STUDIES ON THE TAIL REGION

OF THE

TEMPERATE COLIPHAGE 186 GENOME

A thesis submitted for

the degree of Doctor of Philosophy

at the University of Adelaide by

QING XUE (M. Sc)

Department of Biochemistry

October 1993

In Response to the 5 issues raised by my examiner Gail Christie

Point 1

At this stage of the analyses it is a fact noted that the rare codons commonly appear in the N terminal region of *orf38*, and that the situation is similar for the homologous gene *I* of P2. It is not known whether the gene is efficiently translated.

Corrections to thesis

i) p60 Line 17, P2 gene *J* altered to read P2 gene *I*.

ii) Inserted Line 19:

"……adopted. The sequence from 180 to 530 of P2 gene *I* shows similar use of rare codons."

Point 2:

The criticism of the examiner is valid.

Lines 9-15 of p73 were deleted and the following inserted:

"However neither the strength of the hairpin nor the length of the run of U's lends confidence to any claim for terminator status, but the presence of an inverted repeat is noted."

Point 3:

The examiner has requested further comment. I have therefore inserted the summary page 86 the following:

" V. The polarity studies of Hocking and Egan (1982) clearly suggested the existence of a promoter between genes *H* and *G*, but at this stage no further evidence can be offered for its existence. Dibbens and Egan (1992) were able to detect the late promoter P_{12} by primer extension studies, but not by GalK reporter studies, suggesting the better sensitivity of primer extension studies. As the polarity results suggest that genes *G*, *F*, *E* and *D* are transcribed in infections with either *J*, *I* or *H* amber mutants it seems appropriate to examine the transcripts encoding genes *G*, *E*, *F* and *D* from cells infected with 186 *J*am41 or 186 *I*am40a for evidence of a P_G promoter."

Point 4

No homology with the distal part of P2 gene *H* could be found associated with the downstream (inversion) element in the 186 sequence.

No comment added to thesis.

Point 5

There have been no biochemical studies on the late proteins of 186 performed in the Egan laboratory, nor anywhere else to my knowledge. At the stage of writing my thesis the sequence of P2 homologous with 186 *CP53*, *CP54* and gene *G* were not published, and therefore no comparison could be made.

No comment added to thesis.

TO ALL MY YOUNG FRIENDS

WHO DIED FOR THEIR SOCIETY'S PROGRESS

Nucleotides Are Beautiful

CONTENTS

ACKNOWLEDGEMENTS

I am grateful to the University of Adelaide for the scholarship that made this thesis possible.

I deeply thank the following persons who have been associated with this thesis:

Professor W.H.Elliott, for his permission to study in the Department of Biochemistry. His noble character and genuine academic attitude have been an inspiration to me.

Professor G.E. Rogers, Head of Department, for his excellent and continual efforts to create advanced study and research environment.

Members of Laboratory 19, both past and present. In particular I would like to thank Dr. Justin Dibbens and Dr. Ian Dodd for their enthusiastic, constructive help and good advice in experimental technique, academic discussion and computing and valuable comments on my thesis;

Sivaprasad, Lel Whitbread, Tony Fratini, and Satish Dogra for their help in the specific field, Lesley Crocker for photography;

Jon Anargyros, Janis Soltys, Brian Denton, Jan Somerville, Ros Murrell, and other members of the Department for their excellent technical assistance and friendship.

I would specially pay my sincere tribute to my supervisor Dr. John Barry Egan, for his direction, encouragement, understanding and enthusiasm throughout the course of my Ph.D., and for his critical reading and criticising of this thesis. His succinct style and opimistic philosophy on life are also the precious resources that will benefit me for my whole life.

Finally, I wish to thank my parents and my wife for all their endurance.

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge and belief, this thesis contains material not previously published or written by another person except where due reference is made in the text.

I consent to this thesis being made available for photocopying and loan.

QING XUE

Abstract

The aim of this thesis is to study the genetic organisation and expression of a 7.2 kb tail region on phage 186 genome. In this thesis, the complete sequence of the 7251 bp *Hin*dIII fragment spanning 37.6% to 61.3% of the phage 186 genome is presented. Computer-assisted analysis revealed nine potential coding regions aligned on the *l*-strand of this fragment. By marker rescue test five previously known tail genes *K, J, I, H* and *G* were located in the corresponding coding frames, whose ordering agrees with the previous studies on 186 genetic map. An incomplete open reading frame was designated *orf37c* , while three other coding frames, were designated *CP38, CP45* and *CP54*..

Combining previous polarity results of Hocking and Egan with the analysis on the 7.2 kb *Hin*dIII fragment sequence, two regions were presumed to harbour potential late promoters. The predicted promoter PJ precedes gene J, a leading gene of polarity group JIH, and possesses the typical late promoter consensus sequence of P2 and 186. Galactokinase reporter studies showed the transcription activity from the fragment containing PJ. Primer extension analysis of in *vivo* RNA located the 5'-end of a mRNA at exactly expected +1 position deduced from the late promoter consensus sequence. The studies in this thesis also indicated that the activity of the late promoter PJ is dependent upon the B gene product.

A piece of sequence preceding gene G had been presumed as a late promoter PG based on the polarity studies and sequence analysis, but no transcription activity was shown in GalK assay. Primer extension also failed to indicate the 5'-end of expected RNA. Northern blotting studies revealed that a 6.5 kb transcript could be detected by both probe J and probe G. Most likely this

6.5 kb transcript initiated from promoter PJ and passed through PG region and the downstream genes *G, F, E, D*, and *B* till the terminator TB.

A potential rho-independent terminator T45, which precedes promoter PJ, is proposed. Another potential rho-independent-terminator on the 7.2 kb *Hin*dIII fragment sequence, TI , was strongly suggested by both the DNA sequence feature and the preliminary transcription studies on PJ. This potential terminator is positioned between gene *I* and *CP52* within the PJ transcript. The need of the phage 186 for a great amount of sheath and tube proteins also supports the existence of TI, by which the short transcript responsible for the sheath and tube proteins could be accumulated.

The comparative studies on the 7.2 kb *Hin*dIII fragment sequence offered information on the gene function of the 186 tail region. The sequence presented in this thesis exhibited extensive and high similarity with all the available five tail genes of phage P2, both at the level of DNA and the level of deduced amino acid sequences. Most likely, 186 gene *K* is responsible for the tail fiber protein, *CP45* is for an assembly protein, gene *J* is for the tail sheath protein while gene *I* for the tail tube. The 7.2 kb tail region of 186 also displayed similarities with unrelated phages λ, T3, T7 and some inversion systems, such as phage Mu G-gin, phage P1 C-cin, E.coli plasmid pB15 P-pin , and S.boydii P-pinB. Mu-IR-L-like sequences were located in 186 gene *K* and P2 gene *H* . They are 67% and 90% identical to Mu IR-L sequence respectively. An imperfect IR-R-like sequence was found on the *r*-strand of the 7.2 kb *Hind*III region with 60% identity to the 186 Mu-IR-L sequence. A DNA flow possibly happened between E.coli gyrase *A* and phage 186 gene *K*, since they in part share similarity in DNA sequence. The impressive similarities and parallel ordering of the subregions between the 186 tail fibre genes and other relative systems hint that these phages may have employed the same cellular component for their tail fiber function in the early stage of their evolution.

ABBREVIATIONS

a.a.	- amino acids
A_{600}	-absorbance at 600 nm
A_{650}	-absorbance at 650 nm
am	-amber mutant
bp	-base-pair
cfu	- colony-forming units
Ci	-Curie
CP	-Computer predicted
DNA	-deoxyribonucleic acid
EDTA	- ethylenediamine tetra acetate
EM	- electron microscope
g	- gravitational force
gm	-gram
kb	-kilodalton
M	-Molar
mA	-milliamp
mg	-milligram
min	- minute
ml	-millilitre
mM	-millimolar
ng	-nanogram

nm	-nanometre
orf	-open reading frame
orfs	-open reading frames
pfu	-plaque-forming units
RBS	- ribosome-binging site
RNA	-ribonucleic acid
rpm	- revolutions per minute
Tris	-Tris (hydroxymethyl) aminomethane
U	- units
uCi	-microCurie
ug	-microgram
ul	-microlitre
UV	-ultraviolet
V	-volts