# Genome Sequence of Bacteriophage ϕAR29: 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 continuos 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

The initial aim of this project was to characterise the integrative recombination mechanism of bacteriophage ϕAR29, 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 ϕAR29 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 ϕAR29 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 ϕAR29 *attP* was shown to include a 16-bp *att* core region, 117 bp upstream of the previously suggested location. Integration of ϕAR29 was found to occur at the 3'end of an arg-tRNA gene on the AR29 genome (*att*B). Imperfect direct repeats with a consensus sequence (ANGTTGTGCAA) were found surrounding the *att*P 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 ϕAR29 prophage, providing an improved host for integrative plasmids, and for detailed studies of AR29 physiology and ϕAR29 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 |