

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

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

by

**CHUNG-MING HUANG**

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

1990

# TABLE OF CONTENTS

ABSTRACT . . . . .	i
ACKNOWLEDGEMENTS . . . . .	iii
LIST OF PUBLICATION . . . . .	iv
LIST OF FIGURES . . . . .	v
LIST OF TABLES . . . . .	vii

## Chapter I Introduction

1-1 Introduction . . . . .	1-1
1-2 Cellulolytic microorganisms . . . . .	1-3
1-2-1 Fungi . . . . .	1-3
1-2-2 Bacteria . . . . .	1-4
1-3 Biochemical studies . . . . .	1-9
1-3-1 Cellulose structure . . . . .	1-9
1-3-2 Cellulolytic systems . . . . .	1-10
1-4 Genetic studies of cellulase genes . . . . .	1-19
1-4-1 General concept of gene manipulation . . . . .	1-19
1-4-2 Genetic manipulation in cellulase genes . . . . .	1-24
1-4-3 Cellulase gene structure . . . . .	1-26
1-5 Aim of this study . . . . .	1-30

## CHAPTER II MATERIALS AND METHODS

2-1 Bacterial strains and vectors . . . . .	2-1
2-2 Media and growth conditions . . . . .	2-2
2-3 Enzymes and chemicals . . . . .	2-2
2-4 Molecular biological methods . . . . .	2-3
2-4-1 Gel electrophoresis . . . . .	2-3
2-4-1-1 Agarose gel for DNA . . . . .	2-3
2-4-1-2 Polyacrylamide gel for protein . . . . .	2-3
2-4-1-3 Polyacrylamide gel for DNA sequencing . . . . .	2-5
2-4-1-4 Preparation and assembly of glass plates . . . . .	2-5
2-4-2 Preparation of DNA . . . . .	2-6
2-4-2-1 Chromosomal DNA . . . . .	2-6
2-4-2-2 Phage DNA . . . . .	2-6
2-4-2-3 Plasmid DNA . . . . .	2-8
2-4-3 Extraction of DNA from agarose gel . . . . .	2-9
2-4-3-1 Electroelution method . . . . .	2-9
2-4-3-2 DEAE membrane binding method . . . . .	2-10
2-4-4 Restriction digestion . . . . .	2-10
2-4-4-1 Dephosphorylation of DNA . . . . .	2-11
2-4-5 Construction of recombinant phage library . . . . .	2-11

2-4-6	Screening of recombinant for cellulase activities . . . . .	2-12
2-4-7	Hybridization . . . . .	2-12
2-4-7-1	Denatured salmon sperm DNA . . . . .	2-13
2-4-7-2	Preparation of probe . . . . .	2-13
2-4-7-3	Dot hybridization . . . . .	2-13
2-4-7-4	Immunological visualization . . . . .	2-14
2-4-8	Subcloning of insert fragment . . . . .	2-14
2-4-8-1	Transformation . . . . .	2-14
2-4-8-2	<i>Bacillus</i> transformation (protoplast transformation) . . . . .	2-15
2-4-8-3	Transfection . . . . .	2-16
2-4-9	DNA deletion . . . . .	2-16
2-4-10	DNA sequencing analysis . . . . .	2-18
2-4-10-1	Assembly and analysis of sequence data . . . . .	2-18
2-5	Biochemical methods . . . . .	2-19
2-5-1	Partial purification of cellulase enzymes . . . . .	2-19
2-5-1-1	Preparation of crude enzyme extract . . . . .	2-19
2-5-1-2	Gel filtration chromatography . . . . .	2-19
2-5-2	Biochemical assay of cellulase activities . . . . .	2-20
2-5-2-1	Assay of endoglucanase activity . . . . .	2-20
2-5-2-2	Assay of exoglucanase and $\beta$ -glucosidase activities . . . . .	2-20
2-5-2-3	Protein concentration . . . . .	2-21
2-5-2-4	Substrate specificity assay . . . . .	2-21
2-5-2-5	Inhibition and activation assay . . . . .	2-21
2-5-3	Cellular localization of enzymes in <i>E. coli</i> . . . . .	2-22
INDEX	. . . . .	2-23

### CHAPTER III CONSTRUCTION OF GENOMIC LIBRARY AND IDENTIFICATION OF CELLULASE GENE FROM *RUMINOCOCCUS FLAVEFACIENS*

3-1	Introduction . . . . .	3-1
3-2	Results . . . . .	3-3
3-2-1	Construction of genomic library . . . . .	3-3
3-2-1-1	Preparation of partially digested chromosomal DNA . . . . .	3-3
3-2-1-2	Ligation with vector DNA and packaging with lambda packaging extract . . . . .	3-4
3-2-2	The genomic library of <i>R. flavefaciens</i> . . . . .	3-5
3-2-3	Screening the library for cellulase genes expressed in <i>E. coli</i> . . . . .	3-6
3-2-4	Biochemical assay of cellulase activities . . . . .	3-10
3-2-5	The restriction mapping of recombinant clones . . . . .	3-12
3-3	Discussion . . . . .	3-12
3-3-1	Library construction . . . . .	3-12
3-3-2	High copy number of cellulase genes and cellulase families . . . . .	3-14
3-4	Short summary . . . . .	3-15

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

4-1	Introduction . . . . .	4-1
4-2	Results . . . . .	4-3
4-2-1	Localization of cellulase genes in $\lambda$ CM903 . . . . .	4-3

4-2-1-1	The difficulties of subcloning the cellulase genes of $\lambda$ CM903 into plasmid vectors . . . . .	4-3
4-2-1-2	Localization of CMCase and MUCase genes in $\lambda$ CM903 . . . . .	4-3
4-2-2	Homology studies of $\lambda$ CM- and <i>celA</i> gene of <i>C. thermocellum</i> . . . . .	4-5
4-2-3	Localization of enzymes in <i>E. coli</i> cells . . . . .	4-5
4-2-4	ExoIII deletion of $\lambda$ CM903 . . . . .	4-6
4-2-5	Examination of endoglucanase, exoglucanase and $\beta$ -glucosidase gene expression . . . . .	4-6
4-2-6	Partial purification of cellulases from $\lambda$ CM903 . . . . .	4-6
4-2-7	Characterization of cellulase gene products . . . . .	4-9
4-2-7-1	Endoglucanase . . . . .	4-9
4-2-7-2	Exoglucanase . . . . .	4-11
4-2-7-3	$\beta$ -Glucosidase . . . . .	4-11
4-2-8	Substrate specificity assay . . . . .	4-13
4-3	Discussion . . . . .	4-13
4-3-1	Substrate degradation . . . . .	4-13
4-3-2	Plasmid instability . . . . .	4-15
4-3-3	Enzyme expression . . . . .	4-17
4-3-4	Choice of deletion method . . . . .	4-18
4-3-5	Enzyme purification and characterization of gene products . . . . .	4-18
4-4	Short summary . . . . .	4-19
Chapter V Nucleotide sequencing and analyses of a 7.3 kb cellulase gene fragment		
5-1	Introduction . . . . .	5-1
5-2	Results . . . . .	5-1
5-2-1	Nucleotide sequence results . . . . .	5-1
5-2-3-1	First open reading frame . . . . .	5-2
5-2-3-2	Second open reading frame . . . . .	5-15
5-2-3-3	Third open reading frame . . . . .	5-20
5-2-3-4	Fourth open reading frame . . . . .	5-23
5-2-3-5	Fifth open reading frame . . . . .	5-26
5-3	Discussion . . . . .	5-29
5-4	Short summary . . . . .	5-32
Chapter VI FINAL DISCUSSION AND CONCLUSIONS		
6-1	Use of the <i>E. coli</i> lambda system in the study of rumen microorganism . . . . .	6-1
6-2	The cellulase of ruminococci . . . . .	6-2
6-3	The cellulase gene system in ruminococci . . . . .	6-4
6-4	Applications of cellulases and cellulase genes . . . . .	6-10
6-5	Study of cellulase genes of rumen microorganisms . . . . .	6-10
6-6	Conclusion . . . . .	6-12
BIBLIOGRAPHY . . . . .		B-1

## ABSTRACT

A genomic library of *Ruminococcus flavefaciens* 186 was constructed using the lambda vector  $\lambda$ NM1149. It constituted  $2.1 \times 10^4$  recombinant clones which was large enough to represent the entire genome of this bacterium. From this library, 26 CMC<sup>+</sup> clones were identified after screening about 2500 recombinant clones. These CMC<sup>+</sup> clones were divided into four groups according to their insertion fragment size. Partial restriction maps of these clones have been achieved. Eight representative clones from these four groups showed different mapping patterns. One of the three 9 kb insert clones ( $\lambda$ CM903) was selected for further study based on its ability to express all three types of cellulase activities.

The locations of endoglucanase and exoglucanase genes in  $\lambda$ CM903 were determined. Two separate fragments from  $\lambda$ CM903 were subcloned and named  $\lambda$ CMEH1 (CMC<sup>+</sup>) and  $\lambda$ CMEH2 (MUC<sup>+</sup>). Cross hybridization experiment using *celA* gene of *C. thermocellum* and the 26 recombinant clones showed no significant homology. However, different degrees of homology were found among the 26 ruminococcal clones.

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

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

Using *ExoIII* deletion, the locations of endoglucanase and exoglucanase genes were determined in  $\lambda$ CM903. The internal *HindIII-HindIII* fragment of 7.3 kb from  $\lambda$ CM903 was sequenced. Five ORFs were detected using computer software (UWGCG) analyses.

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

## ACKNOWLEDGEMENTS

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

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

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

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

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

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

## LIST OF PUBLICATIONS

### Publications or abstracts arising from this thesis

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

## LIST OF FIGURES

### Chapter I

- Fig. 1-1. TEM of *R. flavefaciens* attach to Avicel.  
Fig. 1-2. Basic structure of cellulose component.  
Fig. 1-3. Organization of cellulose molecules.  
Fig. 1-4. Diagram of cellulase acting in cellulose chain.  
Fig. 1-5. Competition between endo- and exo-glucanase for adsorption.  
Fig. 1-6. Substrate stereospecificity of two exoglucanase.  
Fig. 1-7. Proposed mechanism for the endoglucanase-catalysed cleavage of  $\beta$ -1,4-linkages in cellulose.  
Fig. 1-8. Hydrolysis of cellulose by bacterial cellulytic enzyme system.  
Fig. 1-9. A general diagram of gene cloning procedure.  
Fig. 1-10. The organization of functional domains in two cellulase families.

### Chapter III

- Fig. 3-1.  $\lambda$ L47 bacteriophage vector.  
Fig. 3-2.  $\lambda$ NM1149 bacteriophage vector.  
Fig. 3-3. Gel electrophoresis of recombinant clones.  
Fig. 3-4. Congo red method for detection of CMCase clones.  
Fig. 3-5. Methylumbelliferone fluorescent method for detection of MUCase and MUGase clones.  
Fig. 3-6. Restriction maps of eight representative  $\lambda$ CM-clones.  
Fig. 3-7. A summary in construction of genomic library of *R. flavefaciens* 186.

### Chapter IV

- Fig. 4-1. A diagram of *ExoIII* deletion method.  
Fig. 4-2. Subclones of  $\lambda$ CM903 and genes location.  
Fig. 4-3. A deletion strategy of  $\lambda$ CM903.  
Fig. 4-4. Gene location of  $\lambda$ CM903.  
Fig. 4-5. Partial purification of cellulase from culture of  $\lambda$ CM903 with serial Bio-Gel P columns.  
Fig. 4-6. A pH profile of cellulases of  $\lambda$ CM903.  
Fig. 4-7. A temperature profile of cellulases of  $\lambda$ CM903.

## Chapter V

- Fig. 5-1. Nucleotide sequence of 7.3 kb fragment of  $\lambda$ CM903.
- Fig. 5-2. Open reading frames identification from 7.3 kb fragment of  $\lambda$ CM903 using "Frames" of UWGCG.
- Fig. 5-3. Nucleotide sequence and deduced amino acid sequence of ORF1.
- Fig. 5-4. G+C bias of 7.3 kb fragment using "Codonpreference" of UWGCG.
- Fig. 5-5. Comparison of ORF1 with *cenA* of *Ce. fimi*.
- Fig. 5-6. Nucleotide sequence and deduced amino acid sequence of ORF2.
- Fig. 5-7. Comparison of ORF2 with other exoglucanase genes.
- Fig. 5-8. Alignment with PTS rich region of ORF1, ORF2 and ORF5.
- Fig. 5-9. Nucleotide sequence and deduced amino acid sequence of ORF3.
- Fig. 5-10. Comparison of ORF3 with other  $\beta$ -glucosidase genes.
- Fig. 5-11. Nucleotide sequence and deduced amino acid sequence of ORF4.
- Fig. 5-12. Comparison of ORF4 with other signal sequences of proteases genes.
- Fig. 5-13. Nucleotide sequence and deduced amino acid sequence of ORF5.
- Fig. 5-14. Comparison of ORF5 with other xylanase genes.

## Chapter VI

- Fig. 6-1 A diagram in application of binding domain of cellulase gene.

## LIST OF TABLES

### Chapter I

- Table 1-1. Cellulase-producing microorganisms.  
Table 1-2. Fermentation characteristics of *F. succinogenes* and *Ruminococcus* sp.  
Table 1-3. Families of cellulases.

### Chapter II

- Table 2-1. Bacterial strains and vectors.  
Table 2-2. Medium for *Ruminococcus*.  
Table 2-3. M9 medium.

### Chapter III

- Table 3-1. Cellulase activities of  $\lambda$ CM clones.

### Chapter IV

- Table 4-1. Cross hybridization of  $\lambda$ CM- clones and *celA* of *C. thermocellum*.  
Table 4-2. Cellular locations of cellulase enzyme in *E. coli*.  
Table 4-3. Summary of partial purification of endoglucanase, exoglucanase and  $\beta$ -glucosidase.  
Table 4-4. Inhibition factors of cellulase enzyme.  
Table 4-5. Specific substrate degradation by  $\lambda$ CM- clones.

### Chapter V

- Table 5-1. Comparison of consensus promoter sequences and SD sequences.  
Table 5-2. Codon frequency of ORF1, ORF2, ORF3, ORF4, ORF5 and other cellulase genes.  
Table 5-3. Summary of homology percentage of *ren* with other cellulase genes.

### Chapter VI

- Table 6-1. Summary of the characterization of gene products and deduced peptides and comparison with other cellulase proteins.  
Table 6-2. Summary of ORFs of 7.3 kb fragment.