

**Genome Sequence of Bacteriophage
 ϕ AR29: a Basis for integrative
Plasmid Vectors**

Shawn Ginn Ming Seet

B. Sc., Hons. Biotech., MU.

This thesis is presented for the degree of
Doctor of Philosophy
at Murdoch University

March, 2005

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

Shawn Ginn Ming Seet

ACKNOWLEDGEMENTS

I would firstly like to thank Prof. Keith Gregg for his never ending support throughout my study of philosophy. During these last four years, Keith has performed beyond his means to provide funds for the continuation of his students' research as well as consistently meeting datelines. His guidance has also allowed me to open up my mind and think outside the box. For all that his has done, I will always be indebted to him.

I would also like to say thanks to my partner, Vera Limadinata, for taking this long journey with me. Her presence has been a comforting light in times of darkness. Her patience, humour and love have, and always will, put a smile on my face.

My deepest thanks go to my family, especially to my mother, for being the pillar of strength and continuous love. I would also like to thank them for understanding my absence around house and showing me that family runs deeper than blood.

Being the last member of Rumen Biotech, I like to give a special big thankyou to Tobias Schoep for standing beside me as long as he can. I never forget your encouragements, brain storming sessions and much laughter. Finally, thanks to Scott McNeil, Tom La and Andrew Low for helping me trouble shoot some of my difficulties and providing many funny jokes.

Lastly, sincere thanks to all SABC people for making the night-life in the lab enjoyable.

Table of Content

Abstract.....	1
Commonly Used Abbreviations	2
Chapter 1: Introduction.....	3
1.0: Genetic Manipulation of Rumen Bacteria	3
1.1: Criteria for Genetic Manipulation	5
1.2: Difficulties in Genetic Manipulation of Rumen Bacteria	6
1.3: Transformation Systems.....	7
1.4: Bacteriophages and their Life Cycle	11
1.4.1: Lytic Cycle.....	11
1.4.2: Lysogenic Cycle.....	13
1.4.3: Mechanism of Phage λ and Lysogenic Cycle.....	14
1.4.3.1: <i>The induction of phage λ lysogenic life cycle</i>	16
1.4.3.1.1: <i>Transcription of the Int gene</i>	16
1.4.3.1.2: <i>Maintenance of λ prophage</i>	17
1.4.3.2: <i>The role of protein Cro in the phage λ lytic cycle</i>	19
1.4.3.2.1: <i>Activation of the lytic cycle from the lysogenic state</i>	19
1.4.3.2.2: <i>Lytic cycle from infection</i>	20
1.4.4: Phage Integration.....	21
1.4.4.1: <i>attB site</i>	23
1.4.5: Excision of Prophage DNA from the Bacterial Chromosome	26
1.5: Phage in the Rumen	28
1.6: Bacteriophage as Transformation Tools.....	29
1.6.1: Transformation Systems That are Based on Phage Integrative Recombination Process	30
1.6.1.1 <i>Staphylococcus</i>	31
1.6.1.2 <i>Mycobacterium</i>	32
1.6.1.3 <i>Streptomyces</i>	33
1.6.1.4 <i>Pseudomonas</i>	34
1.6.1.5 <i>Streptococcus</i>	35
1.6.1.6 <i>Listeria</i>	36
1.6.1.7 <i>Enterococcus</i>	36
1.6.1.8 <i>Rhizobium</i>	36
1.6.2 Current developments in integrative vectors.....	37
1.7: Ruminal <i>Bacteroides</i>	43
1.7.1: Characteristics of <i>Bacteroides</i>	43
1.7.2: <i>Bacteroides thetaiotaomicron</i> Strain AR29 and <i>Bacteroides uniformis</i> Strain AR20	43
1.7.3: Difficulties in Transforming <i>B. thetaiotaomicron</i> AR29 and <i>B. uniformis</i> AR20.....	44
1.8: Rumen Bacteriophage ϕ AR29 and Development of a Transformation System for AR29 and AR20.....	45
1.9: Summary.....	48
1.10: Aim of This Project.....	48
Chapter 2: Materials and Methods.....	50
2.0: General Chemicals.....	50
2.1: Specialised Chemicals, Enzymes and Laboratory Equipments	50
2.2: <i>In silico</i> Analysis.....	52
2.3: Bacterial Strains.....	53
2.4: Commonly Used Solutions.....	55
2.5: Bacterial Growth Media.....	56
2.6: General Methods.....	58
2.6.1: Agarose Gel Electrophoresis.....	58
2.6.2: Ethanol Precipitation.....	59
2.6.3: Phenol:Chloroform:Isoamyl Alcohol Extraction	59
2.6.4: Polymerase Chain Reaction (PCR).....	59
2.6.5: Primer Design.....	60
2.6.7: Purification of DNA from Agarose Gels and Solution.....	62
2.6.8: Spectrophotometry for Nucleic Acid Quantification and Monitoring Bacterial Cell Growth	63
2.6.9: Nucleic Acid Isolation	64
2.6.9.1: <i>Plasmid DNA mini-preparations</i>	64
2.6.9.2: <i>RNA isolation</i>	65
2.6.9.3: <i>Removal of DNA contaminants from RNA preparations</i>	66
2.6.10: Restriction Enzyme reaction	67

2.6.11: Dephosphorylation of DNA 5' Terminus	67
2.6.12: Ligation of DNA	67
2.6.13: Preparation Electro-Competent Cells for Bacterial Transformation	67
2.6.13.1: <i>Preparing of Escherichia coli Electro-Competent Cells</i>	67
2.6.13.2: <i>Preparing of Bacteroides Electro-Competent Cells</i>	68
2.6.14: Bacterial Transformation Electroporation	69
2.6.14.1: <i>E. coli Transformation by Electroporation</i>	69
2.6.14.2: <i>B. uniformis Strain AR20 and B. thetaiotaomicron Strain AR29 Transformation by Electroporation</i>	
2.6.15: Screening for Transformed E. coli and Bacteroides	69
2.6.16: Slide Preparation for <i>Bacteroides</i> Cells – The Indian Ink Test	70
2.6.17: DNA Sequencing	70
Chapter 3: Expression of Integrase and Excisionase Genes from Plasmid pBA	73
3.0: Introduction	73
3.0.1: Aim	74
3.1.0: Materials and Methods:	74
3.1.1 Bacterial strains and Plasmids	74
3.1.2: RNA isolation	74
3.1.3: Two-Step Reverse Transcription Polymerase Chain Reaction (RT-PCR)	75
3.1.4: Gene Sequence Analysis	76
3.2: Results:	79
3.2.1: Transformation of <i>Bacteroides thetaiotaomicron</i> strain AR29 with pBA.	79
3.2.2: Detection of Int and Xis RNA transcripts in <i>E. coli</i> strain SCS110.	81
3.2.3: Detection of Int and Xis RNA transcripts in <i>B. uniformis</i> strain AR20.	82
3.2.4: Detection of Int and Xis RNA transcripts in <i>B. thetaiotaomicron</i> strain AR29	83
3.2.5: Sequence analysis of proposed Int and Xis	85
3.3: Discussion	86
3.3.1 Encapsulation of AR20 and AR29	86
3.3.2 <i>In silico</i> DNA sequence analysis of ϕ AR29 Int and Xis with other known recombinase gene in the GenBank	87
3.3.3 Detecting the transcripts of Int and Xis gene in SCS110, AR20 and AR29	87
Chapter 4: Characterisation of the Bacteriophage ϕ AR29 Genome	90
4.0: Introduction	90
4.0.1: Aim	91
4.2: Material and Methods	92
4.2.1: Bacterial strains	92
4.2.2: Phage Genomic DNA Preparation	92
4.2.3: Testing for the Presence of Cohesive Termini (cos site)	93
4.2.4: Sequencing of the ϕ AR29 Genome	96
4.2.5: Predicting ORFs	96
4.2.6: <i>B. thetaiotaomicron</i> AR29 Genomic DNA Preparation and Sequencing	97
4.2.7: Predicting the Position of Regulatory Elements in ϕ AR29 Genome	98
4.2.8: Predicting Translational Frameshifts in the ϕ AR29 Genome	98
4.2.9: Identifying tRNA in ϕ AR29 Genome	99
4.2.10: Cloning and Examination of Lytic Activity of ϕ AR29 Amidase Gene	99
4.2.11: Western Blots	101
4.3: Results:	102
4.3.1: Nucleotide Sequence of the ϕ AR29 Genome	102
4.3.2: Identification of ϕ AR29 Open Reading Frames, Functional Prediction of Genome Organization	105
4.3.3: Identifiable ORFs	109
4.3.3.1: <i>ORF5 (Terminase Large subunit)</i>	109
4.3.3.2: <i>ORF16 (DnaC protein)</i>	111
4.3.3.3: <i>ORF 25 (Transcription response regulator protein)</i>	113
4.3.3.4: <i>ORF 31 (cI protein)</i>	114
4.3.3.5: <i>ORF 35 (Integrase)</i>	116
4.3.3.6: <i>ORF40 (N-acetylmuramoyl-L-alanine amidase)</i>	120
4.3.4: ORFs with significant matches to known genes and proteins	125
4.3.4.1: <i>ORF4</i>	125
4.3.4.2: <i>ORF6</i>	125
4.3.4.3: <i>ORF17</i>	125
4.3.4.4: <i>ORF18</i>	126

4.3.4.5: <i>ORF22</i>	126
4.3.4.6: <i>ORF 23</i>	127
4.3.4.7: <i>ORF51</i>	127
4.3.5: Poorly Matched ORFs	127
4.3.6: tRNA Gene Prediction	128
4.3.7: Frameshift Analysis	129
4.3.8: Regulatory Elements in ϕ AR29 Genome.....	129
4.3.9: Comparison of ϕ AR29 gene arrangement with other bacteriophages.....	132
4.3.10: Locating and Identifying the Integration Site	134
4.3.11: Identifying the Site and Orientation of Integration of ϕ AR29 into <i>B. thetaiotaomicron</i> AR29	139
4.3.12: Amidase Activity Studies	140
4.4: Discussion	142
4.4.1: DNA Sequence of ϕ AR29	142
4.4.2: Characterisation of the ϕ AR29 Genome	143
4.4.3: Prediction of a Pseudo-tRNA (Leu)	146
4.4.4: Coding Frameshifts.....	150
4.4.5: Regulatory Elements	151
4.4.6: <i>Att</i> Sites.....	152
4.4.6.1: <i>AttP</i>	152
4.4.6.2 <i>AttB</i> site	153
4.4.6.3: <i>Core attachment site</i>	155
4.4.7: Determining the Function of ORF40 (amidase)	156
4.5: Conclusion	157
Chapter 5: Chromosomal integration of plasmid pBA in <i>B. thetaiotaomicron</i> AR29	159
5.0: Introduction.....	159
5.0.1: Phage Integrases and Attachment (<i>att</i>) sites	160
5.0.1.1: <i>Serine Integrases</i>	160
5.0.1.2: <i>Serine Integrase Attachment</i>	164
5.0.1.3: <i>Tyrosine Integrases</i>	164
5.0.1.4: <i>Tyrosine Integrase Att site</i>	166
5.0.2: Plasmid pBA	166
5.0.3 Aim:	168
5.1: Materials and Methods	168
5.1.1: Bacterial strains and Plasmids, Sequence Analysis and RNA Isolation	168
5.1.2: Testing for the Expression of Prophage <i>Int</i> Genes from <i>B. thetaiotaomicron</i> AR29	168
5.1.3: Testing for pBA Integration	169
5.1.3.1: <i>Preparation of samples for PCR analysis</i>	169
5.1.3.2: <i>Tests for the presence of integrated plasmid pBA</i>	169
5.1.3.3: <i>Detection of free, non-integrated plasmid pBA in AR20 and AR29.</i>	169
5.1.3.4: <i>Tests for plasmid pBA integration adjacent to prophage in B. thetaiotaomicron AR29.</i>	170
5.1.4: PCR Detection of Unrecombined <i>B. thetaiotaomicron</i> AR29 <i>attB</i> Site.	172
5.1.5: Approach to Curing Lysogenic Phage ϕ AR29 from <i>B. thetaiotaomicron</i> AR29	173
5.1.5.1: <i>Monitoring the removal of ϕAR29 by determining the presence of intact attP, and integrated phage ϕAR29 in AR29 with PCR</i>	173
5.2: Results	173
5.2.1: Identification of <i>Int</i> and <i>att</i> site in pBA	173
5.2.2: Expression of <i>Int</i> from the ϕ AR29 Prophage.....	173
5.2.3: Test for Integration of pBA into AR20 and AR29 Genomes.....	175
5.2.4: Characterisation of pBA Integration and Excision of Lysogenic Phage ϕ AR29	177
5.2.5: Identification of Potential <i>attB</i> Site in Other Bacterial Genomes	181
5.3: Discussion	184
5.3.1: Presence of <i>Int</i> in AR29 Untransformed with pBA	184
5.3.2: Absence of Co-Integrated pBA and ϕ AR29	185
5.3.3: Detection of intact <i>attB</i> and <i>attP</i>	186
5.3.4: Detection of integrated phage and integrated pBA	190
5.3.5: Mechanisms for Excision of ϕ AR29, Integration of pBA, and Curing of AR29	191
Model 1:.....	192
Model 2:.....	193
Model 3:.....	194
5.3.6: The Identification of Potential <i>attB</i> Sites in Other Organisms.....	195

5.4: Conclusions	196
Chapter 6: General Conclusion and Closing Remarks	197
6.0: Characterisation of the integration module of bacteriophage ϕ AR29, cloned in pBA	198
6.1: Complexity of the Phage genomic sequence	199
6.1.1: Mosaic of the phage genome	199
6.2: Current approach of bacteriophage Integrative vectors in mammalian cells as tools for gene for gene therapy	201
6.2.1: Tyrosine integrase	201
6.2.2: Serine Integrase	202
6.3: Future Work	204
6.3.1: Genomic Analysis	205
6.3.2: Refinement of pBA	205
6.3.3: Application of pBA in Other Bacteria	208
6.4: Closing remarks	210
References	211
Appendix 1	231
Figure 3.1A:	231
Figure 3.2A:	237
Appendix 2	239
Table 4.1A:	239
ϕ AR29 Genomic Sequence	242

ABSTRACT

The initial aim of this project was to characterise the integrative recombination mechanism of bacteriophage ϕ AR29, to provide a better understanding for development of the shuttle plasmid pBA as a site-specific *Bacteroides* integration vector. RT-PCR showed that the previously identified ϕ AR29 recombination genes, integrase (*Int*) and excisionase (*Xis*), were transcribed from pBA in *E. coli* SCS110, *B. thetaiotaomicron* AR29 and *B. uniformis* AR20. *In silico* derived amino acid sequences from both genes showed only very low levels of similarity to other known *Int* and *Xis* in GenBank. To improve understanding of the phage recombination system, the ϕ AR29 genome was sequenced. This revealed a 35,558 bp double-stranded DNA genome with GC content of 39.11%. Bioinformatic analysis identified 53 open reading frames (>30 codons) and gene promoters and terminators that allowed the genome arrangement to be compared with other phages. Comparison of deduced gene products with proteins from other phages identified 6 reading frames, allowed tentative identification of 7 others, but left 40 ORFs unidentified. Those with strong homology to known genes were: large terminase subunit (44.66 kDa), *dnaC* (27.94 kDa), helix-turn-helix (*HTH*) transcription regulator (14.69 kDa), *cI* repressor (26.48 kDa), amidase (18.42kDa) and a novel integrase (54.22 kDa). The integrase gene is located 162 base-pairs downstream of the phage attachment (*attP*) core site, rather than the previously suggested location upstream of the integration site. The ϕ AR29 *attP* was shown to include a 16-bp *att* core region, 117 bp upstream of the previously suggested location. Integration of ϕ AR29 was found to occur at the 3' end of an arg-tRNA gene on the AR29 genome (*attB*). Imperfect direct repeats with a consensus sequence (ANGTTGTGCAA) were found surrounding the *attP* core. A review of pBA sequence showed that only the 5' end (435 bp) of the newly identified *Int* gene was cloned in pBA. Despite this, PCR analysis revealed integration of pBA into the AR29 genome. Serial subculturing of pBA-transformed AR29 was able to cure AR29 of the ϕ AR29 prophage, providing an improved host for integrative plasmids, and for detailed studies of AR29 physiology and ϕ AR29 life cycles.

COMMONLY USED ABBREVIATIONS

Abbreviation	Expansion
ABI	Applied Biosystems Inc.
AMV	Avian myeloblastosis virus
ATP	adenosine-5'-triphosphate
AR	analytical reagent
b	bases
bp	base pair
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic Acid
DNAse	deoxyribonuclease
dNTP	deoxynucleotides
ds	double stranded
EDTA	ethylenediaminetetra-acetic acid
IPTG	isopropyl- β -D- thiogalactopyranoside
MCS	multiple cloning site
OD	optical density
ORF	Open Reading Frame
PCR	polymerase chain reaction
pers. comm.	personal communication
RNAse	ribonuclease
SDS	sodium dodecyl sulphate
TAE	Tris-acetate EDTA buffer
TBE	Tris-borate EDTA buffer
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside