INVESTIGATION OF LACTOSE UTILIZATION GENES IN *CLOSTRIDIUM ACETOBUTYLICUM*

by

KERRIE RUTH HANCOCK

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

1988

ABSTRACT

Preparatory to constructing a genomic library of Clostridium acetobutylicum, cell lysis and DNA preparation conditions leading to undegraded DNA were examined. Early to mid-exponential phase cells lysed more readily than cells at late-exponential or stationary phase. Lysis was facilitated by 0.3% w/v glycine in the growth medium. Achromopeptidase was a more effective lytic enzyme than either mutanolysin or lysozyme.

All strains of C. acetobutylicum produced high levels of DNase activity, coinciding with the late-exponential and stationary growth phases. Chromosomal DNA isolated from all strains of C. acetobutylicum was rapidly degraded. This degradation was not prevented by the use of various protein inactivating agents. The adverse effect of oxygen and related radicals on the DNA of this strict anaerobe was considered to be responsible. Undegraded DNA was isolated by protecting the cells to be lysed from oxygen.

A genomic library of C. acetobutylicum NCIB2951 in the cosmid vector pLAFR1, constituting 3,500 recombinant clones, was prepared. Clones from this library complemented various Escherichia coli auxotrophic mutations, showing that C. acetobutylicum genes are expressible in E. coli.

Recombinant clones coding for the ß-galactosidase of NCIB2951 were isolated from the genomic library using the chromogenic substrate X-gal. The lacY mutation of HB101 could not be complemented by these clones, suggesting that a classical lac operon system does not exist in C. acetobutylicum.

The ß-galactosidase (cbg) gene was further subcloned on a 5.2 kb EcoRI fragment, and was expressed when the fragment was cloned in either orientation. Cbg was thus expressed from its own promoter. The cbg gene is inducible by lactose in C. acetobutylicum. When cloned into E. coli, however, this gene was expressed constitutively, the level being unaffected by the presence of the inducer,

IPTG or glucose.

Six strains of C. acetobutylicum possessed a sequence highly homologous to the cloned ß-galactosidase fragment. The ß-galactosidase gene region of NCIB2951 showed only low homology to the DNA from other Gram-positive bacteria (Streptococcus lactis ATCC7962, Streptococcus thermophilus DRI1424, Lactobacillus bulgaricus DRI20056, Lactobacillus helviticus DRI20064), and no detectable homology to DNA from Gram-negative bacteria (E. coli DC272 or Rhizobium loti PN2231).

The ß-galactosidase activity of the 5.2 kb fragment was inactivated by Tn5 insertion at either of two loci. Locus I (400 bp) was approximately 500 bp from locus II (approximately 3.2 kb). Maxi-cell analysis identified a 100 kDa protein as the ß-galactosidase gene product.

The 5.2 kb fragment was sequenced and analyzed. Three ORF's were identified. ORF1 (cbgA) coded for the structural ß-galactosidase gene. Significant amino acid homology was detected with the amino acid sequences of the lacZ, ebgA (E. coli) and lacZ (Klebsiella pneumoniae). ORF2 (cbgR) coded for a small regulatory protein which shared homology with the amino acid sequence of the "0.3 kb gene" from Bacillus subtilis. ORF3 coded for a truncated protein which shared significant homology with the N-terminal amino acid regions of spo0A and spo0F (B. subtilis), two regulatory proteins of the two-component system. Hence, no lac operon exists in C. acetobutylicum.

ACKNOWLEDGEMENTS

I am indebted to my supervisors, Dr Ian Maddox, Dr Lindsay Pearce and Professor Barry Scott for their collective inspiration, encouragement, guidance, advice, and patience during this PhD, and for providing the facilities which enabled this study. Special thanks goes to the following people for their encouragement, advice and friendship which was greatly appreciated.

At New Zealand Dairy Research Institute (NZDRI): Audrey, Graham and Guy for your constant interest and encouragement, thanks. Also Sam and Charlotte.

At Massey: Noemi, Feng Feng, and Sirinda (fellow PhDs at Biotechnology) and Mike, for your friendship.

Everyone at 'Scott Base' and the surrounding area: Caroline, Michelle, Grant, Robert, Gopher, Chris, Eva, Sharon, Trish, Mike, Lawrence, Sally, George, and Leslie for their friendship, sense of humour and for sharing the lab with me (i.e. tolerance). To those involved in the 'conspiracy' concerning my computer and I - the joke was much appreciated!!!

To Eva, thanks for your assistance and tutoring concerning the SDS-PAGE.

Also to 'Garfield' and 'Footrot Flats' for keeping me sane.

I gratefully acknowledge the following grants, without which this work could not have been done: New Zealand Energy Research Development Council, and the New Zealand Dairy Research Institute.

I would also like to thank Dan Legg (NZDRI) and Robert Hickson (Scott Base) for their great patience and help in using the computers to create this thesis. Also thanks to Paul Le Ceve for his expertise with the photos, and Gail Haydock for help with the typing.

To my family, I would like to show my greatest gratitude, for their unending love and support.

Dedicated to my mother

and especially to my late father

LIST OF CONTENTS

Page

ABSTRACT (i)

ACKNOWLEDGEMENTS (iii)

LIST OF CONTENTS (v)

LIST OF FIGURES (xiii)

LIST OF TABLES (xvii)

ABBREVIATIONS (xix)

CHAPTER 1: LITERATURE REVIEW 1

PART I:

1.1 Introduction 1

1.1.1 History 1

1.1.2 Revival of the fermentation process 4

1.2 The Organism 5

1.2.1 Classification 5

1.2.2 Morphological characteristics 5

1.2.3 Cultural characteristics 6

1.2.4 Substrate range and nutritional requirements 7

1.2.5 Life cycle 8

1.2.6 Sporulation 9

Page

1.3 Biochemistry 10

1.3.1 Course of fermentation 10

1.3.2 Metabolic pathways 11

1.3.2.1 Acid producing pathways 11

1.3.2.1 Solvent producing pathways 15

1.3.3 Factors triggering solventogenesis 16

PART II:

1.4 Genetics of C. acetobutylicum 18

1.4.1 Taxonomy of the genus Clostridium 18

1.4.2 Genetics and strain improvement 22

1.4.2.1 Mutagenesis 23

1.4.2.2 Mutagenesis of C. acetobutylicum 23

1.4.3 Plasmids 26

1.4.3.1 Plasmids in C. acetobutylicum 26

1.4.3.2 Construction of shuttle vectors using plasmids from Clostridium species 27

1.4.4 Bacteriophages 29

1.4.5 Genetic transfer systems 30

1.4.5.1 Protoplast regeneration 30

1.4.5.2 Protoplast fusion 32

1.4.5.3 Protoplast transformation 33

1.4.5.4 Conjugation of C. acetobutylicum 34

1.5 Characterization of various C. acetobutylicum genes 36

1.5.1 Introduction 36

1.5.2 Cloning of genes from C. acetobutylicum 37

Page

PART III:

1.6 Lactose utilization 42

1.6.1 The use of whey for the ABE fermentation 42

1.6.2 Whey production in New Zealand 44

1.7 Pathways of lactose utilization in microorganisms 47

1.7.1 Metabolism of lactose in E. coli (the lac operon) 47

1.7.2 Metabolism of lactose by the PEP:PTS system 49

1.7.3 Lactose utilization in C. acetobutylicum 51

1.8 Cloning of lactose utilization genes 52

1.8.1 Lac operon system 52

1.8.2 PEP:PTS system 53

PART IV:

1.9 Aim of the thesis 54

CHAPTER 2: METHODS AND MATERIALS 55

2.1 Bacterial strains and plasmids 55

2.2 Preparation of culture media 59

2.2.1 Liquid media 59

2.2.2 Solid media 60

2.3 Maintenance of cultures 60

2.4 Growth of bacteria 60

2.5 Method of lysis 61

2.5.1 Lysis experiments using lysozyme 61

2.5.2 Lysis experiments using mutanolysin 62

2.5.3 Lysis experiments using achromopeptidase 62

2.6 Batch fermentation run 63

2.6.1 DNase assay 64

		Page
2.7	Isolation of plasmid DNA	64
	2.7.1 Cleared lysate method	65
	2.7.2 Alkaline lysis method	66
	2.7.3 Rapid boiling method	67
2.8	Isolation of genomic DNA	67
	2.8.1 E. coli, Streptococcus, Lactobacillus	67
	2.8.2 C. acetobutylicum	69
2.9	Digestion of DNA with restriction enzymes	70
2.10	Preparation of packaging extracts	71
2.11	Construction of a pLAFR1 gene bank to C. acetobutylicum NCIB2951	72
2.12	Horizontal agarose gel electrophoresis	74
2.13	Determination of molecular weights	74
2.14	Determination of DNA purity	74
2.15	Purification of DNA preparation	74
2.16	Elution of DNA from agarose	75
2.17	Preparation of dialysis tubing	75
2.18	Preparation of [^{32}P]-labelled DNA probes	76
2.19	DNA transfer and hybridization	77
2.20	DNA ligation	78
2.21	Transformation of E. coli	78
2.22	Conjugation	79
2.23	Phospho-ß-galactosidase enzyme assay	79
2.24	ß-galactosidase enzyme assay	80
2.25	Preparation of bacteriophage lambda 467	81
2.26	Mutagenesis of plasmid DNA with transposon Tn5	82
2.27	Maxi-cell analysis	83
2.28	Protein gel analysis	84
2.29	Preparation of Bal 31 deletions	85

		Page
2.30	Large-scale preparation of M13 RF DNA	86
2.31	Preparation and transformation of E. coli cells	87
2.32	Hybridization of M13 plaques	88
2.33	Preparation of recombinant M13 ss DNA templates	88
2.34	Preparation and assembly of glass plates	89
2.35	Pouring and running of sequencing gels	90
2.36	Preparation of the NTP mixes	91
2.37	Preparation of [^{35}S]-labelled M13 templates	92
2.38	Sequencing with Sequenase kit	93
2.39	Formation of the sequence	93

CHAPTER 3: EFFECT OF VARIOUS BACTERIOLYTIC ENZYMES ON THE LYSIS OF C. ACETOBUTYLICUM

3.1	Introduction	94
3.1.1	Bacterial cell walls	94
3.1.2	C. acetobutylicum cell envelope	95
3.1.3	Peptidoglycan	98
3.1.4	Bacteriolytic enzymes	98
3.1.5	Aim	105
3.2	Results	105
3.3	Discussion	106
3.3.1	Effect of glycine of the medium	107
3.3.2	Culture age	107
3.3.3	Peptidoglycan sensitivity and accessibility	109
3.4	Summary	112

CHAPTER 4: PREPARATION OF A GENOMIC LIBRARY OF C. ACETOBUTYLICUM NCIB2951

4.1	Introduction	113
4.1.1	Aim	114

Page

4.2 Results 114

4.2.1 Difficulties in preparation of the genomic library 114

4.2.2 Preparation of the genomic library 120

4.3 Discussion 123

4.3.1 DNase activity 124

4.3.2 Preparation of genomic library 127

4.4 Summary 128

CHAPTER 5:IDENTIFICATION AND CHARACTERIZATION OF LACTOSE UTILIZATION GENES IN C. ACETOBUTYLICUM NCIB2951

5.1 Introduction 129

5.1.1 Lactose utilization of C. acetobutylicum 129

5.1.2 Aim 130

5.2 Results 130

5.2.1 Complementation of HB101 auxotrophic mutations 130

5.2.2 Attempts to isolate the pbg gene from the C. acetobutylicum genomic library 133

5.2.3 Attempts to select for the C. acetobutylicum pbg gene by expression in E. coli 135

5.2.4 Detection of ß-galactosidase gene in C. acetobutylicum 136

5.2.5 Restriction enzyme mapping of pKH107 140

5.2.6 Expression of cbg in E. coli in the presence of different carbon sources. 140

5.2.7 Detection of ß-galactosidase genes in other C. acetobutylicum strains using hybridization 142

5.2.8 Degree of homology of the C. acetobutylicum NCIB2951 cbg gene region with DNA from other bacteria 144

Page

5.3 Discussion 146

5.3.1 Complementation of E. coli auxotrophs with the C. acetobutylicum genomic library 146

5.3.2 Attempts to isolate the C. acetobutylicum pbg gene 147

5.3.3 Cloning of the cbg gene from C. acetobutylicum 148

5.4 Summary 151

CHAPTER 6:TRANSPOSON TN5 SITE-DIRECTED MUTAGENESIS OF C. ACETOBUTYLICUM NCIB2951 ß-GALACTOSIDASE GENE

6.1 Introduction 152

6.1.1 Transposons 152

6.1.2 Aim 153

6.2 Results 154

6.2.1 Site-specific mutagenesis of plasmid pKH107 154

6.2.2 Physical analysis of plasmid derivatives 156

6.2.3 Examination of the ß-galactosidase gene function 165

6.2.4 Maxi-cell analysis 165

6.3 Discussion 168

6.3.1 Tn5 insertional specificity 168

6.3.2 Identification of the two gene loci for ß-galactosidase expression 170

6.3.3 Gene products of the ß-galactosidase gene region 172

6.4 Summary 172

CHAPTER 7:NUCLEOTIDE SEQUENCE AND GENETIC ORGANIZATION OF THE 5.2 KB ECORI ß-GALACTOSIDASE FRAGMENT

7.1 Introduction 173

7.1.1 M13 biology 173

7.1.2 Modification of M13 for sequencing 173

Page

	7.1.3	Generation of insert templates	176
	7.1.4	Aim	177
7.2	Results		177
	7.2.1	Preparation of insert DNA	177
	7.2.2	Nuclease Bal 31 digestion	180
	7.2. 3	Nucleotide sequence results	181
7.3	Discussion		214
	7.3.1	Analysis of the 5.2 kb fragment	215
	7.3.2	Nucleotide content and codon preference analysis	220
	7.3.3	Amino acid content	227
7.4	Summary		229

CHAPTER 8:FINAL DISCUSSION AND CONCLUSIONS 230

BIBLIOGRAPHY 235

LIST OF FIGURES

		Page
1.1	Typical outline of an ABE fermentation	12
1.2	Biochemical pathways in C. acetobutylicum	13
1.3	Phylogenetic relationship between Clostridium	21
1.4	Utilization of waste whey within New Zealand	43
1.5	Pathways of lactose utilization in bacteria	45
1.6	The lactose operon in E. coli	46
3.1	Structure of the cell wall of C. acetobutylicum	96
3.2	The general structure of cell wall peptidoglycan and enzyme sites	97
3.3	Lysis results of C. acetobutylicum cells of various growth phases grown in the absence of glycine using lysozyme	100
3.4	Lysis results of C. acetobutylicum cells of various growth phases grown in the presence of glycine using lysozyme	101
3.5	Lysis results of C. acetobutylicum cells of various growth phases grown in the absence of glycine using mutanolysin	102
3.6	Lysis results of C. acetobutylicum cells of various growth phases grown in the presence of glycine using mutanolysin	103
3.7	Lysis results of C. acetobutylicum cells of various growth phases grown in the absence of glycine using achromopeptidase	104
4.1	Detection of extracellular DNase activities in various strains of C. acetobutylicum	115
4.2	Levels of intracellular and extracellular DNase activity in a C. acetobutylicum fermentation NCIB2951	116
4.3	Analysis by gel electrophoresis of a random selection of cosmid clones from the genomic library of C. acetobutylicum NCIB2951	122

Page

5.1 Analysis by agarose gel electrophoresis of cosmids isolated from the C. acetobutylicum NCIB2951 genomic library which complement the proA2, leuB6, and thi auxotrophic mutation present in E. coli HB101 131

5.2 Hybridization of L. casei pbg gene region to an EcoRI genomic digest of C. acetobutylicum NCIB2951 134

5.3 Restriction enzyme digests of cosmids that expressed ß-galactosidase activity 136

5.4 Restriction enzyme map of the 5.2 kb EcoRI fragment containing the ß-galactosidase gene region of C. acetobutylicum NCIB2951 139

5.5 Hybridization of C. acetobutylicum NCIB2951 ß-galactosidase gene regions from pKH107 to EcoRI genomic digests of other C. acetobutylicum strains 143

5.6 Hybridization of the 5.2 kb EcoRI fragment containing the ß-galactosidase gene region of C. acetobutylicum NCIB2951 to genomic and plasmid DNA of other Gram-positive and Gram-negative bacteria 145

6.1 Agarose gel electrophoresis of EcoRI digests of Tn5-mutagenized pKH107 derivatives 155

6.2 Diagram of a generalized pKH107 derivative containing a Tn5 insertion within the 5.2 kb EcoRI fragment, showing the area used as probe 157

6.3 Hybridization of the 3.2 kb EcoRI/HindIII fragment from pSUP202 to plasmids carrying the mutagenized 5.2 kb EcoRI ß-galactosidase fragment of NCIB2951 (A) 158

6.4 Hybridization of the 3.2 kb EcoRI/HindIII fragment from pSUP202 to plasmids carrying the mutagenized 5.2 kb EcoRI ß-galactosidase fragment of NCIB2951 (B) 159

Page

6.5 Hybridization of the 3.2 kb EcoRI/HindIII fragment from pSUP202 to plasmids carrying the mutagenized 5.2 kb EcoRI ß-galactosidase fragment of NCIB2951 (C) 160

6.6 Hybridization of the 3.2 kb EcoRI/HindIII fragment from pSUP202 to plasmids carrying the mutagenized 5.2 kb EcoRI ß-galactosidase fragment of NCIB2951 (D) 161

6.7 Map of pKH107 containing the 5.2 kb EcoRI ß-galactosidase fragment of NCIB2951 with the positions of the Tn5 inserts shown 162

6.8 Autoradiograph of a 6% SDS-PAGE gel showing the plasmid-encoded proteins produced by in vivo maxicell [^{35}S]-methionine labelling 167

7.1 Diagram of the M13 life cycle 174

7.2 Diagram of modified M13 175

7.3 Sequencing strategy for the 5.2 kb EcoRI fragment 178

7.4 Nucleotide sequence of the 5.2 kb EcoRI fragment 182

7.5 Diagram of analysis of sequence using FRAMES 189

7.6 Nucleotide sequence and deduced amino acid sequence for ORF1 (cbgA) 192

7.7 Comparison of various ß-galactosidase amino acid sequences 197

7.8 Comparison of various ß-galactosidase simplified amino acid sequences 200

7.9 Nucleotide sequence and deduced amino acid sequence for ORF2 (cbgR) 205

7.10 Comparison of the amino acid sequences of CbgR and "0.3 kb gene" 207

7.11 Comparison of the simplified amino acid sequences of CbgR and the "0.3 kb gene" 207

Page

7.12 Nucleotide sequence and deduced amino acid sequence for ORF3 210

7.13 Comparison of the amino acid sequences of ORF3, Spo0A, and Spo0F 212

7.14 Comparison of the simplified amino acid sequences of ORF3, Spo0A, and Spo0F 213

LIST OF TABLES

		Page
1.1	Carbohydrates fermented by C. acetobutylicum	7
1.2	Homology studies between Clostridium species	18
1.3	Classification of Clostridium species	20
1.4	Mutants of C. acetobutylicum	24
1.5	Plasmids detected in C. acetobutylicum	27
1.6	Summary of experiments into shuttle vectors	28
1.7	Cloned genes of C. acetobutylicum	38
2.1	Bacterial strains used in this project	55
3.1	Differences in the cell wall components between Gram-positive and Gram-negative bacteria	94
3.2	Morphology of C. acetobutylicum cells at various stages of growth	108
4.1	Investigation of C. acetobutylicum DNA stability	119
5.1	Expression of C. acetobutylicum ß-galactosidase activity in E. coli	141
5.2	Detection of ß-galactosidase activity in various strains of bacteria using lactose selective medium supplemented with X-gal	144
6.1	ß-galactosidase activity in Tn5-mutagenized derivatives of pKH107	163
7.1	Plasmids used for the preparation of Bal31 deletion derivatives of the 5.2 kb EcoRI fragment	179
7.2	Table of Shine-Dalgarno and putative promoter sequences for the three ORF's	190

Page

7.3 Percentage (G+C) of the genomic DNA of specific bacteria + ß-galactosidase genes 215

7.4 Codon usage in different ORF's of genes from Clostridium species, compared to B. subtilis and E. coli 222

7.5 Comparison of codons used in ß-galactosidase genes from specific bacteria 224

7.6 Frequency (%) of amino acid utilization in proteins encoded by specific bacteria 228

ABBREVIATIONS

kb	kilobases
pfu	plaque-forming units
Ap	ampicillin
Cm	chloramphenicol
Km	kanamycin
Tc	tetracycline
SDS	sodium dodecyl sulphate
PEG	polyethylene glycol
kDa	kilodaltons
MDa	megadaltons
RF	replicative form
TTC	2,3,5 triphenyl tetrazolium chloride
X-gal	5-bromo-4-chloro-3-indoyl galactopyranoside
SS	single stranded
DS	double stranded
IPTG	isopropyl thiogalactopyranoside
ORF	open reading frame
SD	Shine-Dalgarno sequence