefin (O'Callaghan et al., 1972). Briefly, 50µl of nitrocefin solution (5.2mg nitrocefin in 500µl DMSO and
9.5ml of 50mM potassium phosphate, pH 7.0 ) was incubated with 50μl of enzyme sample and 900μl of 50mM potassium phosphate buffer. Immediately after mixing the samples, the increase in optical absorbance was followed in the spectrophotometer. A 10mM solution of nitrocefoic acid has an OD( 486nm ) of 1.55 in a 1cm cuvette.

II.5.4. Glucose-6-P-DHase assay

Glucose-6-P-DHase activity was assayed by measuring the increase in optical absorbance at 340nm resulting from the reduction of NADP. The following solutions were mixed and equilibrated at 30°C prior to the addition of enzyme: 900μl of 55mM Tris-HCl, pH 7.8, 3.3mM MgCl₂ plus 33μl of 6mM NADP and 33μl of 0.1M glucose-P. The reaction was started by the addition of 33μl of enzyme sample, and the optical absorbance was recorded after 30sec intervals. Glucose-6-P-DHase activity was expressed in relative units: change in absorbance (340nm) per min.

II.5.5. Congo red plate assay

Bacterial colonies or recombinant lambda phage particles were grown at 37°C on agar plates containing the appropriate growth medium and antibiotics, and 15g of low viscosity CMC per liter of agar medium. After removing the colonies and the top agar containing the lambda plaques, the plates were incubated with Congo red solution (2mg/ml), (Teather and Wood, 1982). Finally, the plates were destained with 1M NaCl till zones of clearing were visible.
II.6. Electron microscopy

A culture of JM101 cells harboring the plasmid pTZP-cenC was grown at 30°C in LB plus 100μg of ampicillin/ml to an optical absorbance at 600nm of 1.0. IPTG (1mM final concentration) was added to one half of the culture and incubation continued for 90 min. Then, the cells were harvested from both halves by centrifugation and washed twice with 0.2M potassium phosphate, pH 7.2. The cells were fixed at room temperature for 60 min in phosphate buffer containing 2% glutaraldehyde. Finally, the cells were washed three times with 6.84% sucrose in phosphate buffer. Embedding of the cells, preparation of the cross-sections and microscopic analysis was performed by Dr.D.G.Scraba at the University of Alberta, Edmonton.

II.7. DNA and RNA Methodology

II.7.1. Isolation of C.fimi DNA

Genomic DNA from C.fimi was isolated essentially as described by Maniatis et al. (1982). A culture was grown at 30°C in LB low salt culture medium to a density of 0.5 to 1.0x10^9 cells/ml. The chromosomal DNA was recovered after lysing the lysozyme pretreated cells with 0.5% SDS (65°C, 30 min). RNA was subsequently digested with RNase T1 (400 units/ml). After the pronase incubation step, proteins were removed by several phenol extractions (till interface was clear), followed by ether extractions till the lysate became clear. Residual ether in the DNA preparation was evaporated by blowing N2 gas over the DNA sample while gently shaking on a Vibrax shaking platform. During the whole procedure special care was taken to avoid unnecessary shering of the chromosomal DNA. Purity of the preparation was estimated by measuring the optical absorbance at 260nm and 280nm.
The DNA was stored at -20°C till further use.

II.7.2. Isolation of plasmid and bacteriophage DNA

Small samples of plasmid DNA were isolated by a modification of the alkaline lysis procedure (Birnboim and Doly, 1979). For further purification of small scale plasmid DNA preparations, a NACS Prepac™ (BRL) chromatography step was included in the procedure.

Plasmid DNA on a large scale was isolated according to the method described by Maniatis et al. (1982). Routinely, the final purification step involved the concentration and purification of the DNA by CsCl density gradient centrifugation in the presence of ethidium bromide (EtBr).

The isolation of single stranded M13 and pTZ DNA has been described elsewhere (Messing, 1983; UBS Genescribe-Z™ protocol).

Recombinant lambda L47.1 DNA was isolated as follows. Top agar containing phage particles from single plaques together with E.coli NM538 indicator cells was spread on LB agar plates to give confluent lysis. Top agar containing phage particles was harvested and processed essentially as described by Maniatis et al. (1982). The phage particles were purified on a saturated CsCl density gradient and then dialyzed against 50mM Tris-HCl, pH 8.0, 20mM NaCl, 10mM MgCl₂. The DNA was finally recovered by the SDS-boiling method (Maniatis et al., 1982) and purified by organic extractions.

II.7.3. Preparation of template DNA for sequencing

The procedure described in the UBS Genescribe-Z™ manual was
followed for the isolation of single stranded pTZ DNA. Superinfection of *E. coli* JM101 cells harboring recombinant plasmids with helper phage M13K07 led to the accumulation in the culture medium of single stranded pTZ DNA enveloped in M13 virion proteins. To obtain template DNA of high purity, an alkaline sucrose gradient centrifugation step was added as the final step in the DNA purification procedure. After PEG precipitation of the phage-like (but not infectious) particles and resuspension in 0.5ml of 10mM Tris-HCl, pH 7.2, 1mM EDTA, the samples were overlaid onto a sucrose step gradient (four steps ranging from 5 to 20% sucrose) containing 10mM Tris-HCl, pH 7.2, 0.8M NaCl and 0.2M NaOH. The polyallomer tubes containing the gradients were centrifuged for 15h at 25K rpm in a Beckman SW 50.1 rotor. Fractions of ~0.5ml were collected from the bottoms of the tubes, and the optical absorbance at 280nm and 260nm was determined (Fig.3). The DNA was reproducibly recovered in the lower third of the gradients. Samples purified as described above and stored for several months as dry pellets at -20°C were just as satisfactory for sequencing as fresh preparations. The alkaline sucrose gradient centrifugation step eliminated the problem of high background in sequencing *C.fimi* DNA.

II.7.4. Construction of plasmid deletions for sequencing

A series of deletions in plasmid pTZ18R-8/5-5 was generated by following the Dale protocol (Dale et al., 1985). The procedure is outlined in Fig.4.

A second set of deletions was created using Bal31 exonuclease (New England Biolabs) after linerizing the plasmid DNA with BamHI. The optimal amount of exonuclease and the digestion times required for the generation of the desired deletion fragments were determined empirically by analyzing the extent of the deletions on
FIGURE 3. Purification of template DNA by alkaline sucrose gradient centrifugation. Approx. 0.5ml fractions of an alkaline sucrose gradient were collected dropwise after puncturing a hole in the bottom of the tube. A.) Optical absorbance at 260nm and 280nm of individual fractions were measured, and the resulting R-values (OD[260nm]/OD[280nm]) were plotted (o-o) against the location of the fractions in the gradient (in mls). The concentrations of DNA (µg DNA/ml) in the fractions based on optical absorbance at 260nm are graphically illustrated as dotted bars. B.) Aliquots of each sample equivalent to 0.2µg of single stranded DNA were electrophoresed on an agarose gel and stained with EtBr; m, lambda-HindIII DNA size standards.
FIGURE 4. Generation of deletions following a modification of the Dale procedure. Single stranded pTZ18R-8/5-5 DNA was hybridized with a 2-fold molar excess of primer R23 at 50°C for up to 3h. The primer R23 ( 5'CGACTCACTATAGGGAATTCCCC-3' ) was especially designed to reconstitute the EcoRI site ( ER1 ) of the multiple cloning site in pTZ18R and to hybridize to 14 nucleotides of the lac z gene proximal to the EcoRI site. Then, the DNA was completely digested with EcoRI. The extent of plasmid linerization was checked on an agarose gel, and the DNA was redigested with EcoRI if necessary. This step proved to be critical in avoiding the generation of large numbers of transformants harboring undeleted plasmids. DTT and BSA were added to the reaction mixtures to the final concentrations of 10mM and 200μg/ml, respectively. Exonuclease digestion was initiated by addition of T4 DNA polymerase ( 1 to 5 units per μg DNA; New England Biolabs ). I found that C.fimi DNA was digested more efficiently when the reaction mixture was incubated at higher temperature ( 45°C ). Aliquots containing appropriately deleted fragments were pooled. Terminal deoxynucleotidyl transferase ( TdT ) and dGTP were used to synthesize guanosine homopolymer tails at the 3'-termini of the DNA fragments ( 8 units TdT per ml reaction mixture, 5μM dGTP final concentration, for 20 min at 45°C, followed by heat inactivation ). Fresh primer R23 was subsequently added in a 2 to 4-fold molar excess and annealed at 45°C for 2 to 4h. The recircularized deletion fragments were finally ligated for 12 to 15h at 14°C with 5 to 10 units of T4 DNA ligase and ATP ( 8mM final concentration ), and E.coli JM101 transformants were screened for appropriate deletion mutants by analysis of plasmid DNA on agarose gels. Considerable improvement in the number of transformants was achieved by repetitive cycles of heating the ligation mixtures followed by ligation at 14°C prior to transformation.
primer annealing

- EcoR1
- T4 DNA Pol, Δt

exonuclease deletions

- TdT, dGTP

tailing reaction

- primer band-aid annealing

ligation, transformation
agarose gels. Deletion fragments of the appropriate size were isolated from agarose gels and subcloned into pTZ18R. Subclones containing deletion mutants were screened by analyzing the isolated plasmid DNA on agarose gels.

II.7.5. DNA sequencing

The nucleotide sequences of DNA cloned into pTZ and M13 vectors were determined by the dideoxy chain termination method (Sanger et al., 1977). The following modifications helped to diminish "compressions" inherent to DNA with high G plus C content. The molar ratios of dideoxy- to deoxynucleotides in the nucleotide mixes were adjusted for sequencing G plus C rich DNA (Mizusawa et al., 1986); dGTP was replaced with 7-deaza-dGTP; finally, the sequencing reactions were incubated at elevated temperatures (37 - 45°C) to destabilize secondary structures in the template DNA. The protocol for the electrophoretic analysis of the sequencing reaction products is described by Maniatis et al. (1982).

II.7.6. Construction of the C.fimi DNA-lambda library

C.fimi DNA was partially digested with Sau3A. Fragments ranging in size from 8 to 20kb were isolated from an agarose gel, and the ends of the fragments were dephosphorylated using calf intestinal alkaline phosphatase. The arms of the lambda vector L47.1 were prepared by ligating the cohesive ends followed by total digestion with BamHI and isolation of the arms (free of the stuffer fragment) from an agarose gel. Ligation of the C.fimi DNA fragments with the lambda DNA arms resulted in the formation of recombinant DNA concatamers which were subsequently packaged in vitro as described elsewhere (Hohn, 1979). The phage particles were plated on the non-permissive host E.coli NM359 (P2) to select
for recombinant phage vectors (first round of amplification). The total yield of viable phage particles amounted to approx. 2x10^5 pfu (7x10^4 pfu/µg of ligated and packaged DNA). The number of plaques on *E. coli* NM538 (permissive host) compared to the yield on the P2-lysogen (*E. coli* NM359) indicated that more than 95% of the phage particles had an insert. The recombinant L47.1 phage particles from the first and subsequent amplifications were stored at 4°C.

II.7.7. Synthesis and purification of oligonucleotides

Oligonucleotides were synthesized by Dr. T. Atkinson of the Department of Biochemistry, University of British Columbia, using an Applied Biosystems automated DNA synthesizer, model 380A. They were purified by preparative polyacrylamide gel electrophoresis and C18 cartridge (Sep-Pack; Millipore Corp.) chromatography (Atkinson and Smith, 1984).

II.7.8. Labeling of DNA with ^32_p

Phosphorylation of synthetic oligonucleotides with γ-^32_p-ATP and T4 polynucleotide kinase (PNK) are described by Zoller and Smith (1983). For the labeling of 5'-ends, double stranded DNA fragments were treated with calf intestinal alkaline phosphatase prior to phosphorylation with γ-^32_p-ATP and PNK. To label 3'-ends of DNA fragments, the Klenow fragment of *E. coli* DNA polymerase I was used in a fill-in reaction with α-^32_p-dNTPs. When required, the labeled fragments were further purified by non-denaturing PAGE and subsequent electrophoresis (Maniatis et al., 1982).
II.7.9. Screening methods for recombinant bacteriophage and plasmid vectors

The screening for CMCase activity to identify positive lambda particles by the Congo red plate assay was accomplished by following the method outlined above and as described elsewhere (Kotoujansky et al., 1985) except for the following modification. Plates containing the phage plaques were replicated to Millipore cellulose discs (0.45μm). The discs containing phage particles and E.coli NM538 cells were placed on fresh LB plates till plaques became visible (master plates). The original plates were then used in the Congo red plate assay after scraping off the top agar. Finally, positive clones were identified by matching the haloes with the plaques on the master plates.

The method for the screening of the C.fimi DNA-lambda library and recombinant M13 phage particles with $^{32}$P-labeled oligonucleotide probes is outlined by Woods (1984) and Maniatis et al. (1982). Briefly, the phage plaques were grown to approx. 0.5 to 1.0mm in diameter. The phage particles were lifted from the plates with Biodyne membrane circles (plaque-filter lift) and lysed with an alkaline buffer (Woods, 1984). Subsequently, the DNA was bound to the membranes by baking. After washing the membrane circles with a SDS-solution at 68°C for up to 15h (Woods, 1984) and prehybridization at 37°C for 2h, the membranes were incubated with $^{32}$P-labeled oligonucleotide probes in hybridization buffer (Woods, 1984) at 42°C for 12 to 15h. Subsequently, the membranes were washed in 6xSSC, 0.05% Na$^+$-pyrophosphate for 5 min intervals at increasing temperatures untill positive signals on autoradiograms clearly stood out over background. Normally, washing the membranes at the temperature of Tm - 5°C led to the dissociation of non-specifically hybridized oligonucleotide probes. The Tm was estimated according to the equation:
Tm = 16.6 \log [\text{Na}^+] + 0.41 (\text{G + C in \%}) + 81.5 - 500/n

where n is the number of nucleotides in the hybridization probe (Wahl et al., 1987). Hybridization positive plaques were finally picked, purified and rescreened. The clonal recombinant lambda particles were stored in suspension at 4°C.

*E.coli* clones harboring recombinant plasmids were screened by the colony-filter lift hybridization technique, essentially as described above. Generally, prior to the colony-filter lift, single colonies of transformed *E.coli* JM101 cells were transferred in a grid pattern to fresh plates to simplify the identification of positives. Positive clones were stored either as single colonies on agar plates or as 20% glycerol stocks at -20°C.

II.7.10. DNA and RNA dot-blot analysis

0.5 to 2pmoles of plasmid DNA were denatured according to Maniatis et al. (1982), spotted onto Biodyne membranes and fixed to membranes by baking (80°C, 2h). Between 5 and 10μg of total *C.fimi* RNA per dot in 10mM Tris-HCl, pH 7.6, 1mM EDTA were applied to Biodyne membranes. After baking, the DNA and RNA blots were further processed as described in section II.7.9.

II.7.11. Southern transfer analysis

The protocol for Southern transfer analysis was described in detail by Maniatis et al. (1982). Hybridization of the DNA on the membranes with $^{32}$P-labeled DNA probes and the subsequent washing steps were performed as outlined in section II.7.9.
II.7.12. cDNA synthesis by primer extension

20pmole samples of single stranded pTZ18R-S/B DNA were resuspended in 200\textmu l of HincII buffer (BRL). An equimolar amount of primer-8/5 (5'-GTTTCTCGAGGTCATCA-3') was added, then the samples were heated to 90\textdegree C for 5min and incubated at 50\textdegree C for 2 to 4h. Primer extension was initiated by addition of 15 units of the Klenow fragment of \textit{E.coli} DNA polymerase I plus dNTPs (0.25mM final concentration) and DTT (5mM final concentration). After 20min incubation at 50\textdegree C, a further 15 units of Klenow enzyme were added and the incubation continued for another 20min. The reaction mixtures were de-proteinized by extraction with phenol/chloroform and chloroform. The DNA was recovered by ethanol precipitaion and resuspended in 200\textmu l of mung bean nuclease buffer (New England Biolabs). To generate flush-end, double stranded DNA fragments, 50 units of mung bean nuclease (New England Biolabs) were added, and the reaction mixture was incubated at 37\textdegree C for 20min. The reaction was stopped by addition of 8\textmu l of 10\% SDS. Subsequently, the DNA was purified by organic extractions, precipitated and digested to completion with HindIII as verified by agarose gel electrophoresis. Finally, the 220bp blunt-end/HindIII fragment coding for the N-terminus of CenCl/2 was isolated from agarose gels and purified by organic extractions prior to ligation into pTZ18R-PTIS.

II.7.13. Isolation of \textit{C.fimi} RNA

Total RNA was extracted from \textit{C.fimi} exactly as described by Greenberg \textit{et al.} (1987a). Cells of \textit{C.fimi} cultures (up to 100ml) grown in basal medium (Stewart and Leatherwood, 1976) were harvested in the late log phase. Special care was taken to perform the lysis of the cells and the extraction of RNA as
quickly as possible to avoid excessive degradation of RNA by *C.fimi* nuclease. All glassware used for RNA work was either baked at 300°C for 3h or was bought as disposable labware. Distilled water used for the preparation of buffers was treated with 0.2% diethylpyrocarbonate as described elsewhere (Maniatis et al., 1982).

II.7.14. RNA-DNA hybrid protection analysis

The 3'- and 5'-ends of *C.fimi* transcripts were mapped essentially as described elsewhere (Greenberg et al., 1987a). Briefly, 20μg of total *C.fimi* RNA or the same amount of control RNA (yeast tRNA) were precipitated with 32P-labeled double stranded DNA probes and redissolved in 30μl of hybridization buffer (40mM sodium phosphate, pH 6.5, 0.4M NaCl, 0.5mM EDTA, 80% formamide). The samples were heated at 90°C for 10min, then incubated in a 60°C waterbath for 3h. Subsequently, the samples were diluted quickly with 300μl of ice-cold S1 nuclease buffer (30mM Na+-acetate, pH 4.5, 28mM NaCl, 4.5mM Zn++-sulfate) containing 300 to 700 units of S1 nuclease (BRL). After 30min incubation at 37°C, the reactions were stopped by the addition of 75μl of stop buffer (2.5mM Na+-acetate, pH 4.5, 50mM EDTA). After adding 20μg of carrier RNA (yeast tRNA), the protected fragments were precipitated by addition of 400μl isopropanol. Finally, aliquots of the 32P-labeled samples were analyzed by PAGE using denaturing gels with urea (Maniatis et al., 1982).
III. RESULTS AND DISCUSSION

III.1. Isolation and characterization of endoglucanases from 
\textit{C.fimi} culture supernatant

III.1.1. Purification of C3.1 and C3.2

Beguin and Eisen reported the secretion of two classes of 
cellulases by the \textit{Cellulomonas} sp. IIbc (Beguin and Eisen, 
1977). One class comprised enzymes which were tightly bound to 
insoluble cellulose. These enzymes were shown to be endoglucanases 
with respect to their CMCase activities, and glycoproteins by 
their reactivity with periodic acid-Schiff's stain. The second 
class was typified by a third, non-glycosylated cellulase, 
endocellulase I, found free in the culture supernatant of 
\textit{Cellulomonas} IIbc (Beguin and Eisen, 1978). The purification 
of endoglucanase I was facilitated by its ability to bind to the 
dextran Sephadex G-25. This initial purification step not only had 
the advantage of being a convenient way of concentrating the 
enzyme preparation but was also very specific for this cellulase.

For the purification of the soluble endoglucanases, C3.1 and 
C3.2, secreted by \textit{C.fimi}, we adopted and modified this Sephadex 
G-25 affinity chromatography step. The purification procedure is 
summarized in Fig. 5. Advantage was taken of the fact that C3.1 and 
C3.2 demonstrated sufficient binding affinity for Sephadex G-25 in 
the presence of 0.2M ammonium sulfate to allow their purification. 
It is interesting to note that only ~ 2% of the total cellulolytic 
activity present free in the \textit{C.fimi} culture supernatant was 
recovered by this initial purification step (Table III). The 
wash fraction in Table III includes the CMCase activities 
remaining in the culture supernatant after the batchwise 
incubation with Sephadex G-25 and the "wash" of the column prior
FIGURE 5. Purification scheme for C3.1 and C3.2. 10 ml of LB containing 0.2% glucose was inoculated with a single colony of *C.fimi* and incubated at 30°C for 24h. In the next step, two 100 ml volumes of the same medium were inoculated with 1ml of freshly grown cells and incubated as above. Finally, six 2 litre volumes of Leatherwood's medium (Stewart et al., 1976) supplemented with 0.2% Avicel, were inoculated with 20ml of the second culture and incubated for 5 days at 30°C in a rotary air shaker. The clarified culture medium was treated with PMSF (0.5mM final concentration) and NaN₃ added to 0.02%.

After addition of ammonium sulfate to a final concentration of 0.2M, the culture supernatant was incubated batchwise with 600 ml of preswollen Sephadex G-25 (preswollen in starting buffer: 40mM potassium phosphate, pH 7.5, 0.2M ammonium sulfate). Subsequently, the dextran was packed into a column, washed with starting buffer and the bound material eluted with a concave, decreasing salt gradient (40mM potassium phosphate, pH 7.5, 0.2M ammonium sulfate to 0.02% NaN₃). Fractions with enzyme activity (DNS-CMCase assay) were pooled (pool I and pool II). Each pool was fractionated using the MonoQ-FPLC system. The samples were loaded in 20mM sodium piperazine, pH 5.8, 0.2M NaCl (piperazine buffer), and the column was washed with the same buffer. The bound CMCase activities (C3.1 and C3.2) were eluted and recovered by applying a linear, increasing NaCl-gradient (0.02 to 1M NaCl in 20mM sodium piperazine, pH 5.8).

The enzyme preparations were stored in the piperazine buffer containing 0.02% NaN₃ at 4°C. Alternatively, the samples were desalted on a Bio-Gel P-6DG column prior to lyophilization and storage at -20°C.
TABLE III. Flow-chart of the purification of C3.1 and C3.2. Activity was determined by the DNS-CMCase assay. Total activity in the starting material includes all soluble CMC-hydrolyzing components secreted by *C. fimi* when grown under the conditions given in Fig. 5. Units (u) are in μmoles glucose equivalents (CMC-reducing ends) produced per minute.

<table>
<thead>
<tr>
<th>Step</th>
<th>total protein (mg)</th>
<th>activity (u/ml)</th>
<th>total activity (u)</th>
<th>specific activity (u/mg prot.)</th>
<th>recovery of activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>before Seph. G-25</td>
<td>147</td>
<td>0.164</td>
<td>1899</td>
<td>12.9</td>
<td>100</td>
</tr>
<tr>
<td>wash</td>
<td>146</td>
<td>0.121</td>
<td>1500</td>
<td>10.3</td>
<td>83</td>
</tr>
<tr>
<td>pool I</td>
<td>0.34</td>
<td>0.056</td>
<td>8.4</td>
<td>24.3</td>
<td>0.4</td>
</tr>
<tr>
<td>pool II</td>
<td>0.82</td>
<td>0.111</td>
<td>27.1</td>
<td>33.0</td>
<td>1.4</td>
</tr>
<tr>
<td>MonoQ C3.1</td>
<td>0.03</td>
<td>0.116</td>
<td>7.1</td>
<td>236</td>
<td>0.35</td>
</tr>
<tr>
<td>MonoQ C3.2</td>
<td>0.06</td>
<td>0.076</td>
<td>22.0</td>
<td>367</td>
<td>1.2</td>
</tr>
</tbody>
</table>
to elution of the cellulases. The discrepancy in the recovery of the activity (83% in the wash fraction and approx. 2% in pool I and pool II with 15% of total activity not accounted for) is due to the addition of ammonium sulfate to the \textit{C.fimi} culture supernatant prior to the Sephadex G-25 step (ammonium sulfate was shown to decrease the sensitivity of the DNS-CMCase assay). The two major fractions, pool I and pool II, showed similar behaviour on the MonoQ anion-exchange column and were eluted in two separate peak-fractions (C3.1 and C3.2) between 430mM and 440mM NaCl. The yield of C3.1 and C3.2 from a 12 l \textit{C.fimi} culture was typically in the range of 100 - 500 µg protein, depending primarily on the volume ratio of Sephadex to culture supernatant and incubation time of Sephadex with the culture supernatant.

The molecular weights of C3.1 and C3.2 were estimated to be 130'000 and 120'000, respectively, based on their relative mobilities on SDS-PAGE (Fig.6).

III.1.2. Characterization of C3.1 and C3.2

The substrate-bound cellulases from \textit{C.fimi}, CenA and Cex, are glycosylated (Gilkes et al., 1984a, Langsford et al., 1987). The glycan modifications contain exclusively mannose residues, probably linked to threonine side-chains (Gilkes, personal communication). The glycosyl groups may be involved in the binding to cellulose (Chanzy et al., 1984). Although the recombinant, non-glycosylated CenA and Cex enzymes produced by \textit{E.coli} do bind to Avicel, glycosylation might affect the strength of the binding. This effect of glycosylation is suggested by the observation that recombinant CenA and Cex, but not their native counterparts, can be eluted from Avicel with water (Gilkes, personal communication). On the other hand, stability of the enzymes towards extremes of temperature and pH was not affected by
FIGURE 6. SDS-PAGE analysis of C3.1 and C3.2. Samples of MonoQ-FPLC purified C3.1 and C3.2 enzyme preparations were loaded in lanes a and b, respectively. Lane c, aliquot of Amicon concentrated starting material for the purification of the native enzymes (C.fimi culture supernatant); lane m, SDS-PAGE molecular weight standards. The gel was stained with Coomassie blue.

The result of a Western blot experiment (Fig.7) is interesting insofar as none of the soluble proteins present in the C.fimi culture supernatant reacted with Concanavalin A - Horseradish peroxidase (ConA-HRP) conjugate nor did purified C3.1 or C3.2, indicating that these enzymes may not be glycosylated. These data, together with the observation that C3.1 and C3.2 were recovered from the substrate-free culture supernatant of C.fimi grown on Avicel, suggest indirectly that glycosyl chains may be involved in binding of cellulases to crystalline cellulose.

The amino-terminal amino acid sequences of the native enzymes were determined in order to design oligonucleotides for use as screening probes in the cloning of the C.fimi gene(s) encoding C3.1 and C3.2. For the same purpose an internal tryptic peptide of C3.2 (T-115) was sequenced. Table IV lists the amino acid sequencing data. Surprisingly, C3.1 and C3.2 shared the same N-terminal sequence. A sample of purified C3.2 was analyzed for its amino acid composition (Table V). This procedure partially destroyed amino acids T and S and therefore their values are inaccurate. W and C residues are not included in the amino acid composition table since their determination requires specific sample preparations (methanesulfonic acid hydrolysis for W and performic acid oxidation for C).

The preliminary characterization of C3.1 and C3.2 suggested that the two enzymes were closely related to each other. Both proteins were isolated by the same purification scheme. Moreover, their separation on the MonoQ column could only be achieved by eluting with a shallow salt gradient to ensure sufficient resolution of two individual peak fractions. Neither C3.1 nor C3.2 reacted
FIGURE 7. Western blot analysis of purified C3.1 and C3.2. The reactivity of C3.1 and C3.2 with Concanavalin A-Horseradish peroxidase (ConA-HRP) was assessed as follows. After SDS-PAGE separation of *C.fimi* culture supernatant material, the proteins were transferred onto nitrocellulose membrane. The membrane was incubated with ConA-HRP conjugate. Binding of ConA-HRP was detected by addition of HRP color development reagent. Protein samples analyzed were: lane a, 35μg total soluble proteins in *C.fimi* culture supernatant; lanes b and c, 2μg purified C3.1 and C3.2, respectively; and lane d, 0.5μg purified native exoglucanase, gCex. Negative control: SDS-PAGE molecular weight standards.
TABLE IV. Amino-terminal amino acid sequence data of C3.1, C3.2 and the C3.2-internal tryptic peptide, T-115. The sequences were determined by automated Edman degradation. Amino acid sequences up to position number 10 were confirmed by resequencing fresh protein preparations of both enzymes.

<table>
<thead>
<tr>
<th>No. of cycle</th>
<th>PTHAA</th>
<th>No. of cycle</th>
<th>PTHAA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C3.1</td>
<td>C3.2</td>
<td>T-115</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>L</td>
</tr>
<tr>
<td>2</td>
<td>S</td>
<td>S</td>
<td>L/E</td>
</tr>
<tr>
<td>3</td>
<td>P</td>
<td>P</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>I</td>
<td>I</td>
<td>P</td>
</tr>
<tr>
<td>5</td>
<td>G/K</td>
<td>G</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>E</td>
<td>E</td>
<td>D</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>G</td>
<td>P</td>
</tr>
<tr>
<td>8</td>
<td>T</td>
<td>T</td>
<td>Q</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>F</td>
<td>L</td>
</tr>
<tr>
<td>10</td>
<td>D</td>
<td>D</td>
<td>A</td>
</tr>
</tbody>
</table>
TABLE V. Partial amino acid composition analysis of C3.2. The molar contents of individual amino acids were based on the estimated molecular weight of 120'000 for C3.2.

<table>
<thead>
<tr>
<th>amino acid</th>
<th>composition (mol/mol prot.)</th>
<th>amino acid</th>
<th>composition (mol/mol prot.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D + N</td>
<td>99.59</td>
<td>M</td>
<td>6.72</td>
</tr>
<tr>
<td>T</td>
<td>85.98</td>
<td>I</td>
<td>17.66</td>
</tr>
<tr>
<td>S</td>
<td>87.32</td>
<td>L</td>
<td>98.60</td>
</tr>
<tr>
<td>E + Q</td>
<td>136.54</td>
<td>Y</td>
<td>47.89</td>
</tr>
<tr>
<td>P</td>
<td>94.27</td>
<td>F</td>
<td>32.61</td>
</tr>
<tr>
<td>G</td>
<td>145.56</td>
<td>H</td>
<td>19.55</td>
</tr>
<tr>
<td>A</td>
<td>145.50</td>
<td>K</td>
<td>21.57</td>
</tr>
<tr>
<td>V</td>
<td>96.81</td>
<td>R</td>
<td>32.61</td>
</tr>
</tbody>
</table>
detectably with the ConA-HRP conjugate. Also, their respective kinetic parameters for the hydrolysis of the aryl-cellobioside substrate, pNPC, were very similar: the Kms in mM were 0.13 for both enzymes, and the Vmax values in μmoles per min per mg protein were 0.39 and 0.50 for C3.1 and C3.2, respectively. Furthermore, a rabbit serum specific for purified C3.2 exhibited approximately the same reactivity (titre) in ELISA for both purified antigens, C3.1 and C3.2. And lastly, the two cellulases share the same amino-terminal amino acid sequence.

Because of their congruent properties, I was led to postulate at this stage of the project that C3.1 and C3.2 were encoded by the same \textit{C.fimi} gene. In accordance with the accepted nomenclature the gene was designated \textit{cenC} and its recombinant product CenC.

III.2. Molecular cloning of \textit{cenC}

III.2.1. Preliminary experiments

In the following paragraph I describe the appetizer to the \textit{cenC} cloning dinner. Despite its elaborate preparation, it was not a meal in itself. However, its aftertaste contributed considerably to the success of the main course.

A partial Sau3A digest of \textit{C.fimi} DNA was shotgun-cloned into pUC13. For the screening of the plasmid library, a restricted pool of oligonucleotides (JO.2) corresponding to the codons for amino acids number 7 to 11 of the N-terminal sequence of the C3 enzymes was used in a colony-filter lift hybridization experiment. The rationale for using a biased oligonucleotide pool was the previously demonstrated restricted codon usage in \textit{C.fimi} (O'Neill \textit{et al.}, 1986a). Thus far, only 35 out of 61 amino acid codons were found in approx. 6kb of sequenced \textit{C.fimi} DNA. In addition, in over 98% of the codons the third position contained
either a guanosine (G) or cytosine (C). Hence, a biased probe pool was considered appropriate for screening (Fig. 8). JO.2 comprised only four different 15-mers, with all codons ending with a G or C. The screening yielded several hybridization-positive clones; however, none expressed definitive CMCase activity in the Congo Red plate assay. Based on previous results (Wong et al., 1986a, Owolabi, 1988), we knew that E.coli recognizes the transcriptional regulatory sequences of C.fimi only poorly, and therefore, we were not surprised by the very low activity of the JO.2-positive recombinant clones. In order to test for the successful cloning of cenC the nucleotide sequence corresponding to the N-terminal amino acid sequence of C3.1/2 was determined. Consequently, the inserts of several individual clones hybridizing most strongly to JO.2 were completely digested with HaeIII and shotgun-cloned into M13mp10. Subclones to be sequenced were identified by screening with $^{32}$P-labeled JO.2. Fig. 8 shows the sequence of the best matching clone. Note that only one nucleotide did not agree with the probe and only two in total could not be aligned with the amino acid sequence of C3.2. Unfortunately, none of the clones screened with JO.2 contained the sequence corresponding to amino acids number 1 to 4. It was conceivable that certain amino acids (I_4, E_6 and D_11) had been mistakenly identified. But resequencing of new preparations of C3.1 and C3.2 confirmed the original amino acid sequencing data. This led me to conclude that none of the isolated recombinant clones contained cenC or even part of it, possibly as a result of using a biased nucleotide probe pool for screening. Consequently, for the subsequent cloning experiment (see below) a fully redundant pool of oligonucleotides were used for screening purposes.
FIGURE 8. Partial nucleotide sequence of pUC13-1/43. The nucleotide sequence of the plasmid is in alignment with the sequences corresponding to the N-terminus of C3.2 and the pool of synthetic oligonucleotides, JO.2 (bold letters). Vertical bars indicate discrepancies among the sequences.
<table>
<thead>
<tr>
<th>N-terminal A.A. sequence of C3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁  S₂  P₃  I₄  G₅  E₆  G₇  T₈  F₉  D₁₀  D₁₁  –  P₁₃</td>
</tr>
<tr>
<td>GCN  TCN/AGC  CCN  ATC  GGN  GAG  GGN  ACN  TTC  GAC  GAC  GCC  CCN</td>
</tr>
</tbody>
</table>

| JO.2 |
| C  C |
| GGG  GAC |

| pUC13-1/43 |
| 5'GGG  GCC  GGC  ACG  TTC  GAC  GTC  TCG  TGG etc. |
III.2.2. Cloning strategy

The replacement vector lambda L47.1 was chosen for the cloning of genomic *C.fimi* DNA for several reasons: the cloning capacity (4.7 to 19.6kb for cloning into the BamHI sites) is superior to most common plasmid vectors; it is an expression vector with the expression of recombinant inserts being under the control of the strong lambda leftward promoter $P_L$; and lastly, lambda L47.1 provides a convenient positive selection system (Spi-phenotype) by plating the *in vitro* packaged phage particles on the non-permissive host *E.coli* NM 539 (P2 lysogen). NM 539 allows only the propagation of recombinant L47.1 particles (gam-, red-), (Loenen and Brammar, 1980, Brammar, 1982).

Genomic DNA of *C.fimi* was partially digested with Sau3A and ligated to the BamHI arms of L47.1 to form concatamers, which were subsequently packaged *in vitro* (Hohn, 1979, Amersham, bulletin N.334). To select for and amplify recombinant phage particles, the lambda particles were plated on the P2 lysogen NM 539 (first round of amplification). I adopted the activity screening protocol described for the cloning of an *Erwinia chrysanthemi* cellulase gene for the identification of CMCase positive phage particles in the plaque-Congo red plate assay (Kotoujansky et al., 1985). Lysis of the host cells during phage propagation leads to liberation of cytoplasmic proteins and increases the sensitivity for the detection of CMCase positive clones. Approx. 2x10^4 phage particles plated on NM 538 (permissive host) were screened for expression of cellulase activity in this way.

However, most of the phage particles were screened with oligonucleotides in a plaque-filter lift hybridization assay. Fig.9 lists the pools of oligonucleotide probes prepared for this purpose. Note that the pools J0.3A,B,C and D constitute the fully redundant repertoire of oligonucleotides corresponding to the amino acid sequence depicted in the top panel of Fig.9. The
FIGURE 9. Oligonucleotide screening probes for the lambda L47.1-C.<i>fimi</i> DNA library. The letters with numbers in subscript represent amino acids and their positions in the N-terminal amino acid sequences of the <i>C.fimi</i> enzymes (C3.1/2) and the internal tryptic peptide T-115. The pools JO.3A, B, C, and D constitute the fully redundant repertoire of oligonucleotides corresponding to the amino acid sequence E<sub>6</sub> to D<sub>11</sub> of the C3-enzymes. The probe pool, BM.C, was biased with respect to <i>C.fimi</i> codon usage.

<table>
<thead>
<tr>
<th></th>
<th>E&lt;sub&gt;6&lt;/sub&gt;</th>
<th>G&lt;sub&gt;7&lt;/sub&gt;</th>
<th>T&lt;sub&gt;8&lt;/sub&gt;</th>
<th>F&lt;sub&gt;9&lt;/sub&gt;</th>
<th>D&lt;sub&gt;10&lt;/sub&gt;</th>
<th>D&lt;sub&gt;11&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3.1/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>GGN</td>
<td>ACN</td>
<td>TTC</td>
<td>GAC</td>
<td>GAC</td>
</tr>
<tr>
<td>JO.3A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>GGN</td>
<td>ACN</td>
<td>TTC</td>
<td>GAC</td>
<td>GA</td>
</tr>
<tr>
<td>JO.3B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>GGN</td>
<td>ACN</td>
<td>TTT</td>
<td>GAC</td>
<td>GA</td>
</tr>
<tr>
<td>JO.3C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>GGN</td>
<td>ACN</td>
<td>TTC</td>
<td>GAT</td>
<td>GA</td>
</tr>
<tr>
<td>JO.3D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>GGN</td>
<td>ACN</td>
<td>TTT</td>
<td>GAT</td>
<td>GA</td>
</tr>
<tr>
<td>T-115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>CCN</td>
<td>TAC</td>
<td>GAC</td>
<td>CCN</td>
<td>CAG</td>
</tr>
<tr>
<td>BM.C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>CCG</td>
<td>TAC</td>
<td>GAC</td>
<td>CCG</td>
<td>CAG</td>
</tr>
</tbody>
</table>


oligonucleotide probe pool BM.C was used for the cross-screening of JO.3 positive clones. Two plates with approx. 1.5x10^4 plaques totally were screened with each of the four JO.3 probe pools, and the positive phage particles were purified by plating and isolation of single plaques. Fig.10 shows an example of an autoradiogram of a plaque-filter lift hybridization experiment using ^32P-labeled JO.3A. Significantly, the screening with JO.3A yielded far more positive plaques (10 to 20-fold) than any of the other three probe pools, suggesting that JO.3A contained the correctly matching oligonucleotide.

A selection of JO.3A-positive or CMCase-positive clones was subsequently rescreened using different types of probes. These results are listed in Table VI. Only two clones (I69 and I80) reacted with more than one type of probe, suggesting that the genes cex, cenA and cenC are not closely linked on the C.fimi chromosome. Surprisingly, none of the JO.3A-positive clones showed significant activity by the Congo red plate assay; nor did any of the CMCase active clones react with JO.3A. However, in light of the findings with respect to the CenC-overproducing construct, this phenomenon can be explained (see chapter III.4.). Fig.11 is a diagram of the recombinant clone L47.1-I69 (positive for JO.3A and BM.C but negative for CMCase activity) which I selected for further studies. The approx. length of the C.fimi DNA insert was 10.5kb; and contained two internal BamHI sites. The unique HindIII and EcoRI sites were parts of the sequences in the left and right arms of the cloning vector.

III.2.3. Subcloning and initial characterization of cenC

Initial attempts to subclone the entire HindIII fragment of L47.1-I69 into pUC13 were unsuccessful, probably because of
FIGURE 10. Autoradiogram of a plaque-filter lift hybridization experiment. Approx. 15,000 in vitro packaged and amplified phage particles were plated with permissive host cells, NM538 on two separate plates (150mm in diameter). Plaque material was absorbed onto Biodyne membrane. Subsequently, the membrane was probed with $^{32}$P-labeled J0.3A. A selection of positive clones (three of which are indicated by arrows) were isolated and purified for further characterization.
TABLE VI. Cross-screen of recombinant L47.1 clones.

A selection of positive clones was hybridized with oligonucleotide probes specific for either cex, cenA or cenC. Upper panel: JO.3A- positive clones; lower panel: CMCase-positive clones. a: activity was determined qualitatively on CMC-Congo red indicator plates. b: oligonucleotide probes used to screen recombinant phage plaques; JO.3A and BM.C were labeled by phosphorylation, the probes specific for cex and cenA were radioactively labeled by nick-translation of 5' -sequences of the respective genes.

<table>
<thead>
<tr>
<th>λ L47.1 clones</th>
<th>activity</th>
<th>JO.3A</th>
<th>BM.C</th>
<th>CenA</th>
<th>Cex</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>114</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>124</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>134</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>145</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>157</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>169</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>180</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IV 9</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>II 16</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>II 26</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>II 36</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>II 47</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>II 71</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>II 91</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
FIGURE 11. Diagram of the recombinant clone lambda L47.1-169. The stippled box represents the insert of C.fimi DNA, the lines represent lambda DNA. Restriction enzymes are: B1, BamHI; R1, EcoRI; H3, HindIII and S3A, Sau3A. P_L stands for the leftward promoter of lambda; numbers are in kb. The restriction fragments which hybridized with either JO.3A or BM.C are indicated.
interference of the lambda P_L promoter with plasmid replication. In order to correlate cenC sequences with subfragments of the L47.1-I69 insert, the 5'-end and orientation of cenC were determined in a dot-blot hybridization experiment (Fig.12). For this, the BamHI-HindIII fragment lacking the lambda P_L promoter and BamHI-BamHI fragment of L47.1-I69 were subcloned into pUC13 (to give pUC13-B and pUC13-B/H) and screened either with JO.3A or BM.C. Since the size of cenC required to encode C3.1/2 was estimated to be approx. 3.4kb, the results from the dot-blot experiments indicated that the BamHI-HindIII fragment containing the lambda P_L promoter (Fig.11) was lacking any part of cenC.

In order to prove by nucleotide sequence analysis that the the clone L47.1-I69 contained cenC or at least part of it, a total Sau3A-HaeIII digest of pUC13-B (which hybridized with JO.3A) was shotgun-cloned into M13mp19. Positive clones were identified by plaque-filter lift hybridization using 32P-labeled JO.3A as screening probe. Partial sequences of two subclones are listed in Fig.13. Both sequences overlapped and agreed fully with the amino acid sequence. Note the unusual C.fimi codon, GGA, for glycine at amino acid position number 7. This unexpected finding offers an explanation for the lack of success of the first attempt to clone cenC: the biased pool of oligonucleotides in JO.2 did not include GGA for glycine (Fig.8).

The 7.7kb BamHI-HindIII fragment hybridizing with JO.3A and BM.C (Fig.11) was subsequently cloned into pTZ18R to give plasmid pTZ18R-8. I found the Genescribe vector system most suitable for generating specific constructs since it enabled the isolation of the plasmid as single stranded DNA (Fig.14). DNA fusion sites or specific sequences in constructs can be examined easily, without the need for further subcloning, by DNA sequence analysis using commercially available or specifically prepared sequencing primers.
FIGURE 12. DNA dot-blot hybridization experiment. pUC13-B DNA (I69-B) and pUC13-B/H DNA (I69-B/H), containing the BamHI/BamHI fragment and the BamHI/HindIII fragment of the recombinant lambda clone L47.1-169, respectively, were spotted onto Biodyne membrane and hybridized either with $^{32}\text{P}$-labeled J0.3A (panel A) or $^{32}\text{P}$-labeled BM.C (panel B). Positive control was L47.1-169 DNA; negative controls: L47.1-IV9 (unrelated recombinant clone), L47.1 and pUC13 DNA.

![A panel image showing hybridization results.]

![B panel image showing hybridization results.]

<table>
<thead>
<tr>
<th>Controls</th>
<th>L47.1-IV9</th>
<th>pUC13</th>
</tr>
</thead>
<tbody>
<tr>
<td>L47.1-169</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>L47.1</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I69-B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I69-B/H</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 13. Nucleotide sequence analysis of plasmids M13mp19-B4 and M13mp19-B5. The sequence is aligned with the N-terminal amino acid sequence of the C3 enzymes. The bold letter nucleotide triplet emphasizes unusual *C.fimi* codon for glycine (compare with oligonucleotide probe J0.2 in Fig.8).
<table>
<thead>
<tr>
<th>N-terminal</th>
<th>A₁</th>
<th>S₂</th>
<th>P₃</th>
<th>I₄</th>
<th>G₅</th>
<th>E₆</th>
<th>G₇</th>
<th>T₈</th>
<th>F₉</th>
<th>D₁₀</th>
<th>D₁₁</th>
<th>-</th>
<th>P₁₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.A. sequence of C₃.1/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td>GCC</td>
<td>TCN/AGC</td>
<td>CCN</td>
<td>ATC</td>
</tr>
<tr>
<td>B₄ / B₅</td>
<td>5'G</td>
<td>ATC</td>
<td>GGG</td>
<td>GAG</td>
<td>GGA</td>
<td>ACG</td>
<td>TTC</td>
<td>GAC</td>
<td>GAC</td>
<td>GGG</td>
<td>CCC</td>
<td>etc.</td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 14. Diagram of the Genescribe vector system pTZ18/19-R/U. Ap-R, ampicillin resistance marker gene; ori (f1 / pBR322), origins of replication of plasmids pBR322 and phage f1; lac z, β-galactosidase gene; lac p/o, promoter and operator of the lac operon; T7 promoter, phage T7 promoter in vitro transcription; MCS, multiple cloning site.
( USB, Genescribe-Z™ protocol ). The diagram of the insert in clone pTZ18R-8 (Fig.15) documents various restriction enzyme sites determined by the method of Smith and Birnstiel (1976). By Southern transfer analysis, the 5'-end of cenC, encoding the N-terminus of the mature endoglucanase, was found to be localized on a 900bp SstI-BamHI fragment (Figs.15 and 16). Similarly, the restriction fragment hybridizing with BM.C was identified as a SstI-SmaI fragment (Figs.15 and 16).

III.2.4. Sequence analysis of the 5'-end of cenC

The clone JM101[pTZ18R-8] expressed very low CMCase activity by the DNS-CMCase assay (~0.008u/mg of total cell extract proteins). In order to construct a more efficient expression system (see chapter III.4.) it was necessary to sequence the 5'-end of cenC and the DNA upstream of this region.

Deletions of the insert in pTZ18R-8/5 (Fig.17) were generated using two different strategies: starting at the SstI site of the fragment, the exonuclease activity of T4DNA polymerase was used to create deletion fragments by digesting linear, single stranded plasmid DNA exclusively in the 3' to 5' direction (see Dale's protocol in section II); deletions starting at the BamHI end of the insert were generated using Bal31 exonuclease. Arrows in Fig.17 denote the orientations and the lengths of the DNA sequences determined in each deletion construct.

The purity of the template DNA was crucial to the successful analysis of C.fimi DNA by the chain termination method (Sanger et al., 1977). Definitive data were reproducibly obtained only when the single stranded template DNA was further purified on an alkaline sucrose gradient (see section II). The relative high G plus C content of C.fimi DNA (72%) necessitated additional
FIGURE 15. Restriction enzyme map of recombinant DNA in pTZ18R-8. Localization of restriction enzyme sites was established according to the method of Smith and Birnstiel (1976). Restriction enzymes: B, BamHI, H, HindIII, K, KpnI, M, MluI, P, PstI, Sm, SmaI, S, SstI. No sites were found for BclII, BglII, ClaI, EcoRI, PvuII, SphI, and XbaI. Solid bars indicate the restriction fragments hybridizing with either JO.3A or BM.C. Stippled bars represent the lengths of inserts of two subclones of pTZ18R-8 chosen for further studies.
FIGURE 16. Southern transfer analyses of pTZ18R-8 DNA. Total digests of pTZ18R-8 DNA were resolved on a 1% agarose gel, transferred to Biodyne membranes and probed with $^{32}$P-labeled JO.3A (panel A) or with $^{32}$P-labeled BM.C (panel B). Numbers in lanes refer to restriction enzymes used: 1, KpnI; 2, PstI; 3, MluI; 4, MluI and BamHI; 5, SmaI; 6, SmaI and BamHI; 7, SstI; 8, SstI and BamHI. Lane c, control DNA (pUC12 digested with BamHI), lane m, DNA standards (lambda DNA digested with HindIII and EcoRI).
FIGURE 17. Sequencing strategy for the insert in pTZ18R-8/5-5. The pTZ18R-8/5-5 deletions D1, D2, and D3, were constructed as described by Dale et al. (1985); see also Materials and Methods. Deletions in the opposite direction were created by digestion of the plasmid with Bal 31 exonuclease (after linearization with BamHI) and subsequent subcloning into pTZ19R. The synthetic oligonucleotide, JO.T, was used as sequencing primer to confirm the B28 sequencing data. The arrows indicate the origin, direction and number of nucleotides determined in each sequencing experiment. The cross-hatched bar represents the insert in pTZ18R-8/5-5.
precautions (O'Neill et al., 1986a, Wong, 1986). Secondary structures in template DNA or sequencing reaction products manifested themselves as "compressions". Fig.18 shows an example of compressions typically seen in sequencing *C.fimi* DNA. Substitution of dGTP with the analogue 7-deaza-dGTP (7d-dGTP) in the nucleotide mixes proved to be very effective in destabilizing most secondary structures in the sequencing reaction products. A second class of compressions was due to the formation of secondary structures in the template DNA, causing the DNA polymerase to stall and fall off. Preparation of deletions starting close to extended G plus C stretches and/or incubating the sequencing reactions at higher temperature helped to alleviate this problem.

Fig.19 documents the sequence of the insert of plasmid pTZ18R-8/5. The sequence in bold letters indicates the hybridization site for one of the oligonucleotides in JO.3A. The sequencing data were in complete agreement with the amino acid sequencing results. The stretch of 32 amino acids upstream of the putative leader peptidase cleavage site, A - A, shows typical features of a leader peptide of a secreted protein (Wickner and Lodish, 1985): positively charged amino acids at the N-terminus are followed by a sequence of hydrophobic amino acids; close to the putative leader peptidase cleavage site, amino acids with small side-chains are predominant. The predicted start codon is GUG; no other in-frame start codons could be identified within 200 nucleotides proximal to the GTG. The size of the leader peptide (32a.a.) is similar to other leader sequences of *C.fimi* proteins: Cex, 41a.a.; CenA, 31a.a. and CenB, 32a.a. (Owolabi et al., 1988a, O'Neill et al., 1986a, Wong et al., 1986a).
FIGURE 18. Effect of using 7d-dGTP during sequencing C.fimi DNA. The recombinant DNA in plasmid pTZ18R-8/5-5 was sequenced by the chain-termination method. To test the effect of 7-deaza-dGTP (7d-dGTP) on destabilizing secondary structures in the sequencing products, either normal nucleotide mixes (-7d-dGTP) or nucleotide mixes containing 7d-dGTP instead of dGTP (+7d-dGTP) were used in the sequencing reactions.
FIGURE 19. Nucleotide sequence of the 5'-region of cenC. The reading frame was identified by alignment with the N-terminal amino acid sequence of C3.1/2. Letter code in superscript represents the amino acid sequence. Start codon and sequence corresponding to oligonucleotide screening probe (J0.3A) are indicated in bold letters. The letter N in the nucleotide sequence denotes undetermined nucleotide residues.

```
5' -> 3' CGCGGCGCGGCCC
GTCACGGTGTCGCGGTTCGACGTGCACGGACGCCGGCGCGNTCGTCGACGGCACCGCG
ACGTCGGGCGCGCTGGCGTCGCTACTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNCCCCCGCAGTGACGAGCAGCTGCGCAGCGCGAGGCCGACGCGCGGTTTCGGTGCTG
CCGTGTCCGACCGCGGCGGCGCGACCTTCCCCGGCTGCGACGGCAGCGAGCTGCGATCG
TACGGCAGCTGCGTCAGGGCGCGGCGGCGCTCCTCGCTTTCGCGCCGACGCGGCTC
TCGCTTGGTGAGCCGACGGGCGGCGGCGGTCCGGCCTCCTCCTGCCGGACGCGCTC
ACCGGGCCACGGCGGGGCGGCGGCGTCCCCTGCCGGACCCCTGTCTGGAGCCGTCCGCGAC
CACCGGGCAGGGGTGGAGCGCTCCTACGACAGGGGAGACAGAG

M V S R S S Q A R G A L T A
GTG'GTT'TCT'CGC'AGG'TCA'TCA'CAG'GCG'CGC'GGC'GCG'CTC'ACG'GCC'

V V A T L A L A L A G S G T A
GTC'GTC'GCG'ACG'CTC'GCC'CTC'GCG'CTC'GCC'GCG'CTC'GCG'AGC'ACC'GCC'C

L A A S P I G E G T F D D G P
TC'GCC'GCC'TCG'CCG'GTC'GAG'GGA'ACG'TTC'GAC'GAC'GGG'CCC

E G W V A Y G T D G P L D T S
GAG'GGG'TGG'GTC'GCC'TAC'GCC'ACC'GAC'GCC'CCC'CTC'GAC'ACG'AGC'

T G A L W L A V P A G S
ACG'GCC'GCC'CTG'TGG'CTC'GCC'GTG'CCG'GCC'GGA'TCC
```

BamHI
III.3. Analyses of in vivo cenC transcripts of C.fimi

III.3.1. Mapping of the 5'-ends of cenC transcripts

The expression of cex, cenA and cenB in C.fimi is regulated by the nature of the carbon source in the culture medium (Greenberg et al., 1987a,b). An initial experiment was conducted to determine the optimal conditions for transcription of cenC in C.fimi with respect to the carbon source. Aliquots of total RNA isolated from C.fimi cultures grown in minimal medium containing either glucose (0.2%), glycerol (0.2%) or CMC (1%) were blotted onto Biodyne membranes and hybridized with the 5'-cenC specific, 32p-labeled oligonucleotide JO.T. No cenC transcripts could be detected with RNA isolated from cells grown on glycerol or glucose (Fig.20). Only total RNA from cells grown on CMC contained cenC transcripts at a level which allowed their detection by the dot-blot experiment. If C.fimi does express cenC constitutively then transcripts isolated from glucose or glycerol grown cells were too few for detection by this analysis. All subsequent work was conducted with total RNA isolated from C.fimi cultures grown with CMC.

Fig.21 outlines the preparation of the hybridization probe for the 5'-mapping of cenC transcripts. The denatured, 32p-labeled probe was annealed in the presence of formamide either with total RNA from C.fimi or with unrelated RNA. Subsequently, the samples were digested with S1 nuclease, and the protected fragments were resolved by denaturing polyacrylamide gel electrophoresis. Several fragments approx. 260 nucleotides long were protected by C.fimi RNA (Fig.22). The heterogeneity of the fragments may have resulted from overdigestion of the RNA-DNA hybrids with S1 nuclease rather than representing specific start sites of cenC transcription. For fine-mapping of these fragments an appropriate dideoxy-sequencing ladder was prepared by making a sequencing
FIGURE 20. Dot-blot analysis of total *C.fimi* RNA. Total RNA was isolated from *C.fimi* cultures grown either on 1% CMC (CMC₁ and CMC₂, two individual RNA preparations), 0.2% glycerol (Gly₁) or 0.2% glucose (Glu₁) and blotted onto Biodyne membrane. The ^32^P-labeled oligonucleotide J0.T was used as hybridization probe. In the control lane, equivalent RNA samples were hydrolyzed in 0.2N NaOH prior to blotting.
FIGURE 21. Preparation of the probe for the 5'-cenC transcript mapping. Plasmid pTZ18R-8/5 was digested with BamHI, dephosphorylated with calf intestinal alkaline phosphatase and labeled with T4 polynucleotide kinase plus γ-(32)P-ATP. Subsequently, the DNA was digested with EcoRI, and the fragments were separated by non-denaturing PAGE. Finally, the 900bp probe was electroeluted and concentrated prior to use. The black box represents part of the multiple cloning site in pTZ18R; the stippled box indicates C.fimi DNA. The following restriction enzyme sites are represented: B, BamHI, H3, HindIII, Sm, SmaI, Ss, SstI, and R1, EcoRI.
FIGURE 22. S1 nuclease protection analysis of 5'-ends of cenC transcripts. Protected, $^{32}$P-labeled DNA fragments were separated by denaturing PAGE. Lane 1, intact probe (no S1 nuclease treatment); lane 2, probe hybridized with yeast tRNA and digested with S1 nuclease (control); lane 3, S1 nuclease protected fragments after hybridization with total C.fimi RNA. Standards (M) are $^{32}$P-labeled fragments from single stranded M13 DNA digested with HaeIII.
ladder from the probe in plasmid pTZ19R-B/S. Digestion of the reaction products with BamHI prior to their loading on the gel ensured that the sequencing fragments shared the same start site as the $^{32}$P-labeled hybridization probe. In this way the main protected fragments could be mapped to the nucleotides indicated by arrows in Fig. 23. Note that the arrows indicate the putative start sites of the three dominant fragments of a total of four to five protected fragments.

The diagram in Fig. 24 shows the putative start sites of in vivo cenC transcripts (bold letters) and part of the proximal non-transcribed nucleotide sequence of cenC. The sequence proximal to the putative cenC mRNA start sites was compared for homology with some known bacterial promoters (Rosenberg and Court, 1979, Hawley and McClure, 1983). Two E.coli promoters (lexA and rrnEl P1) showed some degree of homology with the -10 and -35 hexamers of the proposed cenC promoter (Miki et al., 1981 and Gilbert et al., 1979). There was no homology with Gram-positive transcriptional regulatory sequences. However, it should be noted that these sequences have to be examined further and tested for their functionality either by studies with purified C.fimi RNA polymerase or by in vivo analysis. A prerequisite for the latter approach is a transformation procedure for C.fimi which would allow the re-introduction of our postulated C.fimi promoters in order to test them for their ability to initiate transcription.

III.3.2. Mapping of the 3'-ends of cenC transcripts

The 3'-ends of in vivo cenC transcripts were also determined by S1 nuclease protection analysis. Fig. 25 shows the strategy for the preparation of the hybridization probe. The choice of the restriction enzyme fragment to use as probe was based on the
FIGURE 23. Fine-mapping of 5'-ends of cenC transcripts. 5'-labeled DNA probe was treated with S1 nuclease after hybridization with total *C.fimi* RNA (lane 1) or with yeast tRNA (lane 2). A sequencing ladder was prepared by sequencing pTZ19R-B/E with reverse primer followed by complete digestion of the sequencing reaction products with BamHI (to ensure that the fragments share the same 5'-ends as the hybridization probe). Arrows indicate the nucleotides corresponding to the 5'-ends of *C.fimi* transcripts.
FIGURE 24. 5'-region of cenC showing the proposed transcription start sites. The non-transcribed coding strand of cenC proximal to the 5'-ends of in vivo transcripts is compared with known bacterial promoters. The lexA and rrnE1 P1 -10 and -35 hexamers are given for comparison with corresponding cenC sequences. Putative cenC mRNA start sites are shown in bold letters.
5' --> 3'

ACCGAAACGATTCGGTCCGGTGCGCACGCTTCCGCGGGCGGCCGCGGCCGTGCCTACCCTGCCGCCGCAGCACCGGC

-35

TTCCAA

17bp

TAT

ACT

rrnEl P1

TTGC GG

16bp

----

TATAAT
FIGURE 25. Scheme used for the preparation of the probe for the 3'‐cenC transcript mapping. The SmaI/HindIII fragment from pTZ18R-8/5 was first subcloned into pTZ18R (to give pTZ18R-8/5-4) and digested with MluI. The 3'-ends were labeled by a fill-in reaction using Klenow enzyme plus dGTP and α-32P-dCTP. After digestion with HindIII, the 650bp MluI/HindIII fragment was isolated as described in the legend to Fig.21. The black boxes represent parts of the multiple cloning site of pTZ18R; the stippled boxes indicate C.fimi DNA. The following restriction enzyme sites are indicated: B, BamHI, H3, HindIII, M, MluI, Sm, SmaI, Ss, SstI, and R1, EcoRI.
- R1  Ss  
- Sm  
- HindIII  
- pTZ18R  

- Smal  
- HindIII  
- pTZ18R  

- MluI  
- Klenow, α - P-dCTP, dGTP  
- HindIII  
- PAGE  

-3' 32P  
550bp  
-5'
estimated size of cenC and its localization in the plasmid pTZ18R-8/5. The autoradiogram of the S1 nuclease protection experiment revealed two major protected fragments and a smear of fragments of smaller size (Fig. 26). Again the RNA-DNA hybrids might have been overdigested with the nuclease. The two major bands mapped to two guanosine nucleotides depicted in bold letters on the distal end of the pTZ18R-8/5 insert (Fig. 27). The 3'-ends of the transcripts were located very close to the original cloning site of the C.fimi insert adjacent to the left arm of L47.1, indicating that the entire gene for CenC was contained on the plasmid pTZ18R-8/5. The lengths of the transcripts were approximately 3.4kb. Palindromic sequences which are postulated to function as transcriptional terminators are underlined by arrows in Fig. 27. The putative translational stop codon (TGA) can be verified once the entire sequence of cenC is known.

III.4. Overproduction of CenC in E.coli

III.4.1. Construction of pTZP-cenC

None of the subclones of cenC expressed significant CMCase activity as measured in the DNS-CMCase assay. The clone JM101[pTZ18R-8] when induced with isopropyl-β-D-thiogalactopyranoside (IPTG) was most active with a specific activity of 0.008 u/mg protein (units in µmoles glucose equivalents per minute) for total soluble cell extract proteins. Based on the specific activities of the purified C3 enzymes, a one litre culture of JM[pTZ18R-8] grown to an OD600 of 1.0 would produce less than 5 µg of CenC. Therefore, it was necessary to construct a better expression system to facilitate the purification of large quantities of CenC.

The efficiency of expression of recombinant genes depends on
FIGURE 26. Mapping of 3'-ends of *C.fimi* transcripts. The 3'-end labeled DNA probe was hybridized to either total *C.fimi* RNA (lane 1) or yeast tRNA (lane 2) and digested with S1 nuclease prior to electrophoresis on polyacrylamide gel. DNA standards (M) as in Fig. 22.
FIGURE 27. Nucleotide sequence at the 3'-end of cenC. The putative translational stop codon (TGA) of cenC is shown in bold letters. The two solid, vertical arrows indicate the nucleotides corresponding to the 3'-ends of the in vivo cenC transcripts. The locations of palindromic sequences are underlined by arrows. The cloning site (Sau3A/BamHI) of C.fimi DNA into the left arm of lambda L47.1 is indicated in bold letters and underlined.
several factors. Besides the copy number of the gene, the strength of the promoter on the expression vector is an important consideration. In addition, the degree of recognition by the host of heterologous translational regulatory sequences and their spacing relative to the translational initiation codon contributes to the efficiency of expression. Other factors such as codon usage by the host, messenger RNA stability and toxicity of the recombinant product may also play an important role.

I found the lac promoter-operator system (lac p/o) suitable for the construction of an efficient cenC expression system. Besides providing a relatively strong promoter, the expression of any gene under its control is regulated in an appropriate host (e.g. the E.coli strain JM101 which expresses high level of the lac repressor protein constitutively). In case the recombinant protein is toxic for the host, the cells are grown up under repressed conditions. Addition to the culture medium of IPTG will induce expression and the recombinant protein will be synthesized.

The synthetic portable translation initiation site (PTIS) was specially designed for optimal function in E.coli (deBoer et al., 1983). When fused to the gene to be expressed, the initiation of translation occurs with high efficiency, thus providing optimal conditions for the expression of cenC.

Fig.28 outlines the strategy for the construction of pTZP-cenC. The PTIS fragment, which also provided the ATG start codon, had to be fused in frame to the second codon, GTT, of the leader sequence of cenC. The double stranded DNA fragment corresponding to the 5'-end of cenC was synthesized by primer extension using primer-8/5, which hybridized to codons 2 to 7 of the leader sequence. This fragment was fused by blunt-end ligation to the ATG of the PTIS in pTZ18R-PTIS (Fig.29). JM101 transformants were screened in a colony-filter lift hybridization assay using $^{32}$P-labeled J0.3A, and
FIGURE 28. Construction of pTZP-cenC.
The black boxes represent the multiple cloning site (or part of it) of pTZ18R, the lightly-stippled boxes indicate the PTIS sequence, and C.fimi DNA containing cenC or part of it is represented by dark-stippled boxes. The following restriction enzyme sites are indicated: B, BamHI, H3, HindIII, Rl, EcoRI, and Ss, SstI. B/bl with bl standing for blunt-end, denotes the fusion point of cenC with PTIS.
A: PTIS was ligated into the BamHI and EcoRI sites of pTZ18R. The construct (pTZ18R-PTIS) was confirmed by sequence analysis.
B: The 900bp EcoRI/BamHI fragment, encoding the cenC-leader peptide and part of the mature CenC sequence, was subcloned from pTZ18R-8/5 into pTZ18R to give pTZ18R-S/B.
C: In frame fusion of the cenC coding sequence to the ATG of PTIS in pTZ18R-PTIS resulting in pTZ18R-PTIS-5' (see Fig.29).
D: The 3.5kb BamHI/HindIII fragment of pTZ18R-8/5 coding for the C-terminus and most part of CenC (complementing the cenC sequence in pTZ18R-PTIS-5' ) was isolated and cloned into pTZ18R-PTIS-5' to complete the construction of pTZP-cenC. Competent JM101 cell were transformed and the resulting recombinant clones were screened for CMCase activity in the Congo red plate assay. Positive clones (JM101[pTZP-cenC]) were further analyzed (see below).
FIGURE 29. Construction of PTIS-cenC fusion by primer extension.
Primer-8/5 was annealed to single stranded pTZ18R-S/B DNA. The sequence of the synthetic primer corresponded to the second codon and extended over the consecutive five codons of the leader sequence of cenC. Subsequently, double stranded DNA was synthesized by primer extension using Klenow enzyme and deoxynucleotides. In the next step, the single stranded gaps were eliminated by Mung Bean nuclease treatment. The DNA was digested with HindIII, and the resulting blunt-end/HindIII fragment, encoding the N-terminus of the unprocessed CenC, was isolated and cloned into pTZ18R-PTIS. For this step (prior to cloning), pTZ18R-PTIS was linearized with BamHI and treated with Mung Bean nuclease to provide a PTIS sequence with flush ends right after the ATG. The vector was further digested with HindIII, isolated and used for the construction of pTZ18R-PTIS-5' as described above. Recombinant JM101 clones were subsequently screened by colony-lift filter hybridization with $^{32}$P-labeled J0.3A. The correct fusion construct was confirmed by sequence analysis.
the fusion construct (pTZ18R-PTIS-5' ) was confirmed by sequence analysis. Finally, the missing 3'-region of *cenC* was cloned into pTZ18R-PTIS-5' to complete the construct of pTZP-cenC.

Initial attempts to synthesize the full length of *cenC* (except start codon GTG ) by primer extension failed. The primer-8/5 was subsequently tested for its specificity in priming sequencing reactions in pTZ18R-8/5. The sequencing results indicated multiple priming which explained the difficulties in getting full-length, double-stranded *cenC* DNA. The problem of multiple priming was solved by subcloning the 5'-end of *cenC* (approx. 250bp ) into pTZ18R prior to primer extension.

III.4.2. Characterization of the expression of *cenC* in JM101[pTZP-cenC]

The clone JM101[pTZP-cenC] did not grow on LB plates containing 100μg of ampicillin/ml and supplemented with IPTG (1mM) when incubated at 37°C. Only at 30°C or below was growth possible but still at a reduced rate compared to the control clone JM101[pTZ18R]. Apparently, the increased expression of *cenC* had a deleterious effect on cell growth. In liquid culture this clone could only be normally propagated by omitting IPTG (repression of expression of *cenC*) and decreasing the culture temperature to 30°C.

Several subclones, including JM101[pTZP-cenC] were compared with each other for their activity in the Congo red plate assay (Fig.30). Although the size of the colonies of JM101[pTZP-cenC] were markedly smaller than the others, the increase in activity was striking. Quantitation of the specific activity of total cell-extract proteins by the DNS-CMCase assay revealed an approx. 40-fold increase in expression in an induced JM101[pTZP-cenC]
FIGURE 30. Congo red plate assay. Four different JM101 clones were analyzed for their relative CMCase activities. The clones contained the following plasmid constructs (from top to bottom): pTZ18R (no insert), pTZ18R-8, pTZ18R-8/5 and pTZP-cenC. The colonies were grown overnight at 30°C on LB plates supplemented with ampicillin and IPTG.
culture versus an induced culture of JM101[pTZ18R-8] (Table VII).

Table VIII illustrates the toxic effect of induced expression of cenC on JM101[pTZP-cenC] cells in culture. 5 hours after addition of IPTG the number of cells had only doubled while the viability of the cells had dropped below 5%. Cells in stationary phase after 24 hours incubation did not express CMCase activity in the Congo red plate assay, and were sensitive to ampicillin, suggesting that they had lost plasmid pTZP-cenC. The cells were sensitive to phage T4, proving them to be E.coli.

The time course of expression of cenC under induced and non-induced conditions is shown in Fig.31. Overproduction of CenC clearly inhibited the growth of the cells. The induced culture reached its highest cell density after 2 1/2 hours whereas the control culture was still in the exponential phase. The maximum specific CMCase activity was reached after 1 1/2 hours and only slightly decreased thereafter, indicating that this culture ceased to produce CenC after 1 1/2 hours and that the recombinant cellulase remained fairly stable.

In addition, under the phase contrast microscope we noted a distinct difference in size and shape of induced versus uninduced cells. We thought to detect densely packed, granular material in the cytoplasm of these cells when induced with IPTG. A similar phenomenon was observed with a clone overexpressing the exoglucanase gene, cex, in E.coli (O'Neill et al., 1986b). A simple protocol for the purification of these Cex granules together with efficient expression of cex promised to provide an optimal system for the overproduction and isolation of the recombinant exoglucanase. However, only drastic measures could partially re-dissolve the Cex aggregates and the recovered activity was marginal. Electron microscopic analysis of
TABLE VII. Endoglucanase activities of various cenC clones. The cells were grown in LB plus 100µg of ampicillin/ml at 30°C to an OD<sub>600</sub> of 1.0. IPTG was added to a final concentration of 1mM where indicated, and incubation continued for another 1½h. The soluble cell extract proteins were isolated by passing the cells through a French press, and removing insoluble material by centrifugation. The cell extracts were assayed for CMCase activities by the DNS-CMCase assay. 1 unit corresponds to 1 µmole glucose equivalents produced per minute.

<table>
<thead>
<tr>
<th>clone</th>
<th>IPTG (1mM)</th>
<th>specific activity (u/mg protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JM101[pTZ18R-8]</td>
<td>+</td>
<td>0.008</td>
</tr>
<tr>
<td>JM101[pTZP-cenC]</td>
<td>-</td>
<td>0.030</td>
</tr>
<tr>
<td>JM101[pTZP-cenC]</td>
<td>+</td>
<td>0.350</td>
</tr>
</tbody>
</table>
TABLE VIII. Effect of IPTG on growth and viability of JM101[pTZP-cenC]. 50 ml volumes of culture medium containing 100 μg ampicillin/ml were inoculated with JM101[pTZP-cenC] from a glycerol stock. The growth of the cultures were monitored by measuring the optical density at 600nm. Numbers of cells ( #cells ) were counted in a Petroff-Hauser cytometer. To evaluate the CMCase activity of cells in the cultures, a number of single colonies were screened by the Congo red plate assay. The viability of the cells in the culture was determined by comparing the number of cells ( #cells ) with their ability to form colonies on LB, ampicillin plates.

<table>
<thead>
<tr>
<th>culture-time in hrs.</th>
<th>OD(600nm)</th>
<th>#cells / ml culture</th>
<th>% total Amp&lt;sup&gt;r&lt;/sup&gt; colonies with haloes</th>
<th>% viable cells on LB, Amp</th>
</tr>
</thead>
<tbody>
<tr>
<td>JM101[pTZP-cenC] induced with IPTG (1mM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.04</td>
<td>3.5 x 10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>67</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>0.24</td>
<td>4.0 x 10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>63</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>0.24</td>
<td>8.0 x 10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>60</td>
<td>&lt;5</td>
</tr>
<tr>
<td>24</td>
<td>6.70</td>
<td>1.3 x 10&lt;sup&gt;10&lt;/sup&gt;</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>JM101[pTZP-cenC] uninduced (control)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.04</td>
<td>3.2 x 10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>73</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>0.45</td>
<td>2.1 x 10&lt;sup&gt;8&lt;/sup&gt;</td>
<td>80</td>
<td>nd</td>
</tr>
<tr>
<td>5</td>
<td>5.70</td>
<td>4.2 x 10&lt;sup&gt;9&lt;/sup&gt;</td>
<td>82</td>
<td>nd</td>
</tr>
<tr>
<td>24</td>
<td>6.90</td>
<td>1.6 x 10&lt;sup&gt;10&lt;/sup&gt;</td>
<td>85</td>
<td>&gt;90</td>
</tr>
</tbody>
</table>
FIGURE 31. Growth and activity profiles of JM101[pTZP-cenC]. At 0 hours, IPTG was either omitted or added (1mM final concentration) to a culture of JM101[pTZP-cenC] at an OD600 of 0.8. The cells were incubated at 30°C in a rotary waterbath. During a culture period of 5 hours, samples were taken to measure the OD600 (o-o, plus IPTG; x-x, no IPTG) and to determine the specific activities in the total cell extracts (stippled bars, plus IPTG; cross-hatched bars, no IPTG). CMCase activity was determined in the DNS-CMCase assay. Units (u) are in μmoles glucose equivalents produced per minute.
JM101[pTZP-cenC] also suggested the intracellular accumulation of CenC (Fig.32). The recombinant material seemed to aggregate around the chromatin (see arrows) causing the chromosome to be compressed. The cell growth studies together with the data from the electron microscopic analysis suggest that intracellular CenC binds to and aggregates at the chromosome, leading to interruption of cell division and, ultimately, to cell death. This would explain why initial cloning experiments of cenC based on expression of activity failed to produce highly active clones.

*E. coli* is not regarded as an optimal system for high-level expression of secreted proteins. The only natural protein secreted into the medium by *E. coli* is hemolysin (Holland et al., 1986). I was interested in seeing how efficiently JM101[pTZP-cenC] processed and transported the recombinant cellulase into the periplasm. Table IX summarizes the CMCase activity profile of CenC in different cellular compartments. The activities and distributions of periplasmic and cytoplasmic marker enzymes (β-lactamase and glucose-6-P-DH, respectively) were used to detect cross-contamination of the two fractions. The fully induced cells seemed to be more fragile since the preparation of these cells for the osmotic shock treatment resulted in a 50% loss of β-lactamase activity. However, an equivalent loss of CMCase activity was not observed. The high molecular weight of CenC together with its potential to form aggregates, thereby getting trapped in the periplasmic space, could explain the discrepancy. Unlike the secretion of exoglucanase into the culture medium by a "leaky" mutant (Gilkes et al., 1984c), hardly any CMCase activity could be recovered from JM101[pTZP-cenC] culture medium. The intracellular distribution (cytoplasm versus periplasm) of cellulase activity indicated the limited capacity of JM101[pTZP-cenC] to transport CenC across the plasma membrane (see control culture in Table IX). Regardless of the induction state of these
FIGURE 32. Electron micrographs of JM101[pTZP-cenC] cells. Cultures of JM101[pTZP-cenC] were either induced with IPTG (panel B) or grown under repressed conditions (panel A). After embedding the cells in Epon 812 epoxy resin, the cross-sections were prepared by ultramicrotomy. The magnification was 20,000-fold.
TABLE IX. Localization of CenC-activity in JM101[pTZP-cenC]. Cells were grown in the presence or absence of IPTG. Activity of CenC was determined in the DNS-CMCase assay. a: numbers in u/ml cell culture; lunit represents 1nmole products formed (CMC-reducing ends, nitrocefoic acid, NADH₂) per minute. b: recovery of activity (%) with respect to total cell extract.

<table>
<thead>
<tr>
<th>cell culture</th>
<th>enzyme assayed</th>
<th>total cell extract</th>
<th>periplasmic fraction</th>
<th>cytoplasmic fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>JM101[pTZP-cenC] plus IPTG (1 mM)</td>
<td>CenC</td>
<td>123\textsuperscript{a} \textsuperscript{b} (100)</td>
<td>72 (59)</td>
<td>51 (41)</td>
</tr>
<tr>
<td></td>
<td>β-lactamase</td>
<td>132 (100)</td>
<td>54 (41)</td>
<td>10 (8)</td>
</tr>
<tr>
<td></td>
<td>glucose-6-P-DH</td>
<td>4.19 (100)</td>
<td>0.40 (9)</td>
<td>2.71 (65)</td>
</tr>
<tr>
<td>JM101[pTZP-cenC] no IPTG</td>
<td>CenC</td>
<td>7.7 (100)</td>
<td>4.5 (58)</td>
<td>2.5 (32)</td>
</tr>
<tr>
<td></td>
<td>β-lactamase</td>
<td>201 (100)</td>
<td>142 (71)</td>
<td>15 (8)</td>
</tr>
<tr>
<td></td>
<td>glucose-6-P-DH</td>
<td>4.84 (100)</td>
<td>0.54 (11)</td>
<td>3.19 (66)</td>
</tr>
</tbody>
</table>
cells, approx. 60% of the total activity was accumulated in the periplasmic compartment.

The intracellular distribution of the recombinant cellulase was examined further by SDS-PAGE (Fig. 33). The appearance of a dominant ( >10% of total cellular protein ), slow migrating protein band in the cytoplasm correlated with the overexpression of *cenC*. Its apparent molecular weight, however, was ~10kDa larger than for C3.1. Since leader peptide processing of soluble secretory proteins is coupled with their translocation across the plasma membrane, this finding suggests the cytoplasmic accumulation of unprocessed CenC. Protein bands with similar molecular weights as C3.1 or C3.2 were not detected in the periplasmic fraction (lane 2c). The electrophoresis data clearly disagree with the intracellular distribution pattern of CMCase activity; a finding which could be explained by postulating that unprocessed CenC has a considerably lower specific activity than the exported cellulase.

III.5. Initial characterization of purified, recombinant CenC

III.5.1. Purification of recombinant CenC1 and CenC2

Prior to the detailed characterization of recombinant CenC, it was necessary to devise a simple purification scheme. Table X describes the protocol chosen for the purification and lists the flow-chart data. Since the larger portion of the total CMCase activity resided in the periplasmic fraction, the osmotic shockate of a JM101[pTZP-cenC] culture was used as starting material. The first purification step employed immunoadsorption chromatography on an affigel-10-rabbitsC3.2 column. Preliminary data indicated
FIGURE 33. SDS-PAGE analysis of protein samples from different compartments of JM101[pTZP-cenC]. Cells were grown either in the absence (lanes 1a, b, c) or presence (lanes 2a, b, c) of 1mM IPTG. Whole cell extract proteins (lanes 1a, 2a), cytoplasmic proteins (lanes 1b, 2b) or periplasmic proteins (lanes 1c, 2c) were electrophoresed. All lanes were loaded with amounts of proteins standardized to the culture volume, except of lane 2c (5-fold excess in order to detect high molecular weight proteins). Lanes 3 and 4: aliquotes of purified C3.1 and C3.2.
TABLE X. Flow-chart of the purification of CenC1 and CenC2. The starting material for the purification was the osmotic shockate (1 litre) of a 5 litre culture of JM101[pTZP-cenC] grown at 30 °C to OD(600nm) of 1.0 and subsequently induced with 1mM IPTG for 1 1/2 hours. a: units in μ moles glucose equivalents produced per minute. Total activity: activity in units recovered after each purification step.

<table>
<thead>
<tr>
<th>purification steps</th>
<th>total protein (mg)</th>
<th>activity a (μ/ml)</th>
<th>spec. activity (μ/mg prot.)</th>
<th>total activity</th>
<th>recovery of activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>osm. shockate (conc.)</td>
<td>74.2</td>
<td>1.79</td>
<td>1.3</td>
<td>95.0</td>
<td>100</td>
</tr>
<tr>
<td>affigel-10 α C3.2 (1M NaSCN)</td>
<td>0.66</td>
<td>5.15</td>
<td>51.5</td>
<td>34.0</td>
<td>36</td>
</tr>
<tr>
<td>MonoQ ngCenC1</td>
<td>0.050</td>
<td>3.53</td>
<td>278.0</td>
<td>13.9</td>
<td>15</td>
</tr>
<tr>
<td>MonoQ ngCenC2</td>
<td>0.025</td>
<td>2.85</td>
<td>390.0</td>
<td>10.1</td>
<td>11</td>
</tr>
</tbody>
</table>
that CenC was sensitive to low pH (glycine-buffer, pH=2.5) and high salt concentrations (NaSCN > 1.5M). To avoid inactivation of CenC, 1.0M NaSCN was chosen as the eluant, although the recovery of activity was rather low (36%). By analogy with the purification of C3.1 and C3.2, the CenC activity was further purified by MonoQ chromatography. We were surprised to find a similar elution profile as for the native enzymes: two closely spaced elution peaks with CMCase activity. The values for the specific activity of the two preparations, designated as CenC1 and CenC2, were comparable with the respective numbers for C3.1 and C3.2 (see Tables III and X).

The purity of the CenC1 and CenC2 preparations was examined by SDS-PAGE (Fig.34). After the affinity chromatography step, two high molecular weight protein bands were detected, which co-migrated with C3.1 and C3.2. These two bands (CenC1 and CenC2) presumably represented the processed isozymes of the cytoplasmic, recombinant CenC. Lanes d and e were overloaded with samples of the MonoQ purified CenC1 and CenC2 preparations to verify their purity.

III.5.2. Amino acid sequence analysis of CenC1 and CenC2

The purity of the two recombinant cellulase preparations allowed their N-terminal amino acid sequences to be analyzed. 10 PTH-amino acid cycles were performed for each sample (data not shown). The results proved that CenC1 and CenC2 had the same amino-terminus. Furthermore, the sequence corresponded to the amino-termini of C3.1 and C3.2. From these data it can be concluded that E.coli not only exported the two recombinant isozymes but also recognized the same leader-peptidase cleavage site as C.fimi.

The question remains as to how one gene can encode two proteins
FIGURE 34. SDS-PAGE analysis of CenC1 and CenC2. Fractions were analyzed at various stages during the purification of the recombinant cellulases. Lane a: total cell extract proteins from induced JM101[pTZP-cenC] cells; lane b: periplasmic fraction (osmotic shockate); lane c: eluate from the affigel-10-rabbitαC3.2 column; lanes d and e: CenC1 and CenC2 preparations after MonoQ-FPLC chromatography; lanes f and g: purified C3.1 and C3.2, respectively. The gel was stained with Coomassie blue.
with different electrophoretic mobilities. Differential glycosylation can be ruled out since *E. coli* does not glycosylate exported proteins. In addition, the similar mobilities of the corresponding native and recombinant cellulases gave further evidence for the absence of glycosylation of the native enzymes by *C. fimi*. It would appear unlikely that proteolytic cleavage in the two species was responsible for the generation of these isozymes since prolonged incubation did not bring about a shift from the higher to the lower molecular weight forms. The same argument rules out auto-proteolytic cleavage. However, it has been reported that cellulase genes from *Schizopyllum commune* give rise to multiple transcripts of different sizes (Willick and Seligy, 1985). A similar phenomenon in our system would explain the unchanging protein ratio (high to low molecular weight) seen in the various cellulase preparations. If indeed *cenC* encoded two different size classes of transcripts, then the 3'-ends of the shorter transcripts would not have been detected in my S1 nuclease studies because the hybridization probe was too short. Unfortunately, the resolution of RNA species by Northern transfer analysis is generally insufficient for the separation of RNA fragments in the size range of 3.5kb which differing by only 250b. The electrophoretic analysis of *in vitro* transcription and translation products encoded by pTZP-cenC could be used to examine this question. In this way the involvement of the host organism's post-translational processing system could be ruled out.

III.6. Final remarks

The endoglucanases, C3.1 and C3.2, are not the main components of the cellulolytic system secreted by *C. fimi*. Furthermore, unlike CenA and Cex, the two cellulases are not substrate-associated when *C. fimi* is grown on Avicel, suggesting that the natural substrates
for C3.1 and C3.2 are most likely soluble cellodextrans which are released from insoluble cellulose by the action of cellulose-associated cellulases.

C3.1 and C3.2 are closely related to each other. Both enzymes were purified by the same purification scheme (Fig.5). Their $M_r$s on SDS-PAGE were similar (130,000 for C3.1 and 120,000 for C3.2, see Fig.6). Amino acid sequence analyses showed that the enzymes share the same N-terminal sequence (Table IV). Furthermore, the values for the kinetic parameters $K_m$ and $V_{max}$, determined by a preliminary kinetic analysis on pNPC as substrate, provided more evidence for their relatedness (the $K_m$ was 0.13mM for both enzymes, and the $V_{max}$ values, in µmoles per min per mg protein, were 0.39 and 0.50 for C3.1 and C3.2, respectively).

A C.fimi DNA-lambda library was constructed for the cloning of the gene(s) encoding the C3 enzymes. Recombinant phage particles were screened either for expression of CMCase activity by the Congo red plate assay or for hybridization with a selection of synthetic oligonucleotide probes by a plaque-filter lift hybridization assay. It was surprising that the activity screening assay did not detect any of the cenC-positive clones determined by cross-screening with the C3-specific oligonucleotide probes (Table VI). It is possible that the N-termini of the C3 enzymes are crucial for the expression of CMCase activity. In this case, most truncated fusion proteins would be inactive. In contrast, several CMCase active clones contained cenA sequences. Following the same line of reasoning, this finding can be explained by the data of the analysis from deletion mutants which indicated that the active site of CenA resided in the C-terminal portion of the cellulase. In addition, computer analysis of the sequence upstream of the cenC start codon revealed several palindromic sequences which might interfere with the expression of transcriptional fusions by acting as transcriptional terminators. Also, we know that the transcriptional regulatory sequences of C.fimi are
only poorly recognized by *E. coli* (Wong et al., 1986a, Owolabi, 1988). Consequently, recombinant lambda clones which were producing low numbers of fusion transcripts would not be detected by the Congo red plate assay. Finally, in the light of the findings from the CenC overproducing clone (JM101[pTZP-cenC], see below), it is feasible that the production of CenC by recombinant lambda clones interfered with the propagation of phage particles and the formation of lambda plaques and, consequently, the corresponding clones could not be identified by the activity screening assay.

The lambda clone, L47.1-169, hybridized with both oligonucleotide probe pools specific for the N-terminal and internal sequences of C3 (J0.3A and BM.C) and was chosen for further characterization (Table VI and Fig.11). The subclones, JM101[pTZ18R-8] and JM101[pTZ18R-8/5] (Fig.15), containing a 7.7kb and a 4.3kb *C.fimi* DNA insert, respectively, only marginally expressed CMCase activity as determined by the DNS-CMCase assay (Table VII; data not shown for JM101[pTZ18R-8/5]).

The location and boundaries of *cenC* in pTZ18R-8 were defined by analyses of *in vivo* *C.fimi* transcripts. RNA-DNA hybrid protection analysis using S1 nuclease led to the identification of the 5'- and 3'-ends of *cenC* transcripts and the location of the corresponding sites in the insert of pTZ18R-8 (Figs.24 and 27). The sizes of the transcripts were in the range of 3.4kb. In addition, by comparison with known bacterial promoters, putative transcriptional regulatory sequences (-10 and -35 hexamers) were identified (Fig.24). It is intriguing that the postulated *C.fimi* promoter sequences for *cex*, *cenA* and *cenB* (Greenberg et al., 1987a, Greenberg et al., 1987b) and *cenC* do not reveal consensus sequences or striking similarities. It should be emphasized, however, that no definite conclusions can be drawn from these data until the validity of these promoters has been confirmed by functional analyses (either by binding studies with purified
C. fimi RNA polymerase preparations or by C. fimi transformation experiments using an E. coli shuttle vector).

The replacement of the C. fimi transcriptional and translational regulatory sequences with an E. coli promoter and ribosomal binding site considerably improved the expression of cenC. The entire coding sequence of cenC except for the GTG start codon was fused in frame to the ATG of a synthetic translation initiation sequence (PTIS). The transcription was placed under the control of the lac p/o system. The resulting construct, JM101[pTZP-cenC], led to the overproduction and cytoplasmic accumulation of CenC (>10% of total cytoplasmic proteins). Only a small fraction of the recombinant enzyme was exported into the periplasm of E. coli. Fractionation of the cells and analysis of the fractions by SDS-PAGE revealed a dominant band with a Mr of approx. 140,000 in the cytoplasmic fraction but a corresponding protein was not detected in the periplasmic fraction (Fig.33). Conversely, 60% of the total CMCase activity was localized in the periplasm (Table IX). Consequently, the specific activity of the cytoplasmic CenC was estimated to be several orders of magnitude lower than the specific activity of the periplasmic enzyme(s). Since the estimated molecular weight of the cytoplasmic CenC was larger than the molecular weight of C3.1, I concluded that the recombinant protein in the cytoplasm represented the unprocessed form of CenC. The difference in the Mr's of the cytoplasmic enzyme and CenCl (approx. 10,000) would account for more than the leader peptide of the CenC enzymes (the Mr of the 32 amino acids leader peptide is 3116). However, I should emphasize that the Mr of the cytoplasmic CenC is only a rough estimate (see Fig. 32). Proteolytic processing of the leader peptide and/or transport across the cytoplasmic membrane seem to be a prerequisite for expression of activity of some periplasmic enzymes in E. coli. However, it is generally accepted that E. coli is not a good host for the efficient export of recombinant proteins. Certain
constructs expressing cex and cenA also led to the intracellular accumulation of recombinant proteins (O'Neill et al., 1986b, Guo. Z.M., personal communication). It was interesting to note that the overproduction of CenC in JM101[pTZP-cenC] interfered with cell division and led ultimately to cell death (Table VIII). The growth rate of the cells was slowed down even when the cells were grown under repressed conditions (no IPTG added to the cultures). Apparently, even low level expression of cenC caused an inhibitory effect on cell growth.

The fractionation of CMCase activities from the periplasmic compartment of JM101[pTZP-cenC] led to the purification of two recombinant cellulases, CenCl and CenC2. Their molecular weights, as estimated by SDS-PAGE, correlated with the values for C3.1 and C3.2 (Fig.34). In addition, the specific activities of the native and recombinant enzymes with CMC as substrate were comparable (Tables III and X). N-terminal sequence analysis of the CenC enzymes not only proved that CenCl and CenC2 share the same N-termini as the C3 enzymes but also showed that the E.coli leader peptidase recognized the same leader peptidase cleavage site as the corresponding C.fimi enzyme. At this stage it is not clear what the mechanisms were which were responsible for the one gene-two polypeptides phenomenon. The preliminary biochemical data for CenCl and CenC2 rule out post-translational modifications such as differential glycosylation and/or proteolytic processing as an explanation for the formation of these isozymes. Analysis of in vitro cenC transcripts and their products will help to clarify this question.

The construction of the clone JM101[pTZP-cenC] provided a system for the analysis of the recombinant cellulases devoid of other contaminating cellulolytic activities. This clone will allow the purification of sufficient quantities of CenCl and CenC2 for their further characterization, such as the determination of the kinetic parameters and the substrate specificities.

A detailed characterization of the recombinant cellulases
(C. fimi protease studies (Gilkes et al., 1988, Langsford et al., 1987, Arfman et al., 1987), analysis of synergistic properties with other cloned cellulases, crystallography, etc.) requires a better source for the isolation of large quantities of CenC1/2. As described above, the production of processed CenC1/2 by JM101[pTZP-cenC] appears to be limited due to the toxicity of the intracellular recombinant products. For the overexpression of the mature enzymes I suggest that the coding sequence of cenC (lacking the leader sequence) be fused to PTIS using a strategy analogous to the construction of pTZP-cenC. Since the overproduction of CenC did not lead to the formation of insoluble aggregates in the cytoplasm of these clones, such a construct might provide a suitable system for the isolation of large quantities of CenC cellulases.
IV. REFERENCES


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


Teeri, T.T. 1987. In:"The cellulolytic enzyme system of *Trichoderma reesei". Technical Research Center of Finland, Publications 38


United States Biochemical Corporation (UBS). Genescribe-Z™: description and experimental protocols. P.O. Box 22400, Cleveland, Ohio 44122


V. APPENDIX

Flow-chart protocol for the molecular cloning of cenC and the construction of the cenC high-expression vector, pTZP-cenC

1. Purification of the *C. fimi* endo-glucanases, C3.1 and C3.2
2. Determination of the amino acid sequences of the N-termini of C3.1 and C3.2 and the C3.2-internal, tryptic peptide, T-115
3. Synthesis of oligonucleotide probes corresponding to the N-termini of the C3 enzymes (J0.3A,B,C, and D) and to T-115 (EM.C)
4. Construction of a *C. fimi* DNA-lambda library by cloning genomic *C. fimi* DNA into the lambda vector L47.1
5. Screening of recombinant lambda phage particles by plaque-filter lift hybridization using the radio-labeled J0.3 oligonucleotide probes
purification of J0.3A positive recombinant phage particles and rescreening with the radiolabeled oligonucleotide probes, J0.3A and BM.C

selection of the recombinant phage particle, L47.1-I69, for further studies

localization of cenC in the insert of L47.1-I69 by DNA dot-blot analysis using the radiolabeled probes, J0.3A and BM.C

subcloning of the cenC-containing, 7.7 kb HindIII/BamHI fragment into the plasmid vector pTZ18R resulting in the subclone pTZ18R-8

localization of the 5'-end of cenC in the insert of pTZ18R-8 by Southern transfer experiments

nucleotide sequence determination of the 900 bp SstI/BamHI fragment containing the 5'-end of cenC

determination of the boundaries of cenC in pTZ18R-8 and identification of the transcriptional regulatory sequences by RNA-DNA hybrid protection analysis
subcloning of the 900 bp SstI/BamHI fragment containing the 5'-end of cenC into pTZ18R resulting in pTZ18R-S/B

synthesis of the 5'-end of cenC starting with second codon of the cenC leader sequence by primer extension using the synthetic oligonucleotide primer-8/5

insertion of the synthetic translation initiation sequence (PTIS) into the EcoRI and BamHI sites of pTZ18R resulting in pTZ18R-PTIS

fusion of the synthesized 5'-end fragment of cenC with the ATG start codon of the PTIS in pTZ18R-PTIS resulting in the subclone pTZ18R-5'

construction of pTZP-cenC by cloning the 3.4 kb BamHI/HindIII fragment of pTZ18R-8 containing the missing 3'-sequence of cenC into pTZ18R-5'

over-expression of cenC by the E.coli clone JM101[pTZP-cenC]