Molecular Analysis of *Shigella flexneri* Bacteriophage SfV *Bam*HI Fragment A Encoding the Viral Late Region and the Role of SfV Genome in Host Virulence

Dario C. Angeles

A thesis submitted for the degree of Master of Philosophy of the
Australian National University
Division of Biochemistry and Molecular Biology
School of Life Sciences

March 2001
Canberra

Statement of Originality

I declare that this thesis is an original work and is a compilation of research results conducted and interpreted by myself in consultation with my supervisors. To my knowledge, this thesis does not contain any material that has been previously published or presented in a forum, except where references have been cited in the text.

Dario C. Angeles

This thesis is dedicated to the memory of my dear friend

Joel Bayaron

October 28, 1967
To
February 13, 2001

Acknowledgement

I would like to thank my supervisors, Dr. Naresh Verma, for accepting me to work in his laboratory and for his much needed guidance, and Dr. Gwen Allison who has done so much not only for the completion of my project but also towards my personal well-being. Thank you Gwen for all the warmth and care you have extended. Thanks to my lab mates Robyn Smith, Minakshi Vishnoi and Parvin Khinda for their friendship and advices, to John Walker for sharing his expertise on Southern hybridisation techniques and his advices, to Mat Adams for his assistance and kind gestures, and to Harris, Nai and Shalini for the laughters and friendship even for a short moment.

My sincere thanks to Adele Loy for proofreading my thesis manuscript, to Lily Shen and Sally Stowe of the Electron Microscopy Unit in RSBS for teaching me the basics in electron microscopy, and to John Maindonald of the Statistical Consultation Unit for assisting in the statistical treatment of my invasion assay data.

Special thanks goes to my mama, papa, my family, and my dearest Boom-Boom who all served as my inspiration in finishing my studies. Thank you also to all my Pinoy friends in Canberra who made my stay colorful, enjoyable and bearable. And thank you to the Australian government (AusAID) for giving me the opportunity to do research studies at ANU and to experience the beauty and joy of living in Canberra.

Abstract

Shigella flexneri is a pathogen capable of invading the colonic epithelium causing shigellosis or bacillary dysentery. World Health Organisation estimates that there are over a billion cases of diarrhea and dysentery, annually, making the development of vaccine against the different *Shigella* serotypes a WHO priority [WHO., 1991 #529][WHO, 1997 #96]. SfV is a temperate bacteriophage capable of mediating *S. flexneri* O-antigen modification that confers serotype-conversion from Y to 5a. Aside from its role in serotype-conversion, little is known about its biology. The genetic composition of phage SfV needs to be examined in the prospect of discovering other phage-encoded factors which may have greater role in conferring virulent traits. In this project, we have endeavoured to further characterise the SfV genome beyond the serotype-conversion gene region.

To proceed with the molecular characterisation of the SfV genome, the 10.1 kb *Bam*HI fragment A region adjacent to the *pac* site was sequenced using a direct DNA sequencing approach which employed cloning and primer walking combined with Southern hybridisation techniques. Initially, universal forward and reverse primers were used to sequence both strands of the pNV728 insert which carries the initial 5.5 kb portion of SfV *Bam*HI fragment A. Afterwhich, the gaps were filled through primer walking using customised primers. The DNA beyond pNV728 *Sac*I site was sequenced from overlapping fragments detected by probes containing pNV728 nucleotide sequences adjacent to the *Sac*I site. A total of 54 input sequences were assembled to derive the 10109 bases of SfV *Bam*HI fragment A.

Analysis of the 10.1 kb sequence revealed the presence of 1 incomplete and 13 complete open reading frames. Based on database homology search and comparative analysis

between fragment A open reading frames and the non-redundant protein database, this region was found to encode for gene products which share homology to essential major and minor structural proteins of bacteriophages, primarily of the family Siphoviridae. Of the 14 open reading frames predicted, eight were deduced as encoding for the putative terminase subunits, the portal protein, protease, head protein and tail assembly components.

Functional studies of *orf200* and *orf409* were performed to ascertain their identity. Preliminary Western immunoblotting and amino acid sequencing results indicated ORF409 product as the capsid protein. N-terminal amino acid sequence of the 32 kDa phage SfV protein band H, AQGVAQDEKG, was located internal to the ORF409 predicted amino acid sequence suggestive of ORF409 processing as part of the head protein maturation. This process was demonstrated through experiments involving recombinant constructs, one carrying the 2055 bp complete *orf200* and complete *orf409* insert (B876) and the others carrying the 1616 bp incomplete *orf200* and complete *orf409* insert (B877 and B878). Overexpression of B877 and B878 by IPTG induction revealed the accumulation of the 45.8 kDa ORF409 capsid protein which was not observed in B876. The experiments suggest that a complete functional ORF200 protease is required to mediate cleavage of the ORF409 capsid protein to its mature 32 kDa form. A similar head morphogenetic pathway was observed in other bacteriophages including *Pseudomonas* phage D3, *Streptomyces* phage C31 and coliphage HK97. In addition, the striking similarity in the organisation of the DNA packaging and head assembly genes among bacteriophages, supports not only the designation of *orf*200 and *orf*409 as the protease and capsid genes, respectively, but also that of the putative *orf*577 terminase and *orf*367 portal protein genes. The cluster arrangement of these genes denotes a well conserved organisation reflective of their movement as a single module during gene transfer and its origin from a common ancestral source.

In order to assess the role of phage SfV genome in host *S. flexneri* virulence, a SfV lysogen strain wild-type serotype Y strain carrying bacteriophage SfV, a SfV cured serotype 5a strain, and recombinant SFL1339 strains carrying different segments of the SfV genome were produced. These strains were compared based on their ability to invade HeLa cells. The invasion level of the lysogen strain SFL1338 was significantly higher than the invasion level of its serotype Y parental strain SFL1339, and the invasion level of the cured strain SFL1337 was significantly lower compared to the level detected for its serotype 5 parental strain SFL1336. These results suggest that the SfV genome is enhancing the invasive potential of *S. flexneri*. To investigate which portions of the genome are most responsible for the observed increase in cell invasion, isogenic SFL1339 strains were produced containing various SfV genomic segments. SFL1346 containing the serotype-conversion genes $gtrA_{(V)}$, $gtrB_{(V)}$, and *gtrV* showed the most significant difference to the invasion level of the control strain SFL1342 containing only the vector. These experiments implicated that the three-gene cluster was the only SfV gene locus effective in enhancing bacterial invasiveness. Further studies were conducted to determine if the invasion level of SFL1346 would decrease upon the disruption of the glucosyltransferase gene, *gtrV*. Plasmid pNV731 was mutated by inducing a frameshift mutation in the *gtrV* gene, then transformed into SFL1339, designated as SFL1394. The invasion level of SFL1394 compared with SFL1342 and SFL1346, showed partial reduction in its invasive capabilities, suggestive of the influence of $gtrA_{(V)}$, and $gtrB_{(V)}$, in conferring *S. flexneri* invasive traits.

TABLE OF CONTENTS

Acknowledgement 1
Abstract 2
Table of Contents 5
List of Figures 9
List of Tables 12
Abbreviations 13

Chapter 1: Introduction 16
1.1. Genus *Shigella* 17
1.2. Shigellosis 18
1.3. Epidemiology 19
1.4. Pathogenesis 20
1.5. Virulence factors 23
1.5.1. Plasmid-encoded virulence factors 23
1.5.2. Chromosome-encoded virulence factors 24
1.6. Immune responses to shigellosis 27
1.6.1. Mucosal humoral immunity 27
1.6.2. Systemic humoral immunity 29
1.6.3. Cell-mediated immunity 30
1.7. Somatic antigen (O-antigen) of the lipopolysaccharide 31
1.7.1. Structure and function 31
1.7.2. Biosynthesis 34
1.8. Bacteriophage-mediated O-antigen modification 35
1.8.1. Temperate bacteriophages 35
1.8.1.1. *S. flexneri* bacteriophage SfV 37
1.8.1.2. Coliphage λ 39
1.8.2. Serotype-converting *S. flexneri* phages and O-antigen modification 41
1.9. Bacteriophage and their role in virulence 46
1.10. Objectives 47

Chapter 2: Materials and Methods 48
2.1. Culture conditions and growth media 48
2.2. Bacterial strains, plasmids and vectors 48
2.2.1. Bacterial Strains 48
2.2.2. Plasmids and vectors 49
2.3. Bacteriophage induction 51
2.3.1. Phage induction by UV irradiation 51
2.3.2. Phage induction by mitomycin-C treatment 52
2.3.3. Phage induction by *xis* overexpression 53
2.4. Preparation of bacteriophage 53
2.4.1. Phage propagation 53
2.4.2. Phage purification and DNA extraction 55
2.4.3. Plaque assay for phage titration 56
2.5. Bacteriophage detection and identification in lysogen 56
2.5.1. Phage sensitivity test 56
2.5.2. Slide agglutination test 57

2.6. DNA processing and basic manipulations 58
2.6.1. Preparation of plasmid DNA 58
2.6.1.1. Minipreparation of plasmid DNA 58
2.6.1.2. PEG precipitation of plasmid DNA 59
2.6.2. Preparation of chromosomal DNA 60
2.6.3. Preparation of DNA for radiolabelling as probe 61
2.6.3.1. Gel-purified restriction fragments 61
2.6.3.2. PCR extension product 62
2.6.3.3. SfV genomic DNA 62
2.6.4. Agarose gel electrophoresis 63
2.6.5. Quantitation of DNA concentration 63
2.6.6. DNA purification by silica matrix binding 64
2.6.7. Restriction enzyme digestion 65
2.6.8. Dephosphorylation of linearised plasmid DNA 66
2.6.9. Klenow treatment 66
2.6.10. Ligation 67
2.7. Competent cell transformation 67
2.7.1. Preparation of competent cells 67
2.7.1.1. Electrocompetent cells 67
2.7.1.2. Rubidium chloride competent cells 68
2.7.2. Transformation 69
2.7.2.1. Transformation by electroporation 69
2.7.2.2. Transformation by heat-shocking 69
2.8. Polymerase chain reaction 70
2.8.1. Amplification for DNA sequencing 70
2.8.1.1. Checking primers for secondary structures 70
2.8.1.2. Preparing sequencing reactions 70
2.8.1.3. Purification of PCR extension products 71
2.8.2. Amplification of gene using a plasmid template 72
2.8.3. Amplification of gene using SfV genome as template 73
2.8.4. Colony PCR 76
2.9. DNA sequencing and computer analysis 77
2.10. Southern hybridisation 78
2.10.1. Alkali blotting of DNA onto nylon membrane 78
2.10.2. Membrane blocking with non-homologous DNA 79
2.10.3. Hybridisation and autoradiography 79
2.10.4. Membrane stripping 80
2.11. SDS-PAGE and Western immunoblotting 80
2.11.1. Protein assay using trichloroacetic acid precipitation 80
2.11.2. SDS-PAGE 81
2.11.3. Protein band staining 83
2.11.3.1. Coommassie blue staining 83
2.11.3.2. Silver staining 83
2.11.3.3. Ponceau S staining 84
2.11.4. Western immunoblotting 84
2.11.4.1. Electrophoretic protein transfer 84
2.11.4.2. Acetone powder preparation 86
2.11.4.3. Adsorption to remove nonspecific binding 86

2.11.4.4. Western immunoassay 87
2.12. Electron microscopy 88
2.13. HeLa cell culture 88
2.13.1. Cell passage and seeding 88
2.13.2. Cell freezing and thawing 89
2.13.3. Cell counting in haemocytometer 90
2.14. Invasion/Gentamycin killing assay 90

Chapter 3: DNA sequencing strategy for bacteriophage SfV *Bam*HI fragment A 92
3.1. Introduction 92
3.2. Results 95
3.2.1. Bacteriophage SfV fingerprint for confirmation of identity 95
3.2.2. Characterisation of pNV728 96
3.2.3. Sequencing of the 5.5 kb portion of *Bam*HI fragment A 97
3.2.4. Restriction probes detected fragments flanking DNA immediately downstream of the initial 5.5 kb of *Bam*HI fragment A 99
3.2.5. Sequencing downstream of pNV728 *Sac*I site 100
3.3. Discussion 102

Chapter 4: Sequence analysis of *Shigella flexneri* bacteriophage SfV BamHI fragment A which encodes for structural proteins of the late region 105
4.1. Introduction 105
4.2. Results 107
4.2.1. Analysis for the presence of *orfs* in the 10.1 kb sequence 107
4.2.2. Putative functions of the ORFs 108
4.2.3. ORF577 108
4.2.4. ORF367 110
4.2.5. ORF200 111
4.2.6. ORF409 112
4.2.7. ORFs 56, 498, 89 and 116 114
4.3. Discussion 115

Chapter 5: Post-translational cleavage of the 45 kDa major prohead capsid subunit is catalysed by Shigella flexneri SfV bacteriophage-encoded protease 123
5.1. Introduction 123
5.2. Results 124
5.2.1. Protein band H N-terminal sequence located within ORF409 124
5.2.2. Localisation and amplification of the putative protease and capsid Gene insert 125
5.2.3. ORF200 is required to cleave ORF409 gene product 127
5.2.4. Immunoblot analysis confirmed ORF409 as capsid 129
5.3. Discussion 131

Chapter 6: The role of Shigella flexneri bacteriophage SfV genome in host Virulence 137
6.1. Introduction 137
6.2. Results 139

6.2.1. Preparation of SFL124 and SFL1 lysogen strains 139
6.2.2. Preparation of cured wild-type EW595/52 type V strain 140
6.2.3. Preparation of SFL1 isogenic strains transformed with plasmids containing genomic portions of phage SfV 143
6.2.4. HeLa cell invasion assay 144
6.2.5. Invasiveness of *gtrV* mutant (*gtrV*⁻) 147
6.3. Discussion 149

Chapter 7: General Discussion 154
7.1. The serotype-conversion and early region of phage SfV genome 154
7.2. Fragment A encodes for phage SfV structural proteins necessary for viral packaging of DNA and morphogenesis 155
7.3. The role of phage SfV ORF577 putative terminase in DNA packaging 158
7.4. Serotype-conversion genes are phage SfV-encoded virulence factors 160
7.5. Conclusion and future direction 163

Appendix 166
References 169

LIST OF FIGURES

Figure 1.1. Invasion of the colonic epithelium by *Shigella*.

Figure 1.2. Structure of the lipopolysaccharide of *Shigella flexneri*.

Figure 1.3. Structure of *Shigella flexneri* O antigens

Figure 1.4. Schematic illustration of phage λ site-specific recombination

Figure 1.5. Transcriptional and functional map of λ prophage

Figure 1.6. Organisation of *S. flexneri* O antigen modification genes.

Figure 1.7. Hypothetical model for O-antigen glucosylation

Figure 2.1. Schematic map of pUC19

Figure 2.2. Schematic map of pT7-5

Figure 2.3. Schematic map of pTrc 99A

Figure 2.4. Schematic diagram of DNA strand synthesis by formation of phosphodiester bonds

Figure 2.5. Schematic diagram of capillary transfer set-up for alkaline blotting

Figure 2.6. Schematic flow diagram of the invasion assay method

Figure 3.1A. Agarose gel electrophoresis of bacteriophage SfV genome digests

Figure 3.1B. Physical map of bacteriophage SfV

Figure 3.2. Construction of pNV728 and transformation into JM109

Figure 3.3. Analysis of single and double digests of pNV728

Figure 3.4. Recombinant plasmids containing *Hin*dIII fragments of pNV728 and pUC19

Figure 3.5. Restriction map of pNV728

Figure 3.6. Location of *Eco*RV-*Sac*I probe 1 and *Hin*dIII probe 2 relative to pNV728 insert

Figure 3.7. Southern hybridisation of SfV genome digests detected by probe 1

Figure 3.8. Southern hybridisation of SfV genome digests detected by probe 2

Figure 3.9. Analysis of combination sequence and hybridisation results

Figure 3.10. Gelassemble bigpicture of primer location and designation used to sequence the 10.1 kb *Bam*HI fragment A

Figure 3.11. Linear sequence of the 10.1 kb *Bam*HI fragment A with relevant restriction sites

Figure 4.1. Functional map of bacteriophage SfV early region

Figure 4.2. Transcriptional and functional map of the 10.1 kb *Bam*HI fragment A

Figure 4.3. Linear sequence of the 10.1 kb *Bam*HI fragment A showing the *orfs*

Figure 4.4. Multiple sequence alignment of terminase proteins

Figure 4.5. Alignment of protease proteins

Figure 4.6. Comparison of head assembly genes organisation among bacteriophages

Figure 4.7. Alignment of tail component proteins

Figure 4.8. Schematic diagram illustrating phage SfV tail protein homologues

Figure 4.9. Transmission electron micrograph of bacteriophage SfV

Figure 5.1A. SDS-PAGE of PEG precipitated SfV particles

Figure 5.1B. Western immunoblot of SfV proteins hybridised with unadsorbed antisera

Figure 5.1C. Western immunoblot of SfV proteins hybridised with adsorbed antisera

Figure 5.1D. N-terminal amino acid sequence of SfV protein band H

Figure 5.2A. Location and size of *Bam*HI fragment A in the SfV physical map

Figure 5.2B. Transcriptional and functional map of *Bam*HI fragment A

Figure 5.2C. Map of SfV head assembly *orfs* showing the location of the PCR amplified inserts

Figure 5.3. Schematic diagram of pNV769 and pNV770 cloning and their recombinant strains

Figure 5.4A. SDS-PAGE of B866, B876, B877 and B878 at 0 and 60 min post-IPTG induction

Figure 5.4B. SDS-PAGE of B866, B876, B877 and B878 at 180 min post-IPTG induction

Figure 5.5. SDS-PAGE of B876 and B877 proteins IPTG-induced over time

Figure 5.6. Western immunoblot of strains B866, B876, B877 and SfV proteins detected by

Anti-SfV polyvalent antisera

Figure 5.7. Alignment of phage protease sequences showing the putative catalytic sites and interpeptide residues

Figure 5.8. Schematic representation of two types of bacteriophage head assembly process

Figure 5.9. Partial ORF409 DNA and protein sequences showing cleavage site

Figure 6.1. Schematic diagram showing steps and methods used in the curing of SFL1337

Figure 6.2. Colony blot hybridisation of SFL1336 colonies after IPTG induction

Figure 6.3. Southern hybridisation of chromosomal DNA of prepared lysogen, cured and isogenic strains probed with the SfV genome

Figure 6.4. Schematic diagram showing phage SfV segments cloned in SFL1339 isogenic strains

Figure 6.5A. Comparison of the invasion level between SFL1338 and SFL1339

Figure 6.5B. Comparison of the invasion level between SFL1336 and SFL1337

Figure 6.6A. Comparison of the invasion level among SFL1339 isogenic strains

Figure 6.6B. Tukey multiple comparison of ratio values of SFL1339 isogenic strains

Figure 6.7. Diagramatic representation of the frameshift mutation introduced into pNV731

Figure 6.8A. Comparison of the invasion level among SFL1342, SFL1346 and SFL1394

Figure 6.8B. Tukey multiple comparison of ratio values among SFL1342, SFL1346 and SFL1394

LIST OF TABLES

Table 1.1. Biochemical characteristics of species within the genus *Shigella*

Table 2.1. List of *E. coli* strains used in this study

Table 2.2. List of *S. flexneri* strains used in this study

Table 2.3. List of cloning vectors and plasmids used in this study

Table 4.1. Genetic features of phage SfV the 10.1 kb *Bam*HI fragment A *orf*s

Table 4.2. Similarity between the 10.1 kb *Bam*HI fragment A ORFs and database proteins

Table 4.3. Comparison of molecular weight and isoelectric point among phage SfV, DT1and λ

Table 6.1. Analysis of SFL124, SFL1339 and their derivative lysogen strains

Table 6.2. Analysis of putative cured derivatives of EW595/52

ABBREVIATIONS

aa	amino acid
ACL	antigen carrier lipid
Amp	ampicillin
ANGIS	Australian National Genomic Information Service
AGE	agarose gel electrophoresis
APC	antigen presenting cell
att	attachment site
bgt	bactoprenol-glucosyl transferase
bp	base pairs
BSA	bovine serum albumin
cfu	colony forming units
DNA	deoxyribonucleic acid
EDTA	ethylene diamine tetracetic acid
Gal	galactose
Glu	glucose
GluNAc	N-acetyl glucosamine
Gp	gene product
gtr	glucosyl transferase
ics	intracellular spread gene
IFN	interferon
Ig	immunoglobulin
IHF	integration host factor

IL	interleukin
int	integrase
int'	partial integrase
Ipa	invasion plasmid antigen
IPTG	isopropylthiogalactoside
Kan	kanamycin
kb	kilobase
kDa	kilodaltons
LB	Luria-Bertani medium
LMW	low molecular weight
LPS	lipopolysaccharide
μ	micro
MALT	mucosa associated lymphoid tissues
M cells	microfold cells
MCS	multiple cloning site
mM	millimolar
MW	molecular weight
MOI	multiplicity of infection (pfu/cfu)
NK	natural killer
nt	nucleotide
OD	optical density
orf	open reading frame
orf'	partial open reading frame
PAGE	polyacrylamide gel electrophoresis

PBS	phosphate buffered saline
PCR	polymerase chain reaction
pfu	plaque forming units
PEG	polyethylene glycol
pI	isoelectric point
PMN	polymorphonuclear cell
RE	restriction enzyme
Rha	rhamnose
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SfII	serotype-converting phage containing *gtrII* locus
SfV	serotype-converting phage containing *gtrV* locus
SfX	serotype-converting phage containing *gtrX* locus
Sf6	serotype-converting phage containing *oac*
SIgA	secretory immunoglobulin A
SIIDC	Swedish Institute for Infectious Disease Control
TBE	Tris-Borate-EDTA buffer
TE	Tris-EDTA buffer
thrW	threonine tRNA gene
UV	ultraviolet radiation
V	volts
WHO	World Health Organisation

I INTRODUCTION

CHAPTER 1

Introduction

Shigellosis or bacillary dysentery, is a disease of major public health concern particularly in developing countries where the mortality rate is high [Kotloff, 1999 #193]. The World Health Organization (WHO) has estimated an annual morbidity of 200 million people worldwide with at least 650,000 deaths among young children [WHO, 1997 #96]. *Shigella* species are the etiological agent isolated in approximately 10-20% of acute diarrhoeal episodes worldwide [Kotloff, 1999 #193]. Children under five years of age are most susceptible due to undeveloped immune response against *Shigella*, exacerbated by the effects of malnutrition, poor sanitation and lack of clean water [Black, 1993 #43][Cohen, 1991 #528]. Adding to this problem is the increasing level of antibiotic resistant strains isolated in many countries [Brito-Alayon, 1994 #531][Lin, 1992 #532]. As early as 1969, severe outbreaks of *Shigella dysenteriae* type 1 carrying plasmids conferring resistance to tetracycline, streptomycin, chloramphenicol and sulphonamides have been reported [Mata, 1969 #533]. Due to the urgency of the problem, WHO has been focusing on the development of an effective vaccine against *Shigella* [WHO., 1987 #534]. Mass immunisation against *Shigella* would be an economical alternative to the prevention of Shigellosis, if widely and properly administered. Recently, a live attenuated *Shigella flexneri* Y strain (SFL124) with an *aroD* gene deletion, was developed and has been shown to be effective among simian and human volunteers [Karnell, 1992A #17][Karnell, 1992B #81]. This was followed by studies involving the use of SFL124 in

the development of a recombinant vaccine strain which would express various *S. flexneri* serotypes, after realising that immunity to *S. flexneri* is serotype specific [Lindberg, 1993 #13][Hale, 1992 #535][Mavris, 1997 #16][Verma, 1993 #116][Huan, 1995 #15][Guan, 1998 #11][Adhikari, 1999 #112][Adams, 2001 #523].

1.1. Genus *Shigella*

Shigella is a gram-negative, non-motile, nonsporulating, bacteria which can be either an aerobe or facultative anaerobe. Their cultural and biochemical characteristics indicate that they are most closely related to *E. coli* [Koneman, 1988 #363]. The genus *Shigella* is divided into four main species based on antigenic and biochemical characteristics. These are *S. dysenteriae* (subgroup A), *S. flexneri* (subgroup B), *S. boydii* (subgroup C), and *S. sonnei* (subgroup D). Some of the more important distinguishing biochemical chareacteristics of the four *Shigella* species are presented below (Table 1.1).

Table 1.1. Biochemical Characteristics of Species within the genus *Shigella*

(Modified from Koneman, 1988)

	S. dysenteriae	*S. flexneri*	*S. boydii*	*S. sonnei*
Serogroup	**A**	**B**	**C**	**D**
Glucose	+	+	+	+
Lactose	-	-	-	-

	S. dysenteriae	*S. flexneri*	*S. boydii*	*S. sonnei*
Raffinose	-	+/-	-	-
Sucrose	-	-	-	-
Xylose	-	-	+/-	-
Indole production	+/-	+/-	+/-	-
β-galactosidase	-	-	-	+
Ornithine decarboxylase	-	-	-	+

1.2. Shigellosis

Shigellosis or bacillary dysentery is an acute infectious disease of the large bowel. The most common early symptoms are fever, watery diarrhea and cramping abdominal pain [Carpenter, 1982 #364]. Fluid and electrolyte losses also occur due to the enterotoxin acting on the intestinal cells. The dysentery phase of illness then follows, this is characterised by the presence of blood, pus, and mucus in the feces and the onset of stool straining (tenesmus), suggesting penetration of bacteria into the colonic epithelium. Patients excrete the organisms in the feces throughout the acute phase of the disease and for about 4 weeks post recovery. This poses a potential source of infection to others since the principal cause of *Shigella* infection is food or drinking water contaminated with intestinal discharges of infected patients [Mosley, 1962 #365].

Shigella can initiate infection from a small infective dose of 10-100 organisms and spread easily in situations where sanitation or personal hygiene is poor [Mims, 1998 #366]. *S. dysenteriae* is the most virulent serotype since it produces a toxin called Shiga toxin which has enterotoxic, cytotoxic and neurotoxic effects [Olsnes, 1980 #367]. Similar toxins are also produced by some strains of *S. flexneri*, *S. sonnei*, and *E. coli* strains expressing Shiga-like toxins, that are related antigenically to Shiga toxin [Keusch, 1977 #368][O' Brien, 1982 #369]. *S. sonnei* may also undergo antigenic shifting from form I (S) to form II (R), this effects the somatic antigen and is attributed to the presence of a 120 megadalton plasmid [Kopecko, 1980 #370].

1.3. Epidemiology

Shigella infection accounts for 200 million diarrheal cases annually, with children under the age of 5 being the most susceptible [WHO, 1997 #96]. In fact, a recent study reported that 69% of all episodes and 61% of all deaths caused by shigellosis involved children under 5 years of age [Kotloff, 1999 #193].

Bacillary dysentery occurs in most parts of the world but with greatest frequency in Asia, Africa and Central America where it is endemic and occasionally epidemic [Mata, 1969 #373][Rahaman, 1975 #374][Gross, 1979 #375][Khan, 1985 #383][El-Rafie, 1990 #384][Adkins, 1987 #385][Black, 1982 #386][Katouli, 1990 #387]. *S. sonnei* was the predominant organism isolated from Asia, Africa, and Central America and is also

common in diarrheal disease in industrialised countries like the United States [Black, 1978 #376]. In industrialised countries, *S. boydii, S. dysenteriae*, *S. flexneri*, and *S. sonnei* were isolated in 2%, 1%, 16% and 77%, respectively, while for developing countries the percentage of cases was divided at 6%, 6%, 60% and 15%, respectively [Kotloff, 1992 #89]. The most predominant serotype of *S. flexneri* is the 2a type. Other serotypes were also isolated such as 1b, 3a, 4a and 6.

Epidemiological surveys in developing countries have shown increasing isolation of *Shigella* strains which are resistant to antimicrobial therapeutic drugs such as sulfonamide, tetracycline, ampicillin, and trimethoprim-sulfamethoxazole [Sack, 1997 #388]. For example, outbreaks of *Shigella dysenteriae* serotype 1 in developing countries have been observed with the emergence of multi-antibiotic resistant strains, due to indiscriminate use of antimicrobial agents [Macaden, 1980 #378][Murray, 1986 #379]. Therefore, development of an efficacious *Shigella* vaccine against the common serotypes would assist the existing prevention measures and antibiotic treatment which cannot adequately control shigellosis.

1.4. Pathogenesis

The initial step in *Shigella* pathogenesis is bacterial invasion via penetration of the colonic mucosa. The point of entry is the gut membranous epithelial cells (M cells) in the follicular epithelium that overlie the lymphoid follicles [Sansonetti, 1999 #389][Neutra,

1999a #188]. M cells are the preferred adhesion and transport sites of microorganisms rather than the apical pole of the gut epithelial lining since the former have scarce mucus covering, no brush border microvilli and have poor production of glycocalyx which allows enhanced physical contact between the pathogen and the M cell membrane [Neutra, 1999b #390][Jepson, 1998 #391]. Following adhesion to M cells, actin and myosin accumulate at the site of attachment in preparation for the internalisation of *Shigella* by phagocytosis [Clerc, 1987 #392]. An alternative pathway by which the bacteria can gain entry is by insertion of *Shigella* through the basolateral side of epithelial cells [Perdomo, 1994a #41]. Responsive polymorphonuclear neutrophils (PMN) squeeze through adjacent cells creating a gap sufficient for *Shigella* entry into the basolateral side of the epithelium (Figure 1.1).

The next step in the invasion process involves lysis of the phagocytic vacuole membrane, releasing the bacteria into the cytoplasm where it rapidly reproduces [Sansonetti, 1986 #339][Kadurugamuwa, 1991 #393]. Bacteria located at the top of tightly packed actin filaments form protrusion which extend and penetrate into an adjacent cell, thus allowing passage of *Shigella* intercellularly without their release into the extracellular medium [Prevost, 1992 #394]. The presence of actin in cell movement was confirmed when treatment with cytochalasin D prevented polymerisation of actin monomers into filaments that also prevented *Shigella* spread, and when trails of labelled F-actin was shown at one pole of the bacteria [Bernardini, 1989 #135]. This movement of *Shigella* was termed Ics for intra- and intercellular spread, and is aided by the collection of actin

Figure 1.1. *Shigella* invasion of the colonic epithelium. A, *Shigella* enters through the basolateral side with the assistance of polymorphonuclear cells (PMN, dark blue) or through the apical M cell (light blue cell). B, *Shigella* survive and escape PMNs and macrophages of the lymphoid follicle by inducing apoptosis. C. *Shigella* can gain entry through the basal side by cell phagocytosis. D. *Shigella* move around the cell cytoplasm with the aid of actin fibers (intracellular spread) and, E. migrate to adjacent cells through membrane protrusions (intercellular dissemination). F, Invading bacteria instigate apoptosis leading to ulcerations of the intestinal epithelium. (Adapted from Pedromo *et al.* 1994 and Sansonetti, 1999a).

filaments at one pole of the bacteria, as has been similarly observed for *Listeria monocytogenes* [Sansonetti, 1991 #82][Tilney, 1990 #395].

Phagocytosis of invasive *Shigella* by macrophages results in programmed cell death of the macrophage [Zychlinsky, 1992 #397]. Studies involving infection of Peyer's patches in a rabbit ligated loop model, showed numerous apoptotic cells when an invasive *Shigella* is inoculated compared to a non-invasive mutant, in which case apoptosis is only detected at background level [Zychlinsky, 1996 #396]. IpaB (invasion plasmid antigen B) is required for macrophage killing owing to its capacity to bind interleukin-1β (IL-1β) converting enzyme (ICE or caspase 1)[Chen, 1996 #398]. Binding of IpaB to the cysteine protease caspase 1 elicits cleavage of pro-IL-1β, engagement of apoptosis and release of mature IL-1β and other proinflammatory cytokines like TNFα, IL-6 and IL-1 [Sansonetti, 1995 #399].

Inflammation enhances polymorphonuclear leukocyte migration through the epithelium, loosening cellular junctions and promoting *Shigella* entry [Perdomo, 1994b #55]. Apical *Shigella* also stimulate influx of basal PMNs by disrupting the epithelial lining and allowing basolateral invasion of epithelial cells [Beatty, 1997 #400]. Invasion of colonic epithelial cells triggers the expression of additional pro-inflammatory molecules (chemokines) such as IL-8 which aggravates inflammation by attracting more PMNs. This cycle of invasion and inflammation could explain the severity of colonic epithelium destruction which is characteristic of Shigellosis [Rout, 1975 #401].

1.5. Virulence Factors.

1.5.1. *Plasmid-encoded Virulence Factors*

Shigella carry a large 220 kb virulence plasmid containing a 30 kb sequence [Maurelli, 1985 #343] that encodes genes essential in the expression of invasive phenotype [Sansonetti, 1982 #349][Sansonetti, 1983 #402]. The *mxi* (membrane expression of antigens) and *spa* (secretion of protein antigens) operons encode type III secretion proteins that assemble into a flagellum-like structure and deliver *Shigella* effector proteins from the bacterial cytoplasm directly onto the surface or into the cytoplasm of epithelial cells [Menard, 1994 #346]. About 20 proteins are secreted through this assembly. The first set includes IpaB, IpaC and IpaD, proteins necessary for bacterial entry into the host cell [Clerc, 1987 #392]. The flow of proteins through the secretion tube is regulated by IpaB and IpaD. IpaB is also involved in forming a complex with IpaC which inserts into the eukaryotic cell membrane, creating a pore that serves to induce actin polymerisation via IpaC and inject proteins into the cell cytoplasm [Menard, 1993 #61][Menard, 1994 #346]. The second set of proteins secreted by the type III secreton system include IpaA. IpaA can bind vinculin once injected into the cell, mediating maturation of the entry focus composed of a dense meshwork of actin filaments [Sansonetti, 1992 #24][Nhieu, 1997 #19]. The compound Congo red was identified to fully activate the *mxi-spa* secretory system [Bahrani, 1997 #403].

Other proteins encoded by the virulence plasmid are SepA, a serine protease which enhances inflammation of infected tissues and IcsA also known as VirG [Makino, 1986 #404]. IcsA is a 120 kDa surface protein localised at one pole of the bacterial cells that helps to induce actin-dependent motility by binding vinculin and promoting the movement of the F-actin comet tail [Bernardini, 1989 #135]. The complete genetic composition and protein products of *S. flexneri* virulence plasmid pWR100 were recently characterised. The study reported on 25 proteins secreted by the type III secretion apparatus and discovered that pWR100 had blocks of genes which were traced to be initially carried by four different plasmids [Buchrieser, 2000 #536].

1.5.2. *Chromosome-encoded Virulence Factors*

There are several *Shigella* chromosomal virulence genes important for survival in the intestinal tissues and those that can regulate the expression of plasmid-encoded virulence genes. For example, the *iucABCD* and the *iutA* genes encode for the synthesis of aerobactin and its outer membrane receptor, respectively [Lawlor, 1984 #405][Griffiths, 1985 #406]. Aerobactin is a hydroxamate siderophore used by *Shigella flexneri* for the transport of iron and whose role in virulence was studied in transposon-induced mutants [Lawlor, 1987 #407]. The *sodB* gene has a critical role in *Shigella* pathogenicity, since it encodes for a superoxide dismutase that protects *Shigella* from oxygen toxicity. *sod*B

mutants were observed to be negative in the Sereny test and produced little detectable damage in ligated loops [Franzon, 1990 #408].

Other genes have been identified following transposon mutagenesis. The *virR* gene was found to be involved in temperature regulated invasion. Transduction of the virR:Tn10 mutation to wild-type *Shigella* resulted in a strain that was invasive at both 30°C and 37°C. The invasion genes were expressed at 37°C but not at 30°C in the wild-type strain [Maurelli, 1988 #409][Maurelli, 1984 #410]. The *virR* gene encodes for the histone-like protein H1, which induces change in DNA supercoiling thereby modulating gene expression [Hulton, 1990 #411]. Transposon insertion in the virulence associated chromosomal gene, *vacB* resulted in decreased invasion and absence of intercellular spread. This was attributed to decreased amounts of IcsA and IpaB in the *vacB* mutants [Tobe, 1992 #412].

The *S. flexneri* loci required for the biosynthesis of lipopolysaccharide (LPS) are alleles of the *E. coli rfb* and *rfa* loci involved in the expression of the LPS basal core and group somatic antigen, respectively [Formal, 1970 #413][Okada, 1991b #414][Schnaitman, 1993 #110]. LPS basal core synthesis, as specified by the *rfa* locus, starts with 2-keto-3-deoxyoctonate-lipid A and to this, L-glycero-D-phosphate, D-glucose, D-galactose, and N-acetyl-D-glucosamine are sequentially added [Watanabe, 1984b #415]. Transposon insertion in the *rfa* locus increases the mobility of the core constituent of smooth *S. flexneri* 2a LPS, resulting in a mutant exhibiting delayed plaque formation and

Sereny reaction [Okada, 1991a #22]. This suggests that the structure of the LPS core affects the virulence of *S. flexneri*. The *rfb* locus encodes rhamnose synthetase, rhamnose transferase and N-acetylglucosamine transferase, which are all required for the synthesis of the O-antigen repeating unit of the group 3,4 or Y variant antigen of *S. flexneri* serotypes 1a, 2a and 5a, respectively [Simmons, 1971 #416].

The individual repeat units of O-antigen are transferred to a carrier lipid and polymerised by the *rfc* gene product. The completed O-specific side chain is finally attached to the basal core structure by a translocase encoded by genes from both *rfa* and *rfb* loci [Simmons, 1987 #2]. The chromosomal T locus was discovered as the integration site for incorporation of lysogenic phage encoding type-specific conversion genes. Invasion experiments on Y variants expressing only the 3,4 antigen showed decreased virulence [Timokov, 1970 #419].

Other chromosomal loci associated with *Shigella* virulence include the *kcp*A locus responsible for the regulation of bacterial motility plasmid gene *virG* (*icsA*) [Yamada, 1989 #417], and the Shiga toxin *stx* gene. Mutants lacking *kcpA* showed a negative Sereny test with limited intracellular and intercellular bacterial spread in tissue culture monolayers [Makino, 1986 #404]. Conversely, the *stx* gene unique to *S. dysenteriae* 1 encodes a potent cytotoxin that binds to Galα1-4 Galβ (galabiose) glycolipid receptors. This binding results in the inhibition of mammalian protein synthesis due to cleavage of the N-glycosidic bond at adenine 4324 in 28S rRNA [Lindberg, 1987 #418]. Mutations

in the *stx* gene leads to a decrease in vascular damage in the colonic epithelium of orally challenged monkeys [Sekizaki, 1987 #420]. Also, culture filtrates of *S. dysenteriae* 1 have enterotoxigenic activity that causes fluid accumulation in ligated rabbit ileal loops [Eiklid, 1983 #421].

1.6. Immune Responses to Shigellosis

LPS O-antigens and the invasion plasmid antigens of the outer membrane surface have a prominent role in stimulating the immune system. This has been demonstrated in endemic countries where significantly higher anti-LPS and anti-Ipa serum titers were recorded in children between 2-5 years of age [Oaks, 1986 #63]. These strong LPS and Ipa-specific antibody responses ware also observed among currently infected dysenteric patients [Lindberg, 1984 #422][Oberhelman, 1991 #423]. Patient convalescent sera contained antibodies recognising several plasmid-encoded polypeptides such as the 120 kDa VirG protein product, the 78 kDa IpA, 62 kDa IpaB, 43 kDa IpaC and the 38 kDa IpaD protein [Hale, 1985 #345].

1.6.1. *Mucosal Humoral Immunity*

The active "coproantibodies" of the intestinal secretory immune system were first observed in rabbits protected against dysentery after oral immunisation with killed *Shigella dysenteriae* 1 [Brandtzaeg, 1989 #425]. These coproantibodies were identified

in 1965, as the principal determinant of human humoral mucosal immunity and are known as secretory immunoglobulin (sIgA) [Brandtzaeg, 1989 #425]. Synthesis of sIgA has been shown to be triggered in response to several bacterial and viral antigens [Tomasi, 1965 #426]. About half of the total antibody producing cells of the lymphoid system resides in the submucosa and an estimated 20-30 IgA cells per IgG cell were found in the human gut mucosa [Mims, 1995 #427]. Prior to mucosal secretion, the sIgA dimer is complexed with a secretory piece that protects against proteolytic degradation. During mucosal infection, the submucosal sIgA enters the blood via the lymphatics to give increased serum IgA levels [Roitt, 1997 #526].

Local immunity to *Shigella* infection is characterised by the appearance of circulating antibody secreting cells in the peripheral blood and the production of *Shigella* specific sIgA, which can block interactions between the bacteria and the mucosa [Reed, 1971 #39][Oberhelman, 1991 #423][Keren, 1989a #37][Keren, 1989b #428]. Specific antibodies could also mediate antibody-dependent cellular immunity against *Shigella* [Lowell, 1980 #73]. In the mucosal gut-associated lymphoid tissue (GALT), precursor B and T lymphocytes are stimulated after exposure to antigen by antigen presenting cells (APC) and then migrate via the efferent lymphatics through to the systemic circulation. The lymphocytes enter the mucosal effector sites such as the lamina propia of the intestine, bronchi and the genitourinary tract, guided by homing receptor molecules. At these effector sites, B cells clonally expand and mature into IgA secreting plasma cells. This migration from IgA inductive tissues such as the GALT and MALT (mucosa

associated lymphoid tissue) to IgA effector sites is called the common mucosal immune system, and can be primed by oral or intranasal antigen administration [Roitt, 1997 #526]. Local anti-*Shigella* antibody responses have also been detected in intestinal secretions of infected monkeys [Dinari, 1987 #44], dysenteric patients [Oberhelman, 1991 #423] and in colostrum, breast milk, bile, saliva, and tears [Achi, 1992 #430][Schultz, 1992 #431][Cleary, 1991 #125][Cleary, 1989 #432].

1.6.2. *Systemic Humoral Immunity*

Unlike IgA, IgG antibodies generated through natural infection or vaccination have been observed to be serotype-specific [Li, 1994 #429][Cohen, 1991 #38][Cohen, 1992 #433]. This stresses the dominant role of LPS as the primary target antigen for protective immunity and is supported by studies performed on blood and fecal samples which show significantly high *Shigella* LPS-specific responses [Cohen, 1988 #62][Lindberg, 1991 #434][Achi, 1994 #435][Cam, 1993 #436]. Also, antibody levels are observed to be kept elevated in endemic population due to repeated exposure to the microbe, resulting in subclinical infection and boosting the immune response. Studies have reported that protective host immune response to *S. flexneri* is raised against the O-antigen component of the LPS stressing that immune response to O-antigen is serotype-specific and protects against re-infection with an organism possessing the same serotype [Brahmbhatt, 1992 #51][Hale, 1992 #9][Lindberg, 1993 #13].

According to Lindberg and Pal [Lindberg, 1993 #13], the role of Ipa-specific antibodies in host defence against dysentery is not clear since various trial results for Ipa-specific responses are not consistent. For example, very weak anti-Ipa responses were observed in monkeys vaccinated with an invasive *E. coli* strain expressing *S. flexneri* antigens [Oaks, 1986 #63]. However, in another study, strong responses comparable to infections with wild-type strain were noted in monkeys vaccinated with *S. flexneri* vaccine candidate strain that is unable to spread intercellularly [Sansonetti, 1991 #82][Sansonetti, 1989 #437]. Interesting findings were also noted during immunogenicity trials using an attenuated Δ*aroD S. flexneri* Y strain, it was observed that significant LPS specific responses were evident with or without previous *Shigella* infections while only those volunteers with history of bacillary dysentery infection showed Ipa-specific serum responses [Li, 1992 #438]..

1.6.3. *Cell-Mediated Immunity*

There are few reports regarding cell-mediated immunity (CMI) to shigellosis. Two well-studied CMI responses include the proliferative T-lymphocyte reaction induced by *S. flexneri* antigens [Zwillich, 1989 #439] and the sensitivity of *Shigella* infected HeLa cells to natural killer (NK) cell activity. In the latter case, NK cells were shown to be active against *S. flexneri* infected cells but not against uninfected cells [Klimpel, 1986 #328]. Conversely, among volunteers vaccinated with the polysaccharide part of the homologous LPS molecule, significant peripheral lymphocyte proliferation was noted

[Li, 1992 #438]. A similar increase in intraepithelial T cell number was also seen in guinea pigs infected with *Shigella* [Sinha, 1992 #440].

Shigella induces interferon production in infected fibroblasts [Hess, 1989 #64] and inhibits cell invasion by *Shigella* [Niesel, 1986 #66]. Interferon gamma (IFN-γ) has also been shown to suppress the intracellular multiplication of *Shigella in vivo,* while IFN-β mediated HeLa cell resistance to *Shigella* infection [Gober, 1972 #441][Hess, 1987 #442]. It is still uncertain whether the cell mediated immune response by T cells and NK cells are as effective in protecting against dysentery as is the case in viral infections or against tumor cells, however, the observation leading to increased frequency and severity of shigellosis among AIDS patients indicates that $CD4^+$ T cells are important in protecting against *Shigella* infection [Baskin, 1987 #443][Blaser, 1989 #444].

1.7. Somatic Antigen (O-antigen) of the Lipopolysaccharide

1.7.1. *Structure and Function*

The somatic antigen (O-antigen) is the outermost component of the outer structural cell surface layer of Gram-negative bacteria, the lipopolysaccharide (LPS) [Makela, 1984 #121](Figure 1.2). The O-antigen is composed of oligosaccharide repeat units linked to the LPS lipd A component by the core polysaccharide. Lipid A is the innermost hydrophobic component of the LPS. The repeating tetrasaccharide unit of the O-antigen

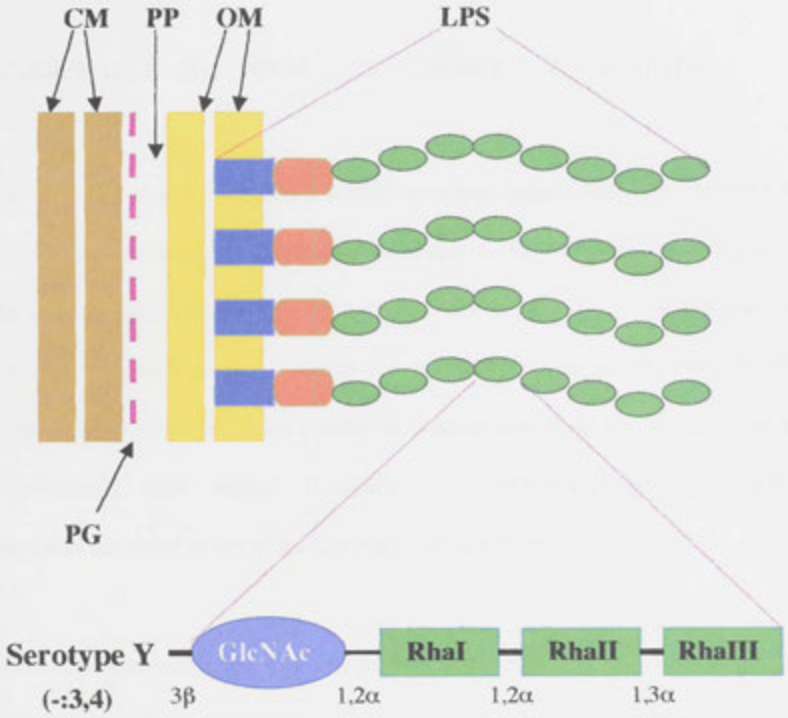


Figure 1.2. Bacterial lipopolysaccharide (LPS). LPS is present in most Gram-Negative bacteria and comprises the lipid A (blue), core polysaccharide (red) and O-antigen (green). The bacterial cell wall components are shown by arrows with their names abbreviated: CM, cytoplasmic membrane; PG, peptidoglycan; PP, periplasm; OM, outer membrane. Each O-antigen unit is a linear polymer of four sugar groups (Rha, rhamnose; GlcNAc, N-acetylglucosamine) forming the basic O-antigen serotype Y (adapted from Reeves, *et al.*, 1996 and Simmons and Roma-Nowska, 1987).

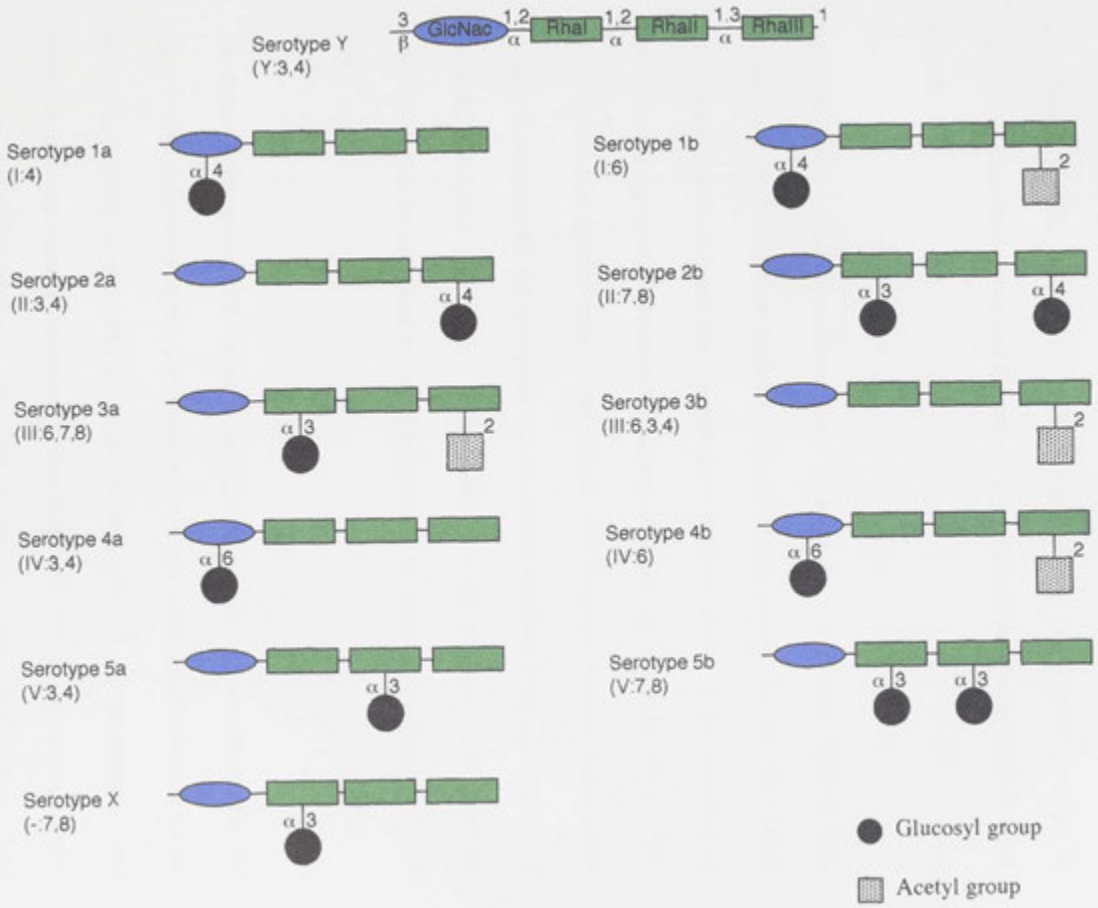
is representative of *S. flexneri* serotype Y and has the basic configuration [Simmons, 1987 #2]:

-3)-β-D-**GlcNAc**-(1→2)-α-L-**RhaI**-(1→2)-α-L-**RhaII**-(1→3)-α-L-**RhaIII**-(1

Variations in type-specific and group-specific antigenic determinants are attributed to either glucosyl or O-acetyl groups being added to specific sugar residues of the repeating unit in a linkage-specific manner [Kenne, 1977 #312](Figure 1.3). For instance, *S. flexneri* serotype Y with group 3,4 antigen was converted to group 6 antigen by *Shigella* bacteriophage SF6 which possesses transferase proteins needed for the O-acetylation of the rhamnoseIII sugar residue [Lindberg, 1972 #445][Lindberg, 1978 #120]. Glucosylation can occur to any of the four sugar residues [Gemski Jr., 1975 #113].

The somatic antigen plays a critical role in *Shigella* virulence. Early experiments on bacterial pathogenesis produced rough *S. flexneri* derivatives (devoid of O-antigen) which were avirulent [Okamura, 1977 #446][Okamura, 1983 #447]. Rough strains were able to invade tissue culture cells and multiply intracellularly but had lost the ability to spread to adjacent HeLa cells and failed to elicit keratoconjunctivitis in the Sereny test. It has been demonstrated that *S. flexneri* O-antigen is able to facilitate adhesion to guinea pig intestinal mucosa *in vitro* [Izhar, 1982 #448]. Izhar *et al.*, (1982) proposed the role of the O-antigen polysaccharide in bacterial colonisation of the colonic epithelium although it is not required for cell penetration. This is supported by experiments involving mutated

Figure 1.3. Chemical composition of O-antigens of the different *Shigella flexneri* serotypes. Modification results from the glucosylation and O-acetylation of the sugar residues N-acetylglucosamine (GlcNac, green) and rhamnose (Rha, blue) comprising the the basic serotype Y O-antigen tetrasaccharide repeat unit. Type-specific (Roman numerals) and group-specific (arabic numeral) antigenic determinants are enclosed in parenthesis.


Serotype Y
(Y:3,4)
3
β
GlcNac
1,2
α
RhaI
1,2
α
RhaII
1,3
α
RhaIII
1
Serotype 1a
(I:4)
α 4
Serotype 1b
(I:6)
α 4
2
Serotype 2a
(II:3,4)
α 4
Serotype 2b
(II:7,8)
α 3
α 4
Serotype 3a
(III:6,7,8)
α 3
2
Serotype 3b
(III:6,3,4)
2
Serotype 4a
(IV:3,4)
α 6
Serotype 4b
(IV:6)
α 6
2
Serotype 5a
(V:3,4)
α 3
Serotype 5b
(V:7,8)
α 3
α 3
Serotype X
(-:7,8)
α 3
Glucosyl group
Acetyl group

S. flexneri LPS. Okada *et al.*, (1991) observed that mutants with altered core structure, no O-antigen, or with decreased length of O-polysaccharide were capable of invading MK2 cells, and multiplying and spreading within infected cells but were unable to spread intercellularly [Okada, 1991a #22]. Okada suggested that the *Shigella* O-antigen is required for intercellular spread and for protecting the bacteria from nonspecific host cell defences.

The O-antigen is critical for virulence since rough strains with exposed lipid A were susceptible to the host defense molecule, the LPS-binding protein [Tobias, 1989 #455]. This plasma protein enhances the crosslinking of bacterial lipid A to the receptors on mononuclear phagocytes and stimulates synthesis of tumor necrosis factor alpha (TNF-α), IL-1 and IL-6 lymphokines. Furthermore, the O-antigen polysaccharide chains function as receptors for the adsorption and invasion of many bacteriophages capable of mediating O-antigenic conversion. This contributes to the diversity of the O-antigen enabling it not to be recognised by antibodies secreted by memory cells [Lindberg, 1973 #456].

The defensive function of LPS was also observed in *Salmonella*. Studies showed that the more virulent *Salmonella* serotype BO does not influence complement activation via the alternative pathway as efficiently as the less virulent DO and CO serotypes. This results in decreased bacterial opsonisation, thus Salmonella BO can evade phagocytosis [Pluschke, 1983 #450][Penn, 1983 #451][Goldman, 1984 #452][Joiner, 1986

#453][Timmis, 1985 #454]. This is a virulence process which may also occur in *Shigella's* counteraction against host defence.

1.7.2. *Biosynthesis*

There are several gene loci in the chromosome that code for the expression of factors necessary for the synthesis of the core and O-antigen polysaccharide. The *rfa* gene cluster encodes factors necessary for LPS core biosynthesis while the *rfb* cluster is involved in O-antigen synthesis and attachment [Bachmann, 1990 #457][Jiang, #458]. The O-antigen monomers are not linked directly to the core polysaccharide. Instead, they are primarily polymerised on a lipid carrier by enzyme products of the *rfb* gene cluster [Whitfield, 1995 #3]. The reactions commence with an N-acetyl-glucosamine-I-phosphatetransferase (Glcρ NAc-I-phosphatetransferase) initiating enzyme Rfe [Meier-Dieter, 1992 #459] that transfers GlcNAc residues found in the O-polysaccharide units of *Shigella dysenteriae* type 1 and *S. flexneri* [Klena, 1993 #460][Yao, 1994 #157]. The *rfb* encoded nucleotide diphosphate sugar synthetases produce activated sugar derivatives which are transferred by glycosyl transferase enzymes to the antigen carrier lipid (ACL) forming the lipid-linked O unit [Robbins, 1971 #461]. Undecaprenol phosphate (und-P) is the lipid carrier derived from C_{55}-polyisoprenoid [Osborn, 1972 #462]. Lipid linked O antigens are sequentially assembled in a blockwise fashion by the polymerase encoded by the *rfc* locus [Collins, 1991 #463]. The Rol protein (regulator of O-chain length) controls the addition of O-antigens to the growing chain by modulating the activity of *rfc* to shift

between its functions of polymerisation and the transfer of O-antigens to the RfaL for ligation [Batchelor, 1991 #464][Bastin, 1993 #465][Morona, 1994 #466].

The *rfb*T and *rfa*L gene products catalyse the ligation of the lipid-linked, polymerised O-polysaccharide to the lipid-A core [Gemski Jr., 1967 #467]. The construction of the LPS molecule occurs on the periplasmic face of the inner membrane [McGrath, 1991 #468] followed by its secretion into the outer leaflet of the outer membrane phospholipids [Raetz, 1990 #469][Lugtenberg, 1983 #470].

1.8. Bacteriophage mediated O-antigen Modification

1.8.1. *Temperate Bacteriophages*

Bacterial evolution has been riddled with the acquisition and loss of properties which are primarily attributed to the exchange of genetic information between various agents. One notable carrier of genetic elements is a bacteriophage, whose lysogenic capabilities provide opportunity for horizontal transfer of genes from one host to another [Calendar, 1998 #302]. Temperate bacteriophages can integrate their DNA into bacterial chromosomes [Campbell, 1993 #472]. In the lytic pathway, the phage infects and multiplies in a suitable host to produce several hundred progeny particles which are released when the bacterial cell lyses. The bacteriophage can assume a latent proviral role in the lysogenic pathway after the integration of its DNA into the circular host chromosome by a site-specific recombination event [O' Gorman, 1991 #473][Stark, 1992 #474]. The integrase

enzyme catalyses a recombination process by binding tightly to a specific DNA sequence on the circular bacteriophage genome resulting in a complex that can now bind to a homologous DNA sequence on the bacterial chromosome. The integrase then catalyses the DNA cutting and resealing reactions around a short region of sequence homology to form a tiny heteroduplex joint at the point of union [Alberts, 1994 #475](Figure 1.4). Several studies have been focused on investigating the molecular mechanism involved in lysis-lysogeny decision [Herkowitz, 1980 #303][Rosner, 1972 #304]. Bacterial host tRNA genes are the preferred sites for bacteriophage chromosome integration because of their presence in multiple copies, thus providing multiple integration sites. The conserved secondary structure of tRNA genes may contain important sequences recognised by enzymes involved in recombination [Reiter, 1989 #476][Marschalek, 1994 #477]

Lysogenic phage DNA is inserted at a specific host chromosomal site, *attB*, by the phage-encoded integrase (Int) [Weisberg, 1983 #485]. The substrates for this reaction are 240 bp of phage DNA specific sequence flanking the crossover point, and a 21 bp host DNA segment. The process involves the N-terminal of integrase molecules binding to so-called "arm sites" which are close to the extremes of the 240 bp *attP* DNA. Then a host protein (integration host factor) binds between the arm sites and the crossover point, bending the DNA so that the catalytic sites on the C-terminal ends are positioned close to the crossover point [Kim, 1992 #486]. Int recognises weakly binding core sites in the 21 bp host sequence surrounding the crossover point. Crossover occurs by the formation of

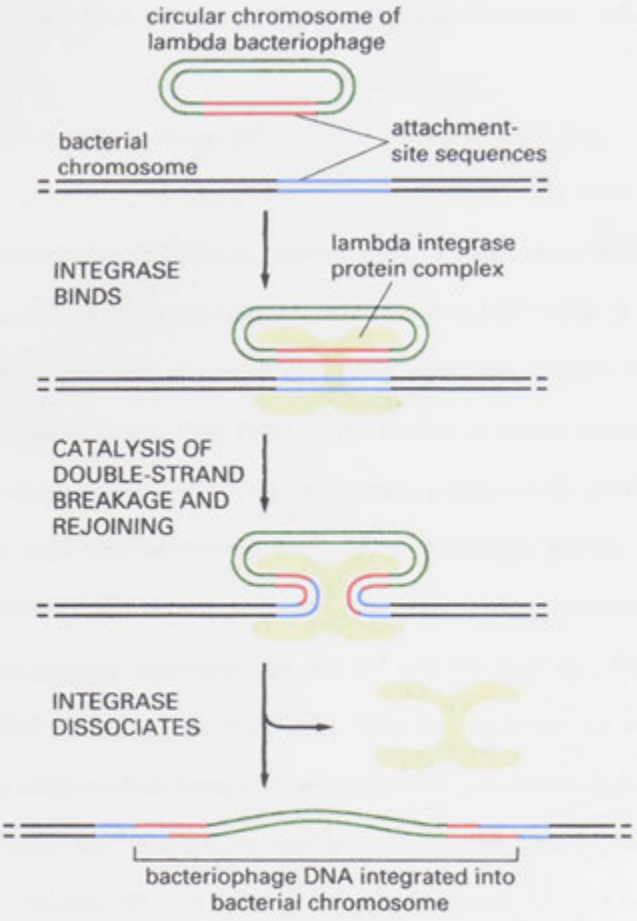


Figure 1.4. The insertion of bacteriophage lambda DNA into the bacterial chromosome. This is an example of site-specific recombination in which the lambda integrase enzyme binds to a specific attachment site sequence on each chromosomal strand, where it makes cuts that bracket a short homologous DNA sequence. The integrase thereby switches the partner strands and reseals them so as to form a heteroduplex joint 7 base pairs long. Reproduced from Alberts *et al.*, 1994.

a crossbridge. Phage DNA is excised from the lysogenic bacterial chromosome when repression is lifted. This process mediated by a phage-encoded protein, Xis, and occurs when Xis binds to λ DNA between the left arm and the core [Thompson, 1987 #487].

1.8.1.1. *S. flexneri Bacteriophage SfV*

Temperate bacteriophage SfV has an isometric head, a long non-contractile tail and belongs to the group B morphology class of Bradley [Bradley, 1967 #179]. It inserts its DNA through site-specific recombination and requires recombinase enzymes belonging to the integrase family [Argos, 1986 #306]. The process of genetic exchange occurs between the attachment site (*attP*) of the circular phage genome and the attachment site (*attB*) of the linear bacterial chromosome. This recombination between phage and bacterial genome occurs when short homologous common core sequences overlap, producing the prophage attachment sites *attL* and *attR* that flank the phage genome [Campbell, 1992 #307][Campbell, 1962 #308]. Other essential factors like the integrase (Int) and the integration host factor (IHF) participate in the recombination event, while excisionase protein (Xis) is more active when the prophage enters the lytic phase [Miller, 1980 #309]. In phage SfV, the *attP*, *xis* and *int* genes located on a 2.2 kb sequence cluster adjacent to the serotype conversion genes have been sequenced and characterised [Huan, 1997A #118]. In addition, the conserved order of this gene cluster is not unique to phage SfV. *Salmonella* Phage P22 and *S. flexneri* phage SfX and Sf6 also exhibited homologous cluster arrangement wherein the *attP* site is immediately adjacent to the

serotype conversion region [Clark, 1991 #119][Leong, 1985 #198][Poteete, 1988 #310][Leong, 1986 #311][Guan, 1999 #151][Verma, 1993 #116].

Another well-characterised region of the SfV genome is the serotype-conversion gene cluster located downstream of the *attP* site. These genes mediate the modification of the LPS O-antigen that creates variation in the O-antigenic specificity. With the exception of *S. flexneri* serotype 6, the O-antigen subunit of *S. flexneri* (serotype Y) consists of a basic repeating tetrasaccharide unit structure: [-3)-β-D-L-GlcpNAc-(1-2)-α-L-RhapI-(1-2)-α-L-RhapII-(1-3)-α-L-RhapIII-(1-] [Simmons, 1987 #2][Kenne, 1978 #111]. In accordance to the position of glucosyl or O-acetyl residues attachment to the repeating hexose unit, the antigenic type (I, II, III, IV, V) and group specificity (3,4; 4; 6; 7,8) of the strain may vary [Kenne, 1977 #312] (Figure. 1.3). The addition of glucosyl moieties in serotype V O-antigen is catalysed by the phage encoded glucosyl transferase enzyme, part of the early genes eliciting O-antigen conversion [Huan, 1997A #118]. The gene operon has been sequenced and contains three continuous open reading frames capable of serotype-converting *S. flexneri* serotype Y (group antigen 3,4) to serotype 5a expressing type V and group antigen 3,4 [Huan, 1997B #117]. O-acetylation of LPS by *S. flexneri* phage Sf6 was also observed to cause O-antigen conversion from group 3,4 antigen to group 6 [Gemski Jr., 1975 #113][Lindberg, 1978 #120][Verma, 1991 #76]. Recent findings have elucidated the role of the three-gene O-antigen glucosylation system in *S. flexneri* bacteriophage SfII and SfX [Mavris, 1997 #16][Guan, 1999 #151]. It has also been proposed that the first two of the three gene cluster be designated as $gtrA_{(type)}$ and

gtrB$_{(type)}$ due to their sequence homology to phages SfII, SfV and SfX. The third glucosyl transferase genes are divergent in primary nucleotide sequence despite their structural similarity [Guan, 1999 #151][Allison, 2000 #199].

The site-specific integration and the serotype–conversion regions discussed earlier are encoded in approximately 5 kb length of the SfV genome. At the left side of these genes is a 15 kb segment of the DNA which has been characterised recently (Allison *et al.*, submitted for publication). It was discovered to encode for the early regulatory genes involved in the phage repression, regulatory, superinfection immunity, and transcription termination systems. Comparative sequence analyses indicate that this 15 kb region of phage SfV shared numerous features of the sequences encoded in bacteriophages of the lambdoid family. Therefore, knowledge of the the molecular profile of phage λ is essential since it could provide a basis or a comparative example in the analysis of unsequenced phage SfV DNA.

1.8.1.2. *Coliphage Lambda (λ)*

Bacteriophage λ -encoded genes have a variety of functions [Campbell, 1994 #164]. For example, upon infection with lambdoid phages, RNA polymerase begins the leftward and rightward transcription from the *pL* and *pR* promoters, respectively (Figure 1.5). The immediate transcription products are the *N* protein on the left and the *cro* protein on the right. The gene product (gp) *N* is able to override ordinary termination signals such

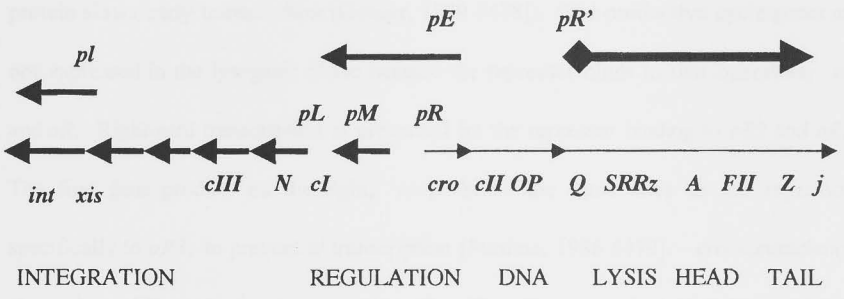


Figure 1.5 . Map of λ prophage showing major gene clusters with some of the genes they contain and major transcripts. Antiterminated transcripts contain arrowheads at terminator sites. The late regulator Gene Q appears between clusters. DNA recognition sites are not shown. Adapted from Campbell, 1994.

as *tL1* and *tR1* allowing transcription to proceed throughout the λ genome [Campbell, 1994 #164]. Thus, gp*N* promotes lysogeny because *cII* and *cIII* (repressor genes) are immediately downstream of *tL1* and *tR1*. The rightward transcript includes genes *O* and *P*, the replication origin within *O* and the late gene activator *Q* (Figure 1.5). If the phage enters the lytic phase, the early protein product made from the *pR* transcript (gp*Q*), stimulates transcription of genes for lysis and virion formation as accumulation of this protein slows early transcription [Goliger, 1989 #478]. The productive cycle genes are not expressed in the lysogenic phase because the repressor binds to two operators, *oL* and *oR*. Rightward transcription is prevented by the repressor binding to *oR2* and *oR3*. The first gene product on the right, *cro*, binds the same sites as the repressor, specifically to *oR3*, to prevent *cI* transcription [Ptashne, 1986 #479]. *cro* accumulation also reduces CII production by repressing *pR*. The *cII* gene product stimulates leftward transcription of *cI* from promoter *pE* within *cII* [Bushman, 1993 #480][Casjens, 1992 #265]. Cells that have a high CII concentration express the repressor, shutting off their *pR* promoter and enter the lysogenic phase.

Replication of λ phage requires the *pR* transcript gene products, gpO and gpP. Replication occurs bidirectionally from an origin within gene *O*. As replication proceeds, virion components are synthesized by the late genes and assembled into proheads and tails [Campbell, 1994 #164]. Packaging starts at the *cos* site where DNA is cut by the terminase protein complex (gpA and Nu1). DNA to the right is packaged into the prohead until the next *cos* site is reached and cleavage occurs. Once the prohead shell

(composed of gpE) is filled, it expands to assume an icosahedral shape before a second major protein, gpD is added [Hendrix, 1992 #482]. Tails are then attached to the heads. The λ tail is made up of a long hollow tube with bent tail fibers at the sides (*stf* and *tfa* genes), terminating in a single fiber encoded by gene *J* [Haggard-ljungquist, 1992 #483]. All the head genes lie at the left end of the map followed by the tail genes. Three adjacent genes of the late operon, *S, R,* and *Rz* encode proteins for lysis [Taylor, 1971 #484]. The S protein is a holin that forms holes in the cytoplasmic membrane disrupting the membrane potential and allowing the endolysin (R protein) to reach and cleave between the N-acetylglucosamines of the rigid murein layer.

1.8.2. *Serotype-Converting S. flexneri Bacteriophage and O-antigen Modification*

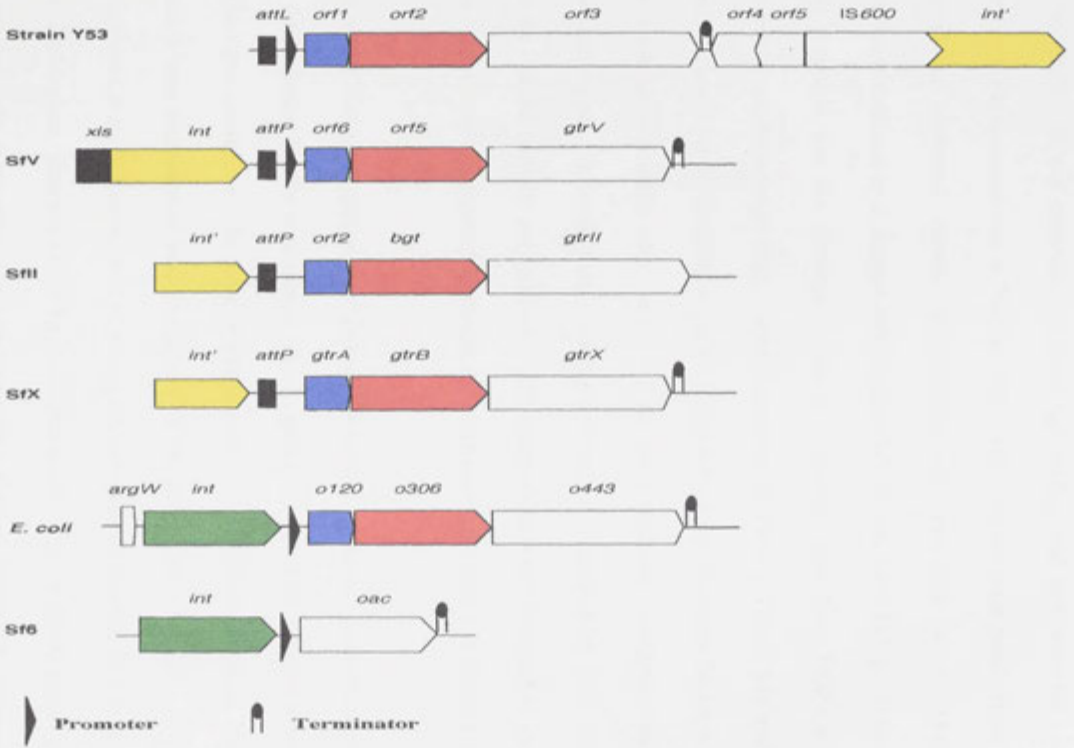
Several *Shigella flexneri* bacteriophages have been known to cause modification of the O-antigenic polysaccharide chain upon lysogeny [Allison, 2000 #199]. Integration of the bacteriophage genome occurs at a specific site identified as the T locus within the *pro* (6 min) and *lac* (8 min) region of the *Shigella* chromosome [Petrovskaya, 1982 #48][Simmons, 1987 #2]. The repeating sugar units of the O-antigen backbone is modified by the addition of an acetyl or glucosyl group. This occurs via the action of phage-encoded acetyltransferase or glucosyltransferase genes [Lindberg, 1977 #488] [Makela, 1984 #121]. The process leads to serotype conversion from the basic O-antigen serotype Y to serotypes 1a, 1b, 2a, 2b, 3a, 3b, 4a, 4b, 5a, 5b, and X (Figure 1.3)[Simmons, 1987 #2] and results in the production of various bacterial serotype

specificities [Gemski Jr., 1975 #113][Lindberg, 1972 #445][Sasaki, 1974 #489][Wright, 1971 #490]. Genetic and phage sensitivity studies showed that *S. flexneri* O-polysaccharide chain with group 3,4 antigen (serotype Y) functions as the receptor for phage adsorption and attachment, thus, O-antigen modification prevents superinfection by other homologous bacteriophages [Lindberg, 1978 #120]. Also, this alteration of surface antigenic epitopes enhances bacterial virulence and survival by evading the host's anamnestic immune response. Due to the occurrence of serotype conversion, therefore, an effective vaccine will need to contain immunogenic epitopes for the different *Shigella* serotypes.

The glucosyl transferase genes of host-induced bacteriophages SfII, SfV, and SfX have been characterised and occur in a conserved three gene cluster. This is located downstream of the phage attachment site (*attP*), preceeded by the integrase (*int*) and excisionase (*xis*) genes (Figure 1.6)[Verma, 1993 #116][Mavris, 1997 #16][Huan, 1997B #117][Huan, 1997A #118][Guan, 1999 #151]. The first two genes in the cluster, designated as $gtrA_{(type)}$ and $gtrB_{(type)}$, have high sequence homology and have been observed to function interchangeably among serotypes. The specific glucosyltransferase genes found in the third position of the cluster, *gtrII*, *gtrV*, and *gtrX*, were less conserved [Allison, 2000 #199].

There are two steps involved in the O-antigen glucosylation mediated by *S. flexneri* bacteriophage SfII [Mavris, 1997 #16]. Initially, the *bgt* product catalyses the transfer

Figure 1. 6 . Comparison of the O-antigen modification gene organization. Represented are the serotype conversion and adjacent factors in the genome of phage SfV (Huan, P.T. *et al.*, 1997A and Huan, P.T. *et al.*, 1997B), SfX (Guan, S. *et al.*, 1998; Guan, S. *et al.*, 1999; and Verma, N.K. *et al.*, 1993), SfII (Mavris, M. *et al.*, 1997), and Sf6 (Clark, C.A. *et al.*, 1991; Verma, N.K. *et al.*, 1991), and in the bacterial chromosome of serotype 1a strain Y53 (Adhikari, P. *et al.*, 1999) and *Escherichia coli* (Blattner, F.R. *et al.*, 1997). Genes encoding proteins showing >85% identity are the same colour. *orf*1, *orf*6, *orf*2 (SfII) and *o120* are *gtrA* homologs; *orf2* (strain Y53), *orf5*, *bgt* and *o306* are *gtrB* homologs; *int* and *int'* are complete or partial integrase genes, respecvtively; xis, excisionase gene; *oac*, O-acetyltransferase gene; *argW*, arginine tRNA gene; *attP*, *attL* and *attR* are phage attachment sites. (Reproduced from Allison and Verma, 2000)


Strain Y53
attL
orf1
orf2
orf3
orf4
orf5
IS600
int'
SfV
xis
int
attP
orf6
orf5
gtrV
SfII
int'
attP
orf2
bgt
gtrII
SfX
int'
attP
gtrA
gtrB
gtrX
E. coli
argW
int
o120
o306
o443
Sf6
int
oac
Promoter
Terminator

of the glucose residue from UDP-glucose onto a bactoprenol lipid carrier molecule, this is followed by a location- and linkage-specific transfer of glucose to the O-antigen sugar units by GtrII. It was observed that genes *bgt* [$gtrB_{(II)}$] and *gtrII* were sufficient to mediate full O-antigen conversion [Mavris, 1997 #16]. On the other hand, three steps involving the sequential transfer of molecules were proposed in the O-antigen glucosylation mediated by *S. flexneri* bacteriophage SfX [Guan, 1999 #151]. Guan *et al.* (1999) proposed that the glucosyl residue is transferred first from UDP-Glc to a membrane bound bactoprenol lipid, a process mediated by GtrB. This is followed by a second step involving the lipid-linked glucose being flipped out (by either the GtrA alone or by GtrA in association with GtrX), across the cytoplasmic membrane onto the periplasmic region. In the third step, the glucosyl residue is attached by GtrX onto the rhamnose I of the partially polymerised O side chain which is still bounded to its lipid carrier. This process is depicted schematically as a hypothetical model in Figure 1.7.

To determine the role of each gene in the serotype conversion casette, Guan et al. (1999), have demonstrated that the three genes, $gtrA_{(X)}$, $gtrB_{(X)}$, and *gtrX* are required to mediate full O-antigen conversion. In order to evaluate the role of individual genes, separate mutations were introduced in each component of the glucosylation cassette. Mutants with deletions in the 5' terminal half of $gtrA_{(X)}$ can only partially convert to a serotype X strain, furthermore, insertion of 470 bp of exogenous DNA into the *Eco*RI site of $gtrB_{(X)}$ also resulted into incomplete serotype conversion since a frameshift mutation was created. This established that *gtrA* and *gtrB* are required for full serotype conversion.

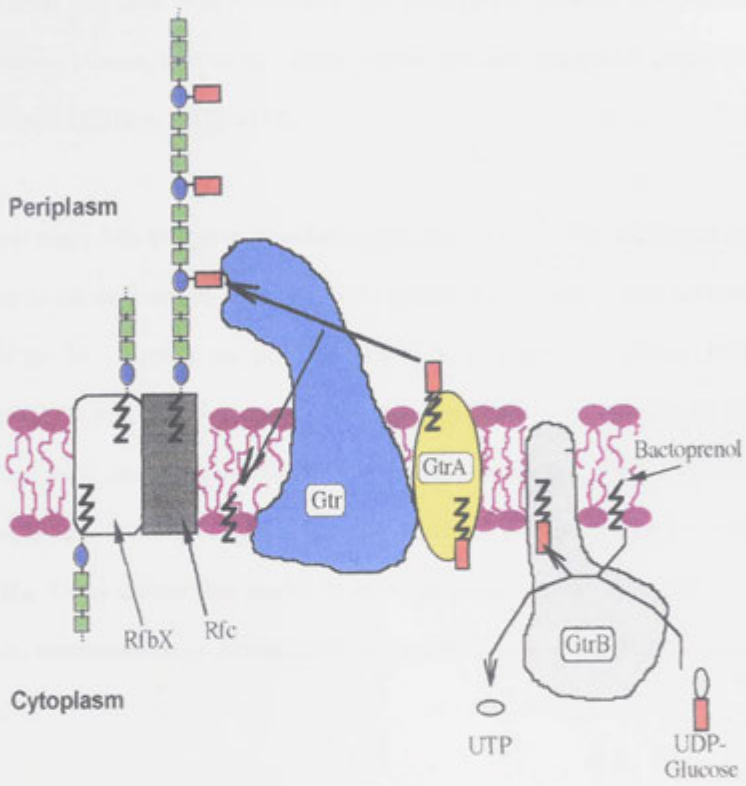


Figure 1.7. Hypothetical model for O antigen glucosylation. The O-antigen repeat unit (Y-serotype antigen) is synthesized in the cytoplasm on the lipid, ACL, "flipped" into the periplasm by the protein, RfbX and polymerised by Rfc. The glucosyl residue from UDP-glucose is transferred to a lipid carrier bactoprenol by GtrB. This lipid-linked glucose is also "flipped into the periplasm by GtrA alone, or in conjunction with the specific glucosyltransferase, Gtr. The group or type specific Gtr protein then transfers the glucose to a particular site on the O antigen chain as it is polymerised by Rfc, and returns the bactoprenol molecule to the cytoplasmic face of the membrane (Reproduced from Guan *et al.*, 1999).

Deletion mutants of *gtrX* did not show any conversion activity, implying this gene is essential for serotype conversion. Thus, the three genes are necessary for full conversion while *gtrX* alone or in combination with either *gtrA* or *gtrB* can only mediate partial conversion [Verma, 1993 #116]. Similar results have been observed in phage SfV [Huan, 1997A #118][Huan, 1997B #117].

Conversely in phage Sf6, an O-acetyl transferase gene, *oac*, catalyses the addition of O-acetyl group to the rhamnose III sugar unit of the polysaccharide chain. This converts serotype Y to 3b, carrying the group antigen 6, 3, 4 (Figure 1.3) [Clark, 1991 #119][Verma, 1991 #76]. The acetyltransferase has been localised downstream of the integrase (*int*) in a similar organisation to the glucosyl transferase gene cluster in bacteriophages SfII, SfV and SfX [Mavris, 1997 #16][Huan, 1997A #118][Guan, 1999 #151]. The 37185 daltons Oac protein is made up of 333 amino acids and is a hydrophobic, membrane-bound enzyme based on hydrophilicity plot prediction [Verma, 1991 #76].

A recent report revealed the presence of a putative three-gene operon from a cryptic prophage which mediates serotype 1a O-antigen modification [Adhikari, 1999 #112]. The predicted proteins ORF1 [($gtrA_{(1)}$] and ORF2 [$gtrB_{(1)}$] show remarkable homology (88-99% identity) to proteins encoded by the serotype conversion loci of *S. flexneri* bacteriophages SfII, SfV, and SfX. No significant sequence similarity was noted for ORF3 (*gtrI*). This gene organisation having two conserved genes followed by a unique

glucosyltransferase gene, is reflective of that observed in bacteriophage SfII, SfV, and SfX (Figure 1.6).

In the O-glucosylation of *Salmonella* species, the addition of glucosyl groups occurs before the lipid-linked O-antigen repeating sugar units have been transferred to the core polysaccharide [Takeshita, 1971 #491], and it involves two major chemical reactions characterised by the step-wise transfer of a glucosyl residue from undecaprenol phosphate (UDP) to a lipid intermediate followed by another transfer step onto the O-antigen subunits [Wright, 1971 #490][Makela, 1973 #492][Nikaido, 1971 #493]. A parallel glucosylation process was observed in *Salmonella* bacteriophage epsilon 34 which can convert *Salmonella* serotype 0,15 to 0,34 by the addition of a glucosyl residue to the galactose sugar in the O-antigen repeat unit [Robbins, 1962 #494][Robbins, 1971 #461]. Conversely in *Salmonella typhimurium*, the chromosome-encoded *oafA* near the *rfa* region produces the factor that mediates O-antigen acetylation [Makela, 1966 #495]. The trans-acylase enzyme product has provided evidence which suggests that, unlike periplasmic glucosylation, the acetylation of the O-antigen occurs in the cytoplasm [Slaugh, 1996 #496]. Yao and Valvano (1994), proposed that acetylation in *E. coli* K-12 takes place at a point prior to the polymerisation of the repeat units, since both the smooth polysaccharide and repeat units are acetyl-modified in the process. This supports the idea of cytoplasmic acetylation as the assembly of repeat units takes place in the cytoplasm before its periplasmic transfer.

1.9. Bacteriophage and their Role in Virulence

Upon lysogeny, temperate bacteriophages can impart properties encoded in its genes that changes the host bacterium, a process termed as lysogenic conversion [Waldor, 1998 #537]. Examples include for exotoxin-encoding genes that are located in the genomes of bacteriophages such as diphtheria toxin, botulinum toxin, Streptococcal erythrogenic toxin, Staphylococcal enterotoxin A, Shiga toxins Stx1 and Stx2, and cholera toxin [Bishai, 1988 #352]. Reports on the filamentous phage phi-CTX revealed that the phage did not only serve as means for the horizontal transfer of virulence genes but its life cycle also played a role in the regulation of the synthesis and secretion of these toxins [Lazar, 1998 #539]. Other bacteriophage-encoded virulence factors include extracellular enzymes such as Streptococcal hyaluronidase, the Lom and Bor outer membrane proteins of phage Lambda, and enzymes that changes the antigenic properties of LPS in *Salmonella* and *Shigella* [Hynes, 1995 #540][Barondess, 1990 #542][Whitfield, 1995 #3]. The spread and persistence of these genes in the bacteriophage genome suggests that these genes provide an evolutionary advantage by enhancing the survival and replication of its host bacteria.

Several studies have assessed the role of LPS in *S. flexneri* invasion and virulence. Rough mutants of *S. flexneri* which lacked O-antigen were shown to be incapable of intercellular spreading despite normal invasion capability [Okamura, 1983 #447]. It was proposed that the smooth surface of invading shigellae is required for spread to adjacent cells and

for protection of intracellular bacteria from the host defense mechanism [Okada, 1991A #22]. In another experiment, the plaque forming ability of *S. flexneri* was not observed in strains that had *galU* or *rfc* mutations [Sandlin, 1995 #543][Sandlin, 1996 #544]. Sandlin *et al.*, (1996) observed that the polar localisation of IcsA on the bacterial surface was altered when *galU*, *rfe*, *rfaL*, and *rfc* mutations were created. Actin polymerisation was affected which caused reduced bacterial intracellular and intercellular movement. Studies of Hong *et al.* [Hong, 1997 #545], showed that both length and distribution of LPS are important for invasion and virulence. The experiments were performed by producing mutations in the chromosomal LPS synthesis genes *rfa*, *rfb*, and *rol*, and in a plasmid-encoded O-antigen chain length regulator pHS-2. Any change in LPS type may have varied effects on the surface characteristics of the mutant that may in turn, show a corresponding effect on the normal organisation of polymerised actin and impair bacterial movement as a consequence [Parker, 1992 #546].

1.10. Objectives

The general objective of this project is to proceed with the sequencing and characterisation of bacteriophage SfV genome in order to discover other gene-encoded properties of phage SfV. The specific objectives are:

- To sequence and analyse phage SfV *Bam*HI fragment A portion of the genome.
- To characterise some fragment A putative genes as to their functionality
- To evaluate the role and effect of phage SfV genome on host virulence.

II MATERIALS AND METHODS

CHAPTER 2

Materials and Methods

2.1. Culture Conditions and Growth Media

Bacterial cultures were generally grown at 37°C overnight in Luria-Bertani (LB) broth and grown on LB agar supplemented with ampicillin, kanamycin or a combination of both for strain selection. LB Congo red agar was used to screen *Shigella flexneri* strains for invasive phenotype, the plates were incubated overnight at 30°C. NZCYM was the liquid medium used for the propagation of phage SfV in its host SFL124. Broth cultures were incubated in a shaking incubator at about 200 rpm. Plate cultures were stored at 4°C while heavy inoculum of strains intended for long-term preservation were resuspended in LB glycerol (1:1) solution before storage in a -80°C freezer.

2.2. Bacterial Strains, Plasmids and Vectors

2.2.1. *Bacterial Strains*

The bacterial strains used and transformed in this study are listed in Table 2.1 and 2.2 with their distinctive characteristics. The host for most ligation products was JM109 derived from *E. coli* K-12 [Yanisch-Perron, 1985 #411]. *Shigella flexneri* strain 124, a live-attenuated vaccine strain in which the *aro*D gene has been deleted from a virulent

Table 2.1. *Escherichia coli* strains used and prepared in this study

STRAIN	DESCRIPTION	SOURCE/REFERENCE
JM109	*rec* A1, *end* A1, *gyr* A96, thi-1, *hsd* R17 (rk-mk+), *sup* E44, *rel* A1	Yanisch-Perron *et al* ., 1985
SY327 lambda *pir*	F' *aro* D (*lac pro*), lambda *pir* (*pir* gene encodes pie protein necessary for the function of R6K origin	Miller and Mekalanos, 1988
B866	P4189- a lysogen with T7 RNA polymerase gene under inducible *lac* UV5 promoter control	Tabor *et al* ., 1985
B869	P4404 containing pT7-5	Tabor *et al* ., 1985
B150	SY327 lambda pir containing pNV731	P. T. Huan
B367	JM109 containing pNV314	P. T. Huan
B377	JM109 containing pNV324	P. T. Huan
B815	JM109 containing pNV724	G. Allison
B823	JM109 containing pNV728	G. Allison
B834	P2780 his dam 3 dam' cloning strain	P. Reeves
B856	JM109 containing pNV749	This study
B857	JM109 containing pNV750	This study
B858	JM109 containing pNV751	This study
B859	JM109 containing pNV752	This study
B860	JM109 containing pNV753	This study
B861	JM109 containing pNV754	This study
B862	JM109 containing pNV755	This study
B863	JM109 containing pNV756	This study
B885	pTrcc 99A containing SfV Xis gene	This study
B876	B866 containing pNV769	This study
B877	B866 containing pNV770	This study
B878	B866 containing pNV770	This study
B879	B866 containing pT7-5	This study
B885	JM109 containing pNV775	This study
B1039	JM109 containing pNV909	This study
B1040	B834 containing pNV909	This study
B1076	B834 containing pNV934	This study

Table 2.2. *Shigella flexneri* strains used and prepared in this study

STRAIN	DESCRIPTION	SOURCE/REFERENCE
SFL1	Wild-type serotype Y	SIIDC*
SFL124	Serotype Y *aro* D	Karnell *et al.* , 1992A
EW595/52	Wild-type serotype 5a	SIIDC*
SFL1339	SFL1 Congo red positive	G. Allison
SFL1342	SFL1339 containing pUC18	G. Allison
SFL1346	SFL1339 containing pNV731	G. Allison
SFL1266	SFL124 containing pUC18	This study
SFL1333	SFL124 lysogen with phage SfV	This study
SFL1334	SFL1012 containing pTrcc 99A	This study
SFL1336	SFL1012 containing pNV775	This study
SFL1337	Cured SFL1012 containing pNV775	This study
SFL1338	SFL1 lysogen with phage SfV	This study
SFL1347	SFL1339 containing pNV724	This study
SFL1348	SFL1339 containing pNV314	This study
SFL1349	SFL1339 containing pNV324	This study
SFL1367	SFL1339 containing pNV909	This study
SFL1394	SFL1339 containing pNV934	This study

*SIIDC, Swedish Institute for Infectious Disease Control

serotype Y strain, was used as a host for phage propagation, plaque assay and as a serotype Y (group antigen 3,4) positive control strain [Karnell, 1992 #364][Karnell, 1993 #363]. Capsid and protease protein were over-expressed in *E. coli* B866, formerly P4189, which is a lysogen with a single copy of the T7 RNA polymerase gene under the control of the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible lacUV5 promoter. T7 RNA polymerase is needed to induce promoter ϕ10 in pT7-5 for the consequent transcription of the genes (see Plasmids and Vectors). Wild-type *S. flexneri* strains EW595/52 and Congo red positive SFL1339 (SFL1, Congo red binding, Pcr^+) were used as hosts for the expression of *xis* gene and invasion assay experiments respectively.

2.2.2. *Plasmids and Vectors*

Table 2.3 lists the plasmid vectors and constructs created in this study. Most works in recombinant DNA preparation involved the plasmid vector pUC19 (Figure 2.1). pNV728 were ligated into this plasmid in order to prepare template DNA for sequencing portions. pUC19 is a small high copy number plasmid (due to the absence of the copy number control gene *rop)* with a pBR322-derived ampicillin resistance gene and origin of replication. It also contains the 146 amino acid terminal portion of the *E.coli lacZ* gene flanking its 54 base pair polycloning site. The former encodes for the amino-terminal fragment of β-galactosidase and this can associate with the host-encoded carboxy-terminal portion of β-galactosidase to form an enzymatically active protein, a process called α-

Table 2.3. List of Plasmids used and constructed in this study

PLASMIDS	DESCRIPTION	SOURCE/REFERENCE
pUC18	pBR322 and M13mp19 derivative; *ori* pBR322, AmpR, *lac* Z+	Yanisch-Perron *et al* ., 1985
pUC19	pBR322 and M13mp19 derivative; *ori* pBR322, AmpR, *lac* Z+ ; polycloning site in opposite orientation to pUC19	Yanisch-Perron *et al* ., 1985
pT7-5	T7 polymerase with *Taq* I/*Xba* I frag. phi-10 promoter plasmid; polylinker region of pUC12; Beta lactamase and ColE1 ori from pBR322	Tabor *et al* ., 1985
pTrcc 99A	pKK233-2 derivative with *trc* promoter upstream of MCS, *rrn* B transcription termination signal, $lacI^q$ and beta lactamase gene	Amann *et al* ., 1988
pNV731	pUC18 with *Xba* I/*Sac* I frag. containing phage SfV 2.8 kb 3-gene casette: gtrA, B, V (*Bam* HI-C)	P. Khinda
pNV314	pUC19 with 3.4 kb *Eco* RI fragment D of phage SfV	P. Huan
pNV324	pUC19 with 7.3 kb *Bam* HI fragment B of phage SfV	P. Huan
pNV724	SfV	G. Allison
pNV728	pUC18 with 5.5 kb *Bam* HI/*Sac* I fragment of phage SfV BamHI frag. A	G. Allison
pNV775	pTrcc 99A with 1 kb Xis gene from pNV324	This study
pNV749	pUC19 with 1.4 kb *Hin* dIII frag. of pNV728	This study
pNV750	pUC19 with 1.2 kb *Hin* dIII frag. of pNV728	This study
pNV751	pUC19 with 2.0 kb *Hin* dIII frag. of pNV728	This study
pNV752	pUC19 with 0.6 kb *Hin* dIII frag. of pNV728	This study
pNV753	pUC19 with 0.3 kb *Hin* dIII/*Sac* I frag. of pNV728	This study
pNV754	pUC19 with 4.2 kb *Pvu* II SfV *Bam* HI frag. A	This study
pNV755	pUC19 with 2.7 kb *Pst* I SfV *Bam* HI frag. A	This study
pNV756	pUC19 with 2.0 kb *Hin* dIII SfV *Bam* HI frag. A	This study
pNV769	pT7-5 with PCR amplified phage SfV complete protease and complete capsid gene nt. 6556-5897	This study
pNV770	pT7-5 with PCR amplified phage SfV incomplete protease and complete capsid gene nt. 6267-5897	This study
pNV775	pTrcc 99A with 1 kb HindIII xis gene from pNV324	This study
pNV909	pNV731 cut and filled at the NcoI site of gtrV gene	This study
pNV934	pNV731 cut and filled at the BclI site of gtrV gene	This study

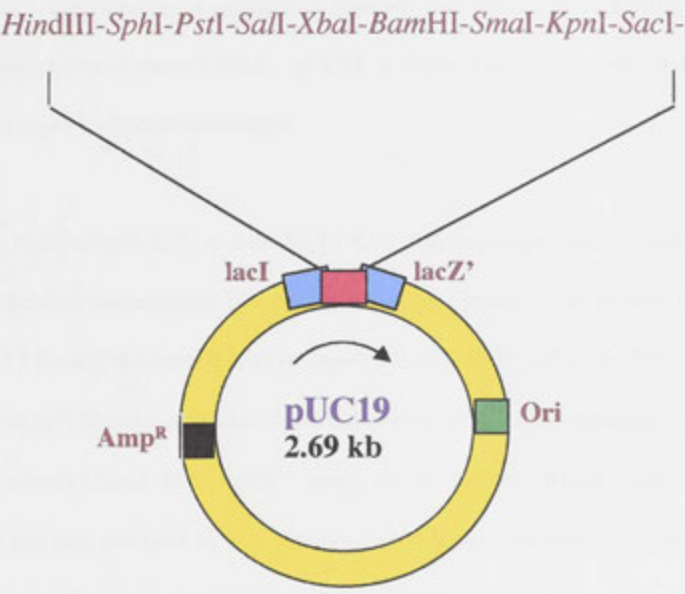


Figure 2.1. Schematic map of pUC19 showing the polycloning site, the *lac* genes, beta-lactamase gene and the origin of replication. Reference: Yanisch-Perron *et al.*, 1985

complementation [Ulmann, 1967 #234]. In the presence of chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, X-Gal [Horwitz, 1964 #235], Lac+ bacteria produced by α-complementation form blue colonies while *lac*- bacteria which carry an insert that disrupts the amino-terminal fragment of β-galactosidase form white colonies. The structure of recombinant plasmids was verified by restriction analysis of minipreparations of plasmid DNA. pUC18 is identical to pUC19 with the polycloning sites arranged in opposite orientation.

Vector pT7-5 (Figure 2.2), a 2404 bp T7 RNA polymerase-promoter system, was used to overproduce bacteriophage SfV capsid and protease protein. The plasmid contains the phage T7 *TaqI*/*Xba*I fragment from nucleotides 22879-22928 (ϕ10), the 70 bp polylinker region of pUC12 and nucleotides 2065-4360 from pBR322 (the β-lactamase gene and the ColE1 origin) [Tabor, 1985 #245]. Strain B869, formerly P4404, hosts the pT7-5 vector and was provided by P.A. Manning of Adelaide University, Australia. Our cloning experiments exploited the specificity of T7 RNA polymerase for its promoters in order to express inserted genes.

Plasmid pTrc 99A (Figure 2.3) expression vector from strain B879 was utilised in the overexpression of the phage SfV *xis* gene. This was the strategy employed in serotype V wild-type EW595/52 to remove the SfV genome which was recombined with its chromosomal DNA [Leffers, 1998 #246]. pTrc 99A is a derivative of pKK233-2

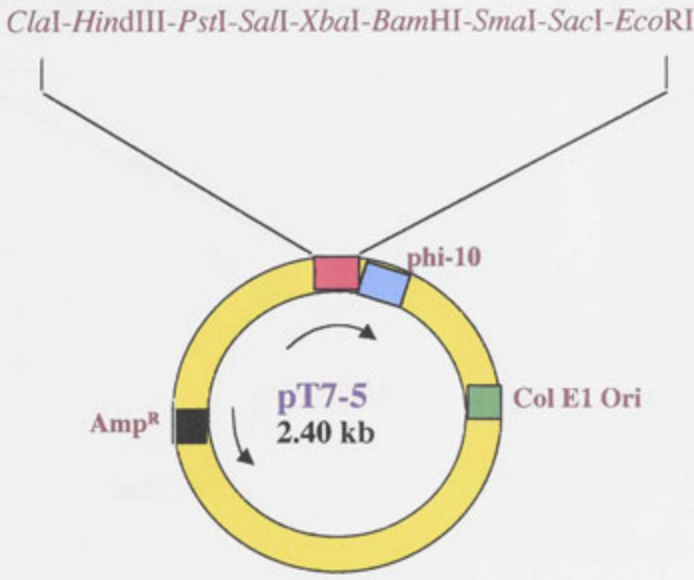


Figure 2.2. Schematic map of pT7-5 showing the polycloning site, the phi-10 promoter gene, the beta-lactamase gene and the ColE1 origin of replication. Reference: Tabor *et al.*, 1985

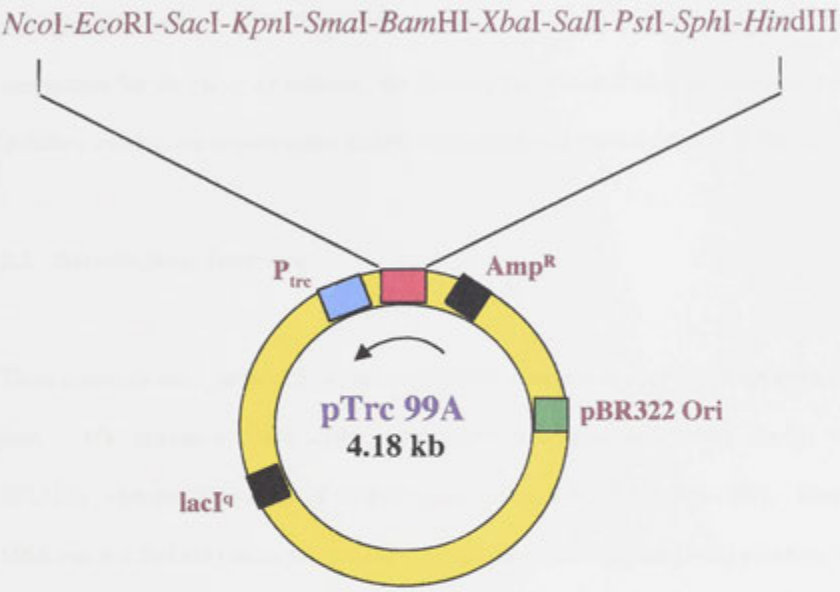


Figure 2.3. Schematic map of pTrc 99A showing the polycloning site, the Ptrc promoter, the *lacI*^{q} repressor, the beta-lactamase gene and the pBR322 origin of replication. Reference: Amann *et al.*, 1988

expression vector with a strong *trc* promoter and a strong *rrnB* transcription termination signal downstream [Amann, 1988 #247]. It contains a multiple cloning site which include a *Nco*I site adjacent to the *trc* promoter which is inducible with 1-5 mM IPTG. pTrc 99A also has a pBR322 origin of replication, *lacI*^{q} repressor region (nt 3055-4135), and a β-lactamase gene (nt 846-1704) conferring resistance to 100 μg/ml ampicillin. In preparation for the curing experiment, the ligated pTrc 99A and *xis* gene construct was initially cloned into non-pathogenic JM109, then subcloned into wild-type EW595/52.

2.3. Bacteriophage Induction

Three protocols were performed throughout the experiments to extract the phage from its host. UV irradiation [Seyedirashti, 1991 #143] was used to confirm identity of SFL1338, a phage SfV lysogen of virulent Congo red positive SFL1 (SFL1339). Phage DNA was purified after induction and digested with enzymes to check for the presence of the typical phage SfV fingerprint (Figure.3.1A and B). While Mitomycin-C treatment [Stanley, 1999 #248][Ehara, 1997 #249] and *xis* gene overproduction [Leffers, 1998 #246] were done to cure EW595/52 by inducing phage SfV out of its host.

2.3.1. *Phage Induction by UV irradiation*

SFL1338 was cultured overnight in LB broth and the next day a 5 ml aliquot of SFL1338 lysogen was pelleted at 4000 rpm in Sorvall RT6000B refrigerated centifuge. The pellet

was resuspended in 2.5 ml of 10mM $MgSO_4$ and the supernatant discarded. The cell suspension was transferred onto a sterile petri dish and exposed to 254 nm UV light source (Phillips UV germicidal lamp) at a distance of 15 cm for 30 seconds. The suspension was then immediately transferred into 15 ml of LB placed in the dark in a shaking incubator for 24-48 hours at 37°C. Culture was examined after 24 hours of incubation to check for occurrence of lysis, extending incubation for a further 24 hours if lysis was not evident. After incubation the cell culture was centrifuged to remove cell debris and the supernatant filtered through a 0.45 µm Millipore filter. The supernatant was then serially diluted for phage titration. Phage identity was confirmed by fingerprinting and colony PCR.

2.3.2. *Phage Induction by Mitomycin-C treatment*

In order to isolate a cured derivative of EW595/52, 10 ml, 100 ml, and 150 ml aliquots of overnight culture were plated onto mitomycin C plates of various concentrations namely 1.0, 0.5, 0.25, 0.125, 0.0625 µg/ml in duplicate. The plates were inoculated using the streak plate technique, and observed for 24-48 hours to check for colonies that survived the toxicity of mitomycin C. The colonies were screened for the absence of the phage through phage sensitivity test (streak and drop method), and slide agglutination test.

2.3.3. *Phage Induction by Xis overexpression*

Recombinant pTrc 99A, *xis* gene plasmid was transformed initially into JM109 host applying the heat-shock method. The plasmid was also electroporated into EW595/52, the wild-type strain intended for phage curing. Transformants were grown at 37°C overnight in 5 ml of 100 µg/ml ampicillin supplemented LB broth. A 50 µl aliquot of the overnight culture was subcultured into fresh 5 ml LB-ampicillin broth (1% subculture) and incubated for a further 2 hours to achieve log phase growth concentration with approximately $1x10^8$ cells per ml. *xis* gene was overexpressed by inducing the vector promoter by the addition of 25 µl of 20 mg/ml IPTG stock (0.4 mM final concentration) to the log phase bacteria culture. At 60 mins and 180 mins post-induction, 10 µl samples were aliquoted and serially diluted ten-fold with 90 µl LB-ampicillin broth. Each dilution from both induction points were plated out on LB-ampicillin plates and incubated overnight at 37°C. Isolates from plates showing 200-500 colonies were picked onto replica plates for colony blotting.

2.4. Preparation of Bacteriophage

2.4.1. *Phage Propagation*

Propagation of bacteriophage SfV was conducted according to the procedures described for the preparation and purification of bacteriophage λ [Sambrook, 1989 #145]. Five ml

of NZCYM media was inoculated with a representative SFL124 colony and shaken overnight at 37°C. A 50 µl aliquot of the overnight culture ($1x10^9$ colony forming unit per milliliter [cfu/ml]) was transferred into 5 ml of fresh NZCYM for a further six hour incubation. The $1x10^8$ cfu/ml culture was then inoculated with a $1x10^7$ plaque forming unit per milliliter (pfu/ml) stock of the phage. The volume of phage used depended on the multiplicity of infection required (MOI, pfu/cfu). The bacteria and phage mixture was incubated at 37°C for 25 minutes and then subcultured into 100 ml of prewarmed NZCYM and incubated in a shaker overnight at 37°C. The next day 2 ml of chloroform was added to the culture in which apparent cell lysis was visible as cell debris settle at the bottom of the flask and incubation continued for another 30 minutes. The mixture was centrifuged twice at 8000g (Sorvall RC plus GSA rotor, 7000rpm), 4 °C for 10 minutes. The supernatant was decanted into a flask, 1 µg/ml of pancreatic DNAse (Boehringer) and RNAse (Boehringer) were added and the solution stirred for 30 minutes at room temperature. NaCl, 5.84g, was dissolved into solution and then the flask was kept on ice for 1 hour. The cold mixture was spun down at 13000g (Sorvall RC plus, GS3 rotor, 9000 rpm), 4 °C for 20 minutes. Ten grams of polyethylene glycol (PEG) 6000 was added to the supernatant, stirred slowly for 15 minutes at room temperature, then incubated on ice for another hour. The mixture was centrifuged at 16,000g (Sorvall RC plus, GS3 rotor, 10,000 rpm) 4°C for 20 minutes and the supernatant discarded. Sedimented bacteriophage particles were suspended in 1 ml SM buffer.

2.4.2. *Phage Purification and DNA Extraction*

The propagated phage in SM buffer was extracted 3 times with equal volume of chloroform to remove PEG and cellular debris. To each 500 µl aliquot of the purified phage preparation, the following were added: 167.5 µl MilliQ water; 37.5 µl 10% SDS (0.5% final concentration), 30 µl 0.5M EDTA (20mM final concentration), 7.5 µl 1 M Tris-HCL pH 8 (10 mM final concentration) and 150 µg Proteinase K from 20 mg/ml stock solution. The mixture was incubated at 55 °C for 30 minutes before extracting twice with phenol and chloroform. Organic phases were saved for chloroform back extraction and mixed with the aqueous phase of the main extract. The aqueous phase was placed in dialysis tubing and incubated overnight at 4 °C against three changes of 1X TE buffer. The purified DNA was collected and stored at 4 °C.

An alternative cesium chloride ethidium bromide gradient centrifugation protocol was used to purify the phage. Using this method, the propagated phage was extracted with an equal volume of chloroform before collecting the aqueous phase to which 0.5 g/ml solid CsCl was added. The mixture was centrifuged at 36,000 rpm for 24 hours in an SW40 rotor Beckman centrifuge. The band containing the phage particles was collected with a needle through the centrifuge tube wall. The CsCl was removed by overnight dialysis against 2L buffer (10mM NaCl, 50 mM Tris-HCl, pH8, and 10 mM $MgCl_2$).

2.4.3. *Plaque Assay for Phage Titration*

A ten-fold serial dilution of the purified phage particle stock in SM buffer was prepared, 10 µl of phage stock and 90 µl of SM buffer per dilution point. To each dilution 100 µl of 3 hour incubated SFL124 log phase growth was added and the mixture incubated at 37 °C for another 20 minutes. The 200 µl content of each tube was pipetted on top of LB agar plate, mixed with 3 ml of 50°C melted 0.5% agar, spread plated then allowed to settle for 5 minutes. The plates were inverted and incubated overnight at 37 °C. The number of plaques were counted and the titer in pfu/ml was derived.

2.5. Bacteriophage Detection and Identification in Lysogen

2.5.1. *Phage Sensitivity Test*

A bacterial clone expressing group antigen acquired through phage directed serotype-conversion can confer immunity to superinfection by the same phage [Clark, 1991 #462]. The phage sensitivity test is based on this premise with experiments involving their sensitivity or resistance to a particular phage species to assess the lytic spectrum of the infecting phage [Huan, 1997B #117]. For our purpose, SFL124 SfV lysogen strains were challenged with phage SfV to confirm successful phage lysogeny, as demonstrated by the resistance to SfV infection. An LB plate was streaked heavily with an overnight culture of *S. flexneri* lysogen test strains and allowed to settle for 10 minutes. Twenty

microliter of purified phage SfV was dropped over the lawn of test bacteria and left to absorb for 20 minutes. Then the plates were incubated at 37°C overnight. The strain was judged sensitive to the phage if a clear zone in which no bacterial growth occurred was seen in the area where the phage had been applied, likewise suggesting unsuccessful phage lysogeny. Alternatively, a streak method was employed by placing a phage drop in a location near the edge of the agar plate. The drop was allowed to flow along the center of the plate by tilting the plate form a diameter line of phage inoculum. The phage suspension was left to dry for 30 minutes. Using a loop, the surface of the test colony was lightly touched and streaked across perpendicular to the phage inoculum line. The plates were inverted and incubated overnight at 37°C. As in the drop method, sensitivity was shown by a clear zone of no growth at the cross point of bacterial streak and phage inoculum. For each test batch, SFL124 was used as the sensitive positive control and the serotype V strain, EW595/52 as the resistant negative control.

2.5.2. *Slide Agglutination Test*

Lysogen and cured strains of *S. flexneri* were screened by the slide agglutination test using sero-type and sero-group specific antisera (Denka Seiken Co., LTD., Japan). This involved a glass slide which was divided into two parts and a 20 µl drop of each polyvalent serum and physiologic saline control (0.9% NaCl) was placed onto each section of the slide. Spontaneous agglutination of antigen was checked by employing physiologic saline, as a negative control. A 20 µl drop of the test bacterium, which had

been densely suspended in physiologic saline, was transferred in the vicinity of both drops of solutions previously placed on the slide. Using a sterile toothpick, the antigen and serum drops and the antigen and physiologic saline were mixed well in their respective sections. The glass slide was tilted back and forth and the mixtures were observed for agglutination. Only agglutinations occurring within one minute were taken as positive and plus signs were used to score the relative strength of clumping. The name of the subgroup of the organism corresponded to the name of the serum which agglutinated the organism.

2.6. DNA Processing and Basic Manipulations

2.6.1. *Preparation of Plasmid DNA*

2.6.1.1. Minipreparations of Plasmid DNA.

Extrachromosomal plasmid DNA was extracted following the alkaline lysis method with slight modification [Sambrook, 1989 #145]. Bacterial culture was grown overnight and then 1.5 ml was pelleted in a microcentrifuge at 13,000 rpm for a minute. The pellet was resuspended in 100 µl of ice-cold solution I (50mM glucose, 25 mM Tris-HCL, pH 8.0, 10 mM EDTA, pH 8.0) by vigorous vortexing. Next, 200 µl of freshly prepared solution II (100 µl 3N NaOH, 150 µl 10% SDS, 1250 µl MilliQ water) was added at room temperature, the solution mixed by inversion and kept on ice for 5 minutes to lyse

the cells. Cold neutralising solution III, 150µl (600 µl 5M KOAc, 115 µl glacial acetic acid, 285 µl MilliQ water) was then added to precipitate cellular debris and chromosomal DNA. The tube was vortexed and placed on ice for 5 minutes and then spun at 13,000 rpm in a microcentrifuge for 6 minutes to obtain the supernatant containing the plasmid DNA. The supernatant was treated with 50 µl of 20 µg/ml RNAse A solution for 20 minutes at 37 °C, then extracted with one volume each of phenol and chloroform followed by a volume of chloroform extraction two times. The plasmid DNA in the aqueous solution was precipitated by adding two volumes of absolute ethanol, incubated at room temperature for 5 minutes, then spun for 8 minutes at 13,000 rpm to collect the DNA. The pellet was washed with 1 ml of ice-cold 70% ethanol. The liquid was drained from pellet before vacuum oven drying at 45 °C for 10 minutes. The DNA pellet was resuspended in 20 µl of MilliQ water and stored at -20 °C freezer.

2.6.1.2. PEG Precipitation of Plasmid DNA as Template for Automated Sequencing.

The volume of purified DNA was increased by dilution with MilliQ water to 50 or 100 µl volume then extracted twice with 1 volume of chloroform. After this, 0.16 volumes of 5M NaCl and 1 total volume of 13% PEG was added to the aqueous phase before it was incubated on ice for 20 minutes. The mixture was microcentrifuged at maximum

speed at 4 °C for 20 minutes. The pellet was rinsed twice with cold 70% ethanol and dried in a vacuum oven for 6-8 minutes.

2.6.2. *Preparation of Chromosomal DNA*

Overnight bacterial culture of 1.8 ml was microcentrifuged at 13,000 rpm for 1 minute. The supernatant was discarded and the cell pellet resuspended in 1 ml of TES buffer (50 mM Tris HCl pH 8.0; 5mM EDTA; 50 mM NaCl) and recentrifuged. The pellet was resuspended in 300 µl of solution containing 25% sucrose and 50 mM Tris HCL, pH 8.0. Next 150 µl of 10 mg/ml lysozyme in 0.25 M EDTA solution was added and the solution placed on ice for 20 minutes. After incubation, 75 µl of 1X TE buffer and 25 µl of lysis solution (10% SDS, 0.25 M EDTA, 1M Tris HCl pH 8.0) was added, followed by addition of 1 mg pronase (Boehringer Mannheim) or proteinase K (Sigma) and incubation at 55 °C for 1 hour. After protein digestion, 600 µl of phenol was added to the clear suspension, which was then gently mixed and centrifuged for 4 minutes. The upper, less viscid, yellowish phenol layer was collected with a tuberculin needle and discarded then another extraction was performed with equal volume of phenol, which was added to the lower, more viscid colorless DNA. The DNA layer was extracted twice with equal volume of chloroform, using a tuberculin needle to collect and discard the bottom, less viscid yellowish chloroform layer. The DNA layer was precipitated by adding 1.5 ml of cold 95% ethanol. Once precipitated, the DNA was spooled out using the hooked tip of

a pasteur pipette and dipped twice into cold 70% ethanol solution and finally transferred into 200 µl of MilliQ water or TE buffer. The DNA was dissolved in a 55 °C water bath for 5 minutes then stored at 4 °C.

2.6.3. *Preparation of DNA for Radiolabelling as Probe*

2.6.3.1. Gel-purified Restriction Fragments The Gigaprime DNA labelling kit (Bresatec, S.A.) employing the oligonucleotide labelling system described by Feinberg and Vogelstein [Feinberg, 1983 #236] was utilised for random primer radio-labelling. For use as a probe, 8 µl of the gel purified 0.6 kb *Eco*RV-*Sac*I and the 0.6 kb *Hin*dIII-*Sac*I, *Bam*HI Fragment A insert was denatured for radiolabelling after quantitation based on its fluorescence intensity in comparison with the SPP1 marker cut with *Eco*RI. The DNA was rendered single-stranded by heating in 100 °C water bath for 5 minutes, snap-chilled on ice for 5 minutes to prevent re-annealing and spun briefly to bring down condensation. To the cold denatured DNA , the following reagents were added: 6 µl of decanucleotide solution; 6 µl of nucleotide-buffer cocktail specific for α-^{32}P-dCTP; 4 µl (50µCi) of α-^{32}P-dCTP (aqueous solution, 3000 Ci/mmol), and 1 µl of Klenow enzyme solution. The tube was kept on ice to discourage premature enzyme activity. The reaction solution was incubated at 37 °C for 15-25 minutes and the reaction stopped by heating at 95 °C water bath for 10 minutes. The labelled probe was added to the hybridisation cylinder with the membrane and prehybridisation solution.

2.6.3.2. PCR Extension Product.

The 1.3 kb serotype conversion gene *gtr*V was amplified by PCR using plasmid pNV323 (contains *gtr*A, *gtr*B, and *gtr*V gene of SfV fragment C) in strain B376 as template. The extension products were purified using the Bresatec DNA purification kit. The concentrated amplified product, 4 µl, was diluted with 4 µl MilliQ water for a total 8µl volume to be denatured for labelling. Similar steps were followed as that in section 2.6.3.1 except that a hybridisation box was used instead of glass cylinder for hybridisation. The probe was used in colony blot hybridisation experiment in chapter 6 to detect successfully cured EW595/52 strain which have lost the gtrV gene it had acquired during phage SfV lysogeny.

2.6.3.3. SfV Genomic DNA.

To check the identity of our lysogen and cured strains prior to the invasion assay experiments, their chromosomal DNA were digested with restriction enzymes *Eco*RV and *Bam*HI, blotted onto Hybond N+ nylon membrane (Amersham Pharmacia Biotech), and probed with genomic SfV DNA to detect which strains contain the SfV genome in its chromosome. For this activity the ~39 kb SfV genomic DNA was degraded to attain 1-2 kb fragment sizes suitable to use as probe. This was achieved by taking 30 µl of dialysed SfV DNA, diluting with 270 µl MilliQ water and sonicating at 8 microns for 10

seconds. The sheared DNA was then purified and resuspended in 15 µl MilliQ water (Bresatec, SA) and visualised on a 0.7% agarose gel to ensure generation of appropriate 1-2 kb fragments for probe preparation. The sonicated 8 µl DNA was denatured and processed as the protocol in section 2.6.3.1.

2.6.4. *Agarose Gel Electrophoresis*

DNA size and concentration were visualised using 0.6-1.0% agarose gel. The agarose powder was dissolved in 0.5X TBE buffer, ethidium bromide at a final 0.5 ug/ml concentration was mixed with the melted agarose. The gel was then poured onto the gel tray and allowed to set usually for an hour. DNA samples were mixed with the loading dye containing 0.25% bromophenol blue (w/v) and 30% glycerol (w/v) and the gel was run at 80 V for 2 hours or before the dye front reach the end of the gel, or 12 V overnight. DNA bands were visualised under UV transillumination and captured by the camera with the Gel-Doc system (Mitsubishi). For each batch, phage SPP1 DNA marker cut with *Eco*RI was used to compare band sizes and fluorescence intensity.

2.6.5. *Quantitation of DNA Concentration*

The amount of DNA in suspension was estimated by comparison of band fluorescence intensity on a agarose gel between a known volume of test DNA and a known volume of phage SPP1 marker cut with *Eco*RI which has predetermined concentration value in

nanogram amount (Bresatec catalogue, 1991). However, a more accurate method was performed by measuring the DNA suspensions absorbance value spectrophotometrically at OD_{260} wavelength (Hitachi U-1100). The suspension was measured in quartz cuvettes and the spectrophotometer was set at zero absorbance using MilliQ water as blank control. The equation applied was:

$C = \phi \times$ dilution factor $\times A$,

where C is the DNA concentration in µg/ml, ϕ is a constant value of 50 for double stranded DNA and 33 for single stranded DNA, and A being the absorbance value measured at OD_{260}.

2.6.6. *DNA Purification by Silica Matrix Binding*

DNA was excised with razor blade and eluted from agarose gel or purified from solution using the Bresatec (Bresatec) or BandPure (Progen) purification kit. The procedure utilises silica matrix to bind DNA, in this technique DNA is bound to glass powder in high salt conditions [Vogelstein, 1979 #232]. The process requires 2.6 volume of the NaI binding buffer to be added to the DNA solution. Next 5 µl of resuspended silica matrix suspension was added and the mixture incubated for 5 minutes at room temperature with regular mixing to enhance DNA binding to the matrix. The silica matrix was then pelleted at 13,000 rpm for 15 seconds and resuspended in ice-cold ethanol wash solution. The

suspension was centrifuged again for 30 seconds and the supernatant discarded. The silica pellet was dried in a 45 °C vacuum oven before resuspending in 15-20 µl of MilliQ water, incubating in a 55 °C water bath for 5 minutes then spinning for 1 minute at 13,000 rpm. The DNA containing supernatant was collected and the pellet discarded.

2.6.7. *Restriction Enzyme Digestion*

DNA was characterised by digestion with restriction endonuclease enzymes used with appropriate buffers and incubation conditions specified by the manufacturer. Master mixes were prepared in large volume digestions comprising of MilliQ water, buffer and the enzyme. Digestions were usually carried out in a 37 °C water bath for 1-2 hour period using 2 Units of enzyme and 0.05-1.0 µg of DNA sample in a 20 µl volume reaction mix. Digestion was arrested by the addition of loading dye, heat inactivation or storage at -20°C. Double digests of DNA samples were carried out using buffer compatibility tables prepared by manufacturers (Amersahm Pharmacia Biotech, Boehringer Mannheim, Bresatec, MBI Fermentas, Promega). The One-Phor-All buffer was also used for this purpose after referring to the chart for buffer compatibility (Amersham Pharmacia Biotech). In some instances, DNA samples were digested first with the enzyme favouring the lower-salt buffer followed by the higher salt buffer and a corresponding increase in the total reaction volume or the DNA was purified after the first digestion using the BandPure or Bresatec purification kit prior to digestion with another enzyme. Digests were usually visualised on varied concentration of agarose gel.

2.6.8. *Dephosphorylation of Linearised Plasmid DNA*

Calf intestinal alkaline phosphatase (CIAP, Boehringer) was used to cleave the 5' phosphate groups from linearised DNA fragments with 5' overhang after a preliminary single digestion to prevent recircularization of the plasmid vector. For a total 50 µl reaction mixture, dephosphorylation was conducted by adding 1-2 units of CIAP, appropriate 1X dephosphorylation buffer and MilliQ water directly to the tube containing the 20 µl restriction enzyme digest, then incubated for 30 minutes at 37°C. The CIAP was inactivated by heating the solution at 75 °C for 10 minutes followed by gel purification of the digested dephosphorylated plasmid vector using the Bresatec purification kit. The process effectively removed traces of CIAP which would have inhibited subsequent ligation reactions.

2.6.9. *Klenow Treatment*

The DNA polymerase I large fragment, Klenow (MBI), was used to fill-in overhanging ends of linear DNA fragments to create blunt ends. Following a one hour digestion with restriction enzymes (*Nco*I or *Bcl*I), one µl of both the Klenow enzyme and dNTP's were added making sure that the restriction enzyme buffer was compatible with the Klenow. The solution was incubated for a further 30 minutes at 37 °C prior to gel purification using the Bresatec purification kit. Alternatively, the DNA was extracted with phenol choloform (1:1) and precipitated with ethanol.

2.6.10. *Ligation*

T4 DNA ligase was used with its corresponding ligase buffer to splice the vector and insert DNA together. The reaction was carried out overnight in a 16°C water bath in a final volume of 10 µl containing 1 µl of 10X ligase buffer and 4 Weiss units of T4 ligase enzyme with a 1:3 vector to insert ratio. Control reactions containing uncut and cut vectors only were included. The ligated recombinant plasmids were transfected into its host or stored at -20°C.

2.7. Competent Cell Transformation

2.7.1. *Preparation of Competent Cells*

2.7.1.1. Electrocompetent Cells.

Electrocompetent strains of B866 (P4189) and SFL1339 were prepared following the method presented by Dower *et al.*, [Dower, 1988 #252]. LB broth, 100 ml was inoculated with 1 ml of overnight bacterial culture (1% inoculum) grown from a representative colony. The culture was grown at 37°C for 3 hours (Log phase) before pelleting the cells by centrifugation at 4°C, 7000 rpm (GSA rotor) for 7 minutes. The pellet was washed twice by resuspending in sterile cold milliQ water, 100 ml the first time and 50 ml the second time, re-pelleting each time at 4°C, 7000 rpm (GSA rotor) for 7 minutes. The pellet was resuspended in 2 ml of sterile cold 10% glycerol solution,

microcentrifuged for 5 minutes at 4°C, 7000 rpm, and resuspended in a final in 400 µl volume of 10% cold glycerol solution. Aliquots of 40 µl were immediately placed at -70°C for preservation and storage.

2.7.1.2. Rubidium Chloride Competent Cells.

The method adapted in preparing *E. coli* JM109 competent cells was that of Rob Hallewell of Chiron Corporation. LB, 100 ml broth was inoculated with 1 ml of overnight bacterial culture (1% inoculum) grown from a representative colony and incubated in a 37°C shaker (250-300 rpm) until OD_{550}= 0.48 (approximately 2.5-3 hours). The cells were chilled on ice and spun for 10 minutes at 4°C, 7000 rpm (GSA rotor). The pellet was resuspended with brief vortexing in 30 ml of ice -cold transformation buffer I (30mM KOAc, 50mM $MnCl_2$, 10mM RbCl, 10mM $CaCl_2$, 15% (w/v) glycerol, pH 5.8) per 100 ml starting culture. The cell suspension was kept on ice for 2 hours, centrifuged for 5 minutes at 4°C, 3000 rpm (GSA rotor), then resuspended as gently as posssible in 4 ml ice-cold transformation buffer II (10mM NaMOPS at pH 7.0 with 1 N NaOH, 75mM $CaCl_2$, 10mM RbCl, 15% glycerol) per 100 ml starting culture. Competent cells were aliquoted in 100 µl quantities into ice-cold Eppendorf tubes and stored at -70°C.

2.7.2. *Transformation*

2.7.2.1. Transformation by Electroporation.

Electroporation is the application of high voltage pulses which permeabilizes the cell envelope and permits efficient uptake of DNA by bacteria [Davis, 1990 #213]. The process involves adding 2 µl of ligation mix to 40 µl of thawed electrocompetent cells. The mixed DNA and cells are then transferred to the pre-chilled cuvette(Bio-Rad). Electroporation was carried out with the Genepulser (Bio-Rad) set at 2.5 kilovolts, 200 ohms and 25 µFD. The cell mixture was immediately added to 1 ml of LB broth and incubated at 37°C for 20-30 minutes to allow gene expression. Afterwhich, 100 µl, 150 µl, and 250 µl aliquot of the cells were spread plated on to plates with appropriate antibiotic supplement for selection, and incubated overnight at 37°C.

2.7.2.2. Transformation by Heat-Shocking.

Aliquots of rubidium chloride competent cells, 100 µl were thawed on ice for 10 to 60 minutes before adding 3 µl of overnight ligation mixture. After 20 minute incubation, heat-shocking followed by incubating the cells for 90 sec. at 42°C or 120 sec at 37 °C. The cells were returned to ice for 2 minutes, then 5 volumes of room temperature LB broth was added before shaking gently at 37 °C for one hour. Aliquots, 50-200 µl were spread on selective nutrient agar plates and these were incubated overnight at 37 °C.

2.8. Polymerase Chain Reaction (PCR)

2.8.1. *Amplification for DNA Sequencing*

2.8.1.1. Checking Primers for Secondary Structures.

The most appropriate primers used were picked by a primer selection program found in the Massachussetts Institute of Technology website (http://www-genome.wi.mit.edu). The computer selection output is based on 200-300 single strand base sequence submitted for analysis. The base sequence was located adjacent to the gene to be amplified ensuring inclusion of RBS and other regulatory genes necessary for expression. Some primers were also manually picked by taking note of reverse complement pairing between the candidate primers 5'-3'sequence and its 3'-5' counterpart. See illustration below:

2.8.1.2. Preparing Sequencing Reactions.

Using 0.5 ml GeneAmp thin-walled PCR tubes, A total 10 µl sequencing reaction mixture was prepared containing 200-500 ng of double stranded DNA template, 2.5 pmol/µl

primer, 4 µl (half-reaction) Big dye terminator cycle ready reaction pre-mix consisting of dye-labelled deoxynucleoside triphosphates, $MgCl_2$ and Amplitaq DNA polymerase FS and MilliQ water made up to 10 µl. The mixture was centrifuged for 15 secs, then 25 cycles of PCR amplification in Perkin Elmer GeneAmp PCR system 2400 was conducted. The primers used in sequencing reactions were typically 18 bp in length containing approximately 50% G+C content and the cycle sequencing parameter setting used were as follows:

- Rapid Thermal Ramp to 96°C
  96°C for 10 sec.

- Rapid Thermal Ramp to 50°C
  50°C for 5 sec.

- Rapid Thermal Ramp to 60°C
  60°C for 4 min.

- Rapid Thermal Ramp to 4°C
  Hold until ready to purify

2.8.1.3. Purification of PCR Extension Products.

After DNA amplification, the PCR tube was centrifuged for 15-20 seconds to bring down condensation. MilliQ water, 10 µl was added to the PCR extension mixture and

the total 20 µl volume added to a tube containing 2 µl 3M NaOAc, pH 5.2 and 50 µl of 95% ethanol and briefly vortexed. The tube was iced for 10 minutes to allow precipitation of DNA , then centrifuged at 4 °C maximum speed for 20-30 minutes in a microcentrifuge. The supernatant was aspirated and discarded before rinsing the pellet with 250 µl of 70% ethanol and vortexing briefly. The tube was re-centrifuged for 10 minutes, supernatant carefully aspirated and discarded then the pellet was dried in a 45 °C vacuum oven for 10 minutes.

2.8.2. *Amplification of Gene using a Plasmid Template.*

pNV323 in strain B376 contains the phage SfV serotype conversion three-gene cassette insert which includes *gtr*A, *gtr*B, and *gtr*V ligated to pUC19 [Huan, 1997B #117]. The 1.3 kb *gtr*V gene unique to the SfV genome was PCR amplified, radiolabelled and used as a probe in colony hybridisation to detect successfully cured strain of wild-type serotype V *S. flexneri* EW595/52. The 20 µl PCR reaction mixture contained 4 µl of 200-500 ng double stranded DNA template (pNV323 was tested in three dilutions: 1 in 50, 1 in 100 and 1 in 200), 10.6 µl of MilliQ water, 1 unit of Taq DNA polymerase and 4.4 µl of the master mix consisting of 2.5 pmol of forward and reverse primers, 2 µl each of 10X Taq polymerase buffer, and 2 mM dNTPs. The 30-mer forward and reverse primers designed by Dr. David Bastin of our laboratory and the cycle parameter setting used to amplify SfV *gtr*V are shown below:

→ **DB2** (Forward) 5'- AGAGAATTCCTACCATTCAACATTAAGGCT-3'

→ **DB3** (Reverse) 5'-AGAGGATCCACATCGCCCAAAATACATCAT-3'

- Initial Denaturation Step

  94°C for 5 min. (1X); Pause here, add enzyme

- Denaturation Step

  94°C for 30 sec. (30X)

- Primer Annealing Step

  50 °C for 30 sec. (30X)

- Extension/Elongation Step

  72 °C for 1 min. (30X)

- Final Extension Step

  72 °C for 7 min. (1X)

The elongation step is estimated to accomplish extension at 1-2 kilobases per minute. Therefore, the one minute elongation time used was sufficient to produce the 1.3 kb SfV *gtr*V coding DNA. The extension products were used as probe for colony hybridisation (see section 2.6.3.2.).

2.8.3. *Amplification of Gene using Whole SfV Genome as Template*

To characterise the functionality of phage SfV protease protein in processing its capsid units, two constructs were developed wherein one contained the complete capsid and protease gene and the other having complete capsid but interrupted gene for the protease. In order to compare expressed resultant protein products, one construct was engineered containing the complete capsid-protease gene inserted into the IPTG inducible expression vector pT7-5 and the other having the complete capsid and only the carboxy-terminal half of the protease gene. The first task was to amplify the inserts for cloning. Primers were designed to contain an *Eco*RI restriction site at the 5' terminus and a *Bam*HI site at the 3' terminus. The cleavage sites corresponded to the appropriate order in the multiple cloning site of pT7-5 that would promote ligation in the correct orientation similar to the promoter's transcriptional direction. The 20 µl PCR reaction mixture contained 2 µl of 200-500 ng double stranded whole SfV genome template, 12.6 µl of MilliQ water, 1 unit of Taq DNA polymerase and 4.4 µl of the master mix consisting of 2.5 pmol of forward and reverse primers, 2 µl each of 10X Taq polymerase buffer, and 2 mM dNTPs. The 26-27 mer forward and reverse primers and the cycle parameter setting used to amplify the 2055 bp (nt 6706 to 4651) SfVcomplete protease-capsid gene and the 1616 bp (nt 6267 to 4651) incomplete protease-capsid gene are shown below:

→ **Proteastart** 5'-AAT **GAATTC** ATCTGACGGGGCTTTTAC-3'

*Eco*RI

→ **Proteasemiddle** 5'-AAT **GAATTC** GCACGCTGAATCTCTCAG-3'

*Eco*RI

→ **Capsidend** 5'-AAT **GGATCC** GACTAATCAACCACCAAC-3'

*Bam*HI

- Initial Denaturation Step

  94°C for 5 min. (1X); Pause here, add enzyme

- Denaturation Step

  94°C for 30 sec. (30X)

- Primer Annealing Step

  49 °C for 30 sec. (30X); Complete Protease-capsid

  48 °C for 30 sec. (30X); Incomplete Protease-capsid

- Extension/Elongation Step

  72 °C for 1 min. and 15 sec. (30X)

- Final Extension Step

  72 °C for 7 min. (1X)

The elongation step is estimated to accomplish extension at 1-2 kilobases per minute. Therefore, the one minute and 15 seconds elongation time used was sufficient to produce the 2 kb complete protease-capsid gene and the 1.6 kb incomplete protease and capsid gene. The extension products were spliced with the overexpression vector pT7-5 and cloned in its host strain B866 (P4189)(see Chapter 5). Another application which used the UV induced and purified phage genomic DNA as template was the amplification of its inherent *gtr*V gene by colony PCR in order to ascertain its identity.

2.8.4. *Colony PCR*

The protocol used was adapted from the method of Schuch and colleagues [Schuch, 1997 #191] with some modifications. The phage SfV serotype conversion gene *gtr*V which is unique to the phage was amplified from a lysogen colony to ensure its acquisition of phage SfV DNA into its chromosome. A single isolated colony of subcultured candidate SfV lysogen was resuspended in 25 µl of 0.5mM NaOH then incubated for 30 minutes at room temperature. After incubation, 25 µl of 1M Tris-HCl, pH 8 was added to the NaOH-treated cells and this was immediately diluted by the addition of 450 µl MilliQ water. A 4 µl aliquot of this DNA preparation was used as template in a PCR reaction mixture identical to section 2.8.2. The 30-mer forward and reverse primers designed by Dr. David Bastin and the cycle parameter setting used to amplify SfV *gtr*V gene were also the same.

2.9. DNA Sequencing and Computer Analysis

Purified extension products (see Section 2.8.1.2) were sequenced in the ABI Automated DNA sequencer, Model 373A which employs the dideoxy method developed by Sanger *et al.* [Sanger, 1977 #233]. The method takes advantage of DNA polymerase's ability to incorporate analogues of nucleotide bases by using 2',3'-dideoxynucleotides as substrate. When a 3'-dye labelled dideoxynucleotide triphosphate (dye terminator) is incorporated at the 3'-end of the growing chain, elongation is terminated selectively at A,C,G or T because the added triphosphate residue lacks a 3'-hydroxyl group (Figure. 2.4). The PE Applied Biosystem sequencer detects fluorescence from four different dyes that are used to identify the A,C,G and T extension reactions. Each dye emits light at a different wavelength when excited by an argon ion laser. All four colors and therefore all four bases can be detected and distinguished in a single gel lane or capillary injection (Automated DNA Sequencing, Chemistry guide: Applied Biosystems, 1998).

Simple text base sequence and its electropherogram output were compared and edited on submission to Telnet 2.6 using the Australian National Genomic Information Service (ANGIS) sequence analysis software package. Phage SfV *Bam*HI Fragment A contigs were assembled into a single contiguous sequence using the Fragment Assembly menu of the GCG program package (2D ANGIS), and submitted to various WEBANGIS programs like WAG's FastA and FastX for nucleic acid and protein database similarity searches, respectively, Map program to extract the amino acid sequence in six reading

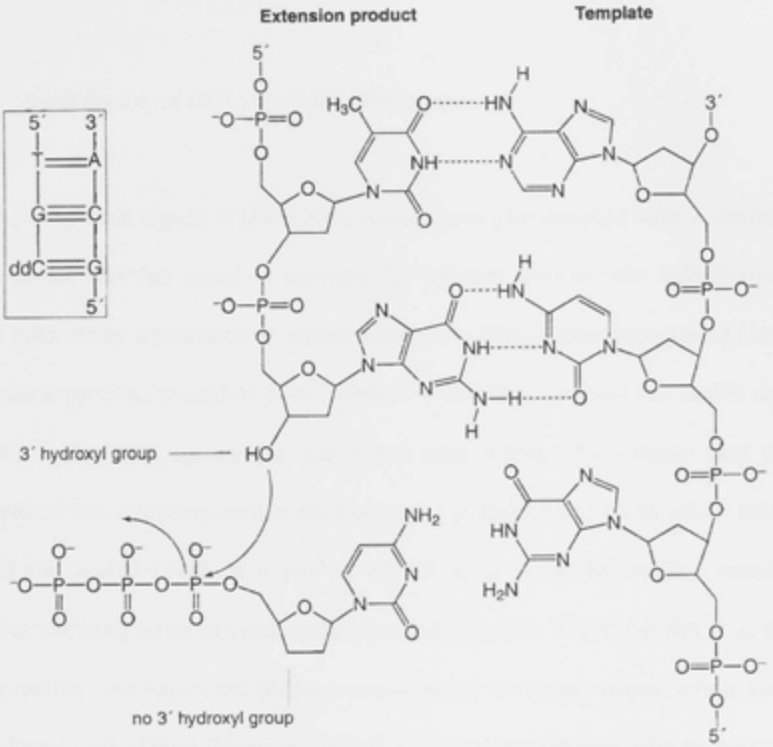


Figure 2.4. Schematic diagram of DNA strand synthesis by formation of phosphodiester bonds. Dideoxycytidine triphosphate (ddC) was used in place of deoxycytidine triphosphate (dCTP) to terminate chain elongation. Inset shows the representation of the process. Source: The Perkin-Elmer Corporation automated sequencing manual, 1998.

frames, Mapsort to locate important restriction sites, Frames to detect likely open reading frames, Eclustalw to perform multiple sequence alignment.

2.10. Southern Hybridisation

2.10.1. *Alkali Blotting of DNA onto Nylon Membrane.*

Initially, restriction digests of DNA were prepared and photographed with a standard marker on the side that served as reference for fragment analysis after hybridisation. This is followed by a procedure of alkaline blotting the DNA digests onto a solid nylon membrane support as prescribed in the Hybond+ (Amersham) protocol for nucleic acid transfers. The 0.6% agarose gel was treated with 0.25M HCl solution until the bromophenol blue dye component of the loading dye changed from blue to yellow color. The gel was soaked briefly in a tray of MilliQ water while the capillary transfer apparatus was being set up on a rectangular glass dish half-filled with 0.4M NaOH as the blotting buffer. An improvised platform was placed in the buffer topped with a glass plate where a wick of three Whatman 3MM filter paper sheets saturated with the blotting buffer were laid across the platform making sure their ends touched the buffer pool. The agarose gel with a small diagonal cut at the upper right-hand corner was placed face down on the 3MM filter paper wick and a similar sized, corner cut and MilliQ water saturated Hybond N+ nylon membrane was layered carefully on top of the gel. The membrane was topped with three sheets of same-sized Whatman 3MM filter saturated with the

0.4M NaOH blotting buffer, and a 5 cm high stack of absorbent paper towels, glass plate, and finally a one kg. weight. A schematic representation of the set-up is presented in Figure. 2.5.

2.10.2. *Membrane Blocking with Non-Homologous DNA*

This pre-hybridisation step involved blocking the DNA blotted membrane with 25 ml of pre-hybridisation solution (Final concentration: 5X SSPE, 5X Denhardt's solution, 0.5% SDS and 100 µg/ml of denatured snap-chilled Herring sperm DNA) for 2 hours in a hybridisation oven (Scientronic HB900) at 65°C. The membrane was placed with blot side facing inside the glass rolling cylinder bottle or facing up when the hybridisation box was used.

2.10.3. *Hybridisation and Autoradiography*

After incubation with the pre-hybridisation solution, the denatured radiolabelled DNA probe (Section 2.6.3) was added into the hybridisation container with the membrane then re-incubated at 65°C in the hybridisation oven for at least 12 hours. Then the membrane was washed twice 10 minutes each with 2X SSPE containing 0.1% SDS at 45 °C, with a final wash with 1X SSPE with 0.1% SDS for 15 minutes. The amount of radioactivity was measured on the surface of the membrane with a Geiger counter to determine if further washes are required. Membranes registering up to 15 counts per second reading

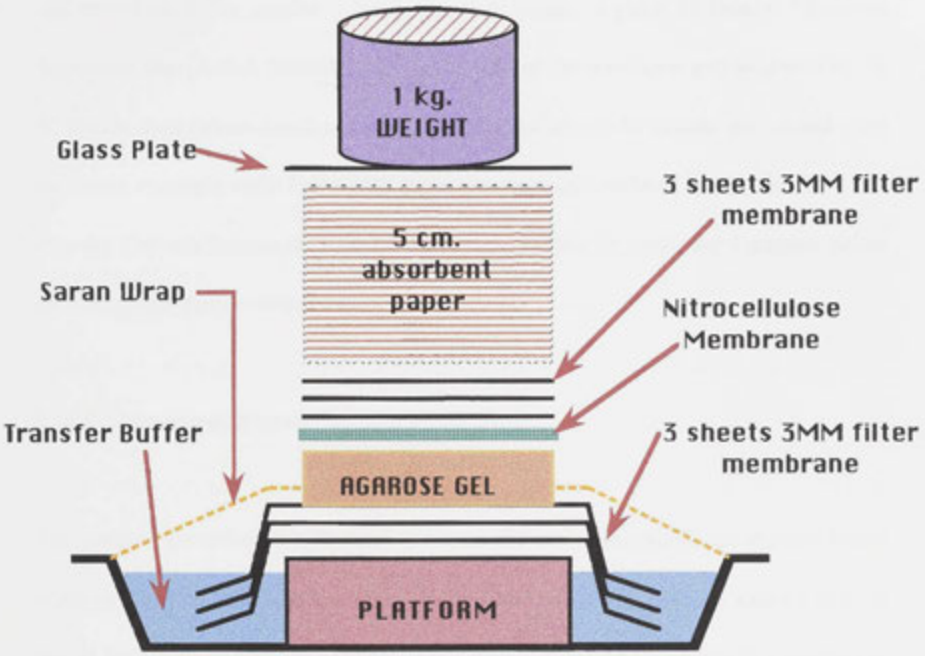


Figure 2.5. Schematic diagram of the capillary transfer set-up for alkaline blotting of DNA onto Hybond N+ nitrocellulose membrane.

were not washed further while membranes emitting 20 counts per second or higher were subjected twice to stringent washes with 0.1X SSPE, and 0.1% SDS at 65°C for 10 minutes each wash. After the final wash, the membrane was covered with Cling wrap, and placed on a film casette. Inside the dark room, a piece of Biomax Maximum Sensitivity film (Kodak Scientific Imaging) on top of the membrane and incubated at -70 °C for the appropriate length of time. Membranes with 5-15 counts per second were incubated overnight while those with higher counts were incubated for few hours. The exposed film was fed into the Kodak X-OMAT 1000 Film Processor for 5 minutes before the radiograph was developed.

2.10.4. *Membrane Stripping*

The method prescribed by Hybond N+ manufacturer (Amersham) to remove bound probe on the membrane was followed. It required rinsing with cupious amount of 0.4N NaOH for 2 hours at 65 °C in a shaking water bath, followed by a rinsing in a solution of 0.1X SSC, 0.1% SDS and 0.2M Tris HCl, pH 7.5 for 1-2 hours in a 45 °C shaking water bath. The desired radioactivity reading was a background reading of <2 counts per second for successful stripping.

2.11. SDS-PAGE and Western Immunoblotting

2.11.1. *Protein Assay- Trichloroacetic Acid (TCA) Precipitation*

Protein precipitation was performed to quantitate the amount of phage particle protein in the sample in order to determine the optimum volume for gel loading. The assay was based on TCA precipitation of protein samples whose spectrophotometric absorbance reading was plotted against bovine serum albumin (BSA) standard curve to identify the concentration of protein [Peterson, #254]. The first step was to prepare the BSA solution series for the standard curve. Amounts of 0 µl, 10 µl, 20 µl, 30 µl , up to 100 µl were aliquoted from a 1 mg/ml BSA stock solution at -20 °C into separate test tubes then filling up each tube content up to 0.5 ml with MilliQ water. The phage stock test samples of 2 µl, 4 µl, 6 µl, 8 µl, and 10 µl amounts were also transferred into different test tubes. To each BSA and test samples, equal volume, 0.5 ml of Reagent A of copper tartrate carbonate [CTC][Sigma], 10% SDS, 0.8N NaOH, and MilliQ water was added, after vortexing the solution was incubated for 10 minutes at room temperature. Next, 0.25 ml each of reagent B (2 ml Folin [Sigma] and 10 ml MilliQ water) was added and after vortexing the solution was incubated for 20 minutes at room temperature. The samples were then placed in cuvettes or microtiter plate wells for spectrophotometric reading of absorbance values at OD_{750}. The concentration of the phage stock test sample was derived from the corresponding BSA standard curve.

2.11.2. *SDS-PAGE*

The phage SfV and protease-capsid experiment bacterial clone samples were usually run on 12% (unless otherwise specified) sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). The samples were first denatured by boiling for 5 minutes

in SDS-PAGE sample loading buffer (2% SDS, 50mM Tris-HCl, pH 6.8, 10% glycerol, 0.1% bromophenol blue)[Laemmli, 1970 #255]. Typically, sample mixtures of up to 30 µl were prepared for each well of the mini-gel set-up (Bio-Rad) and up to 70 µl for the large gel. The low molecular weight (LMW) protein marker sample mixture from the calibration kit (Amersham Pharmacia Biotech) was boiled with 20 µl of the sample loading buffer. Using a pasteur pipette attached to the auto-pipettor, the prepared resolving gel mixture was poured carefully in between two ethanol cleaned glass plate sandwich bounded at its side and bottom edges with spacer combs and sealed with melted 1% agarose gel. The glass chamber was filled up to approximately one cm below the well comb then overlayed with MilliQ water. The gel was allowed to polymerise for 45 minutes to 1 hour before the top water was decanted and the set resolving gel surface blotted dry with a piece of 3MM Whatman filter paper. The stacking gel was poured next on top of the resolving gel , then the comb was immediately inserted through the stacking gel ensuring no bubbles were trapped in the stacking gel and around the teeth of the comb. Once the stacking gel was set after 30 minutes, the comb was removed and the gel assembly was placed in the running tank, attached to the vertical apparatus, and filled to the appropriate level with 1X SDS-PAGE running buffer. Samples were loaded onto each well with a Hamilton syringe and electrophoresed at a constant voltage (180 V) until bromophenol blue reaches the bottom of the gel (approximately 45 minutes for the mini-gel and 4 hours for the large gel). After the run, the gel was either stained or the proteins blotted onto membrane for Western immunoblotting.

2.11.3. *Protein Band Staining*

2.11.3.1. Coommassie Blue Staining.

The stacking gel was cut out after electrophoresis and the resolving gel stained in coommassie blue stain (Sigma)(0.05% coommassie blue R250, 40% Methanol, 10% acetic acid) sufficient to cover the submerged gel. The pan was covered with cling wrap then placed overnight at low setting on a Bio-Line orbital shaker (Edwards Instruments Company, N.S.W.). The coommassie stain in the pan was replaced with distilled water after overnight incubation then destained in the microwave (Panasonic) at high setting three times for 10 minutes. The gels were photographed under light illumination, sealed with water in a heat-sealing bag, dried on 2 pieces of 3MM Whatman paper on a gel drier (Bio-Rad), or placed in a Western blotting transfer apparatus.

2.11.3.2. Silver Staining.

The procedures followed were adapted from the protocols developed by Merril and his colleagues [Merril, 1981 #256][Merril, 1982 #257][Merril, 1984 #258]. After the electrophoretic run the proteins were fixed by soaking the resolving gel in 400 ml of 40% methanol in 10% acetic acid solution on orbital shaker for 30 minutes, followed by soaking in 400 ml of 10% ethanol in 5% acetic acid for 15 minutes. The gel was then transferred into 200 ml of oxidizer (Potassium Dichromate and nitric acid) at room temperature for 5 minutes, and washed twice with distilled water for 5 minutes before

soaking in 200 ml of silver reagent (Silver Nitrate solution) for 20 minutes. The gel was washed in distilled water for 1 minute and then soaked in 200 ml of developer solution (Sodium Carbonate and Paraformaldehyde) for approximately 5 minutes or until the bands reach the desired intensity in relation to the background. The reaction was stopped with a 5 minute soak in 5% acetic acid.

2.11.3.3. Ponceau S Staining.

Some protein gels were visualised for the presence of protein bands after Western blotting and before hybridisation. The membrane was soaked in a working solution of Ponceau S stain,3-hydroxy-4-(-2-sulfo-4-[-4-sulfophenylazo]-phenyl-azo)-2,7naphthalenedisulfonic acid (Sigma) for 10 minutes with gentle agitation or until the protein bands were visible. The membrane filter was washed with several changes of MilliQ water at room temperature, then the lane and protein positions were marked with pencil before proceeding with the Western blot assay. The working Ponceau S stain solution was a 1:10 dilution of the stock solution (2 g Ponceau S, 30 g trichloroacetic acid, 30 g sulfosalicylic acid, made up to 100 ml with MilliQ water) in 9 parts of MilliQ water.

2.11.4. *Western Immunoblotting*

2.11.4.1. Electrophoretic Protein Transfer. There were two electrophoretic transfer system utilised in our experiments. The first one was carried out according to the method of Towbin *et al.* [Towbin, 1979 #259]. Assembly components (the gel,

sandwich plates, four support pads, nitrocellulose membrane cut to the size of the gel, four sheets 3MM Whatman filter paper also cut to the size of the gel) were immersed for 20 minutes in the Western blot transfer buffer (50mM Tris-HCl, 250mM glycine, 0.1% SDS, and 20% methanol). The transfer assembly were arranged in the following order from the bottom: black plate of the sandwich plate, two support pads, two sheets of Whatman paper, the gel, the membrane, two sheets of Whatman paper, two support pads then closed on top with the white plate of the sandwich holder. The sandwich assembly was submerged in the Bio-Rad transblot electrophoretic tank half-filled with the transfer buffer in such an orientation that the black side of the sandwich plate was positioned nearest to the the negatively charged cathode (black) and the white side nearest the positively charged anode (red). For a minigel transblot, the ice casette was placed next to the sandwich assembly and run at room temperature for 1.5 hours at 40 V or 15 V overnight. The large transblot was electrophoresed in the cold room at 15 V overnight.

Another instrument used was the Millipore dry blot apparatus. The gel, ten sheets of 3MM Whatman filter paper and a sheet nitrocellulose membrane (Hybond-C extra for chemiluminescence system) were soaked in the Millipore blot buffer (11.72 g glycine, 28.24 g Tris-HCl, 1.5 g SDS and 800 ml methanol were made up to 4000 ml with MilliQ water). The soaking time was 30 minutes for the minigel and 1 hour for the large gel. From the base plate (anode) the components were stacked as follows: five 3MM Whatman paper, membrane, gel, five 3MM Whatman paper, capped with cathode lid. Electroblotting progressed for 1.5 hours at a constant current of 0.8 mA cm^{-2} gel supplied.

For A Single Minigel: 8.5 x 5.5 cm = 46.7 cm^2

46.7 cm^2 x 0.8 mA per cm^2 = 37 mA constant current supplied

For Two Minigels: 74 mA constant current supplied

For Large Gels: 9.0 x 10 cm = 90 cm^2

90 cm^2 x 0.8 mA per cm^2 = 72 mA constant current supplied

2.11.4.2. Acetone Powder Preparation.

Overnight LB cultures of SFL124, 1.5 ml, were pelleted at 6000 rpm for 5 minutes and resuspended by vortexing for 30 seconds in 500 µl of iced 0.9% NaCl solution then incubated on ice for 5 minutes. One ml of cold (-20°C) acetone was mixed vigorously into each tube of bacterial saline suspension, this was then incubated at 0°C for 30 minutes with occasional vortexing. Next the solution was centrifuged at 10,000 rpm for 10 minutes and the pellet was resuspended in fresh cold (-20°C) acetone, vortexed and incubated at 0°C for 10 minutes. The pellet was sedimented again at 10,000 rpm for 10 minutes the supernatant discarded, then air dried for 1-2 hours.

2.11.4.3. Adsorption to Remove Nonspecific Binding.

Polyclonal antisera usually contain a small proportion of antibodies that will bind either specifically to contaminating antigens on the membrane blot preparation (anti-bacterial

antibodies in the sera) or nonspecifically to charged antigenic residues. To minimise the presence of competitor antibodies, their binding sites need to be blocked by exposing them to saturating amount of competitor protein (bacterial whole cell acetone powder containing disrupted bacterial antigens) that is not the antigen of interest (phage SfV capsid protein) [Harlow, 1988 #260]. To reduce the amount of anti-SFL124 (strain used to propagate phage SfV particles) antibodies in our polyclonal rabbit antiserum raised against whole phage SfV particles, acetone powders were prepared and added into the antiserum to a final concentration of 1%. This serum-acetone powder mixture was incubated for 30 minutes at 4°C then spun for 10 minutes at 10,000 rpm. The supernatant was collected and used as the primary antibody reagent for our immunoassay.

2.11.4.4. Western Immunoassay. The protein blotted nitrocellulose membrane was incubated at room temperature for one hour with the blocking buffer (5% skim milk [Carnation]-5 g skim milk powder in 100 ml TBS-T) on a rotating platform shaker. The membrane was then incubated with the primary antibody, rabbit anti-SfV polyclonal antisera diluted 1:100 (50 µl of adsorbed antisera in 5 ml of TBST-1% skim milk), for 2 hours at room temperature on a rotating platform. The membrane was then washed 3X for 10 minutes each in TBS-T or PBS-T (0.1% Tween 20 in TBS or PBS) then incubated for one hour with the secondary antibody (Sigma Immunochemicals) (Goat anti-rabbit IgG conjugated with horseradish peroxidase) diluted 1:12000. The membrane was again washed 3-4 x 10 minutes in TBS-T or PBS-T. The membrane was developed using the

BM chemiluminescence blotting substrate-POD kit (Boehringer Mannheim) by soaking for 1 minute in the detection solution containing solution A and B at 100:1 ratio. Inside the dark room, a sheet of Biomax X-ray film (Kodak) was placed on top of the membrane and exposed for 1-2 minutes then developed in the Kodak X-OMAT 1000 processor.

2.12. Electron Microscopy

Negative staining was performed on phage SfV particles to visualise at high magnification (20K, 50K, 100K). Purified phage particles, 20 µl, was dropped on the surface of parafilm M (American National Can). A carbon coated copper grid was laid on the surface of the phage solution for 1 minute to allow absorption of phage particles on the 300 mesh copper grid. The grid was drained of excess phage solution and laid on top of a MilliQ water drop for a 5 seconds wash. The grid was drained again then transferred onto the surface of a 2% sodium phototungstate, pH 7.0 to stain for 15-30 seconds then drained again and air-dried for 10 minutes. The phage particles were visualised with a Hitachi JEOL X electron microscope at 80-100 kV.

2.13. HeLa Cell Culture

2.13.1. *Cell Passage and Seeding*

Sterile PBS and Growth media (500 ml RPMI 1640 without glutamine, 50 ml Fetal Calf Serum [10%], 5 ml Glutamine [2mM], and 100 µl Gentamycin [2 µg/ml]) were warmed

in a 37°C water bath. The confluent monolayer was washed and drained twice with 5 ml PBS before 2 ml of Trypsin-EDTA was added to detach cells from the flask wall. Trypsin-EDTA treatment was done in a 37°C 5% CO_2 incubator for 5 minutes. The flask was tapped to ensure cell detachment and 5 ml of growth medium was added to inhibit trypsin. The cells were transferred into a tube and centrifuged at 1000 rpm for 5 minutes (Sorvall). The supernatant was discarded and the cells resuspended in 10 ml of growth medium with 2 µg/ml gentamycin for tissue culture flask seeding, however, 25 ml of growth medium with antibiotic was usually used to resuspend the cells for 6-well plate seeding. The resuspended cells, 20 µl were also mixed with 20 µl tryphan blue for cell counting in haemocytometer (see below). Two milliliters of $1x10^6$ cells per ml was seeded in a new tissue culture flask in which 20 ml of growth medium with 2 µg/ml gentamycin was also added. For the 6 well plate seeding, 2 ml of the 25 ml cell suspension was transferred into each well if invasion assay was to be performed after 24 hours, or 1 ml of cell suspension and 1 ml of growth medium was placed into each well if invasion assay was performed after 48 hours. The plates and flasks were incubated in a 37°C 5% CO_2 incubator and checked everyday for cell growth.

2.13.2. *Cell Freezing and Thawing*

The cells were harvested for cell passage/subculturing and resuspended in 5 ml cold 10% DMSO. The cell suspension was aliquoted at 1 ml amounts per vial then placed on ice. The vials were incubated for 24 hours at -70°C freezer before transferring into liquid

nitrogen. Cells were thawed immediately at 37°C water bath then transferred to 10 ml of warm growth medium, mixed and centrifuged at 1000 rpm for 5 minutes. The pellet was resuspended in 10 ml growth medium and aliquoted accordingly into tissue culture flasks prior to incubation.

2.13.3. *Cell Counting in Haemocytometer*

The coverslip was positioned carefully on top of the Neubauer haemocytometer and the counting chamber was charged with the 1:1 cell-tryphan blue suspension. Live cells in the WBC squares, not absorbing the tryphan blue dye were counted once using the L-rule. 100-300 cells were counted to obtain a statistically significant cell count.

For example,

131 cells counted in 3 large WBC squares or 48 small WBC squares (16x4)

Therefore, 131 cells x 48/16 x 10^4 per ml (one small square= $1/16x10^{-4}$ cm^3) and cm^3 = ml, for aqueous suspension

= 131 cells x 3 x 10^4 per ml

= 3.9 x 10^6 cells per ml x 2 (1:1 dilution with tryphan blue)

= 7.8 x 10^6 cells per ml

2.14. Invasion/ Gentamycin Killing Assay

LB broth inoculated with bacterial test strains were incubated overnight at 30°C. The following day, 0.1 ml was subcultured for 2 hours in 37°C incubator shaker. Meanwhile,

confluent HeLa cells grown in 6-well plates were washed twice with 2 ml PBS prior to inoculation. Three ml of log phase culture was spun at 13,000 rpm for one minute and the cell pellet resuspended in 4.5 ml of RPMI-FCS without antibiotic. Each inoculum preparation represented one test strain and 2 ml of this was added to 2 wells, one to be gentamicin treated and the other an untreated well. The plate was then incubated in the 37°C 5% CO_2 incubator for 1.5 hours to allow adsorption and invasion of *Shigella* strain. Following the incubation, 1 ml of inoculum from each well was discarded and replaced with 1 ml of RPMI-FCS- 500 µg/ml gentamicin for the treated wells and 1 ml of RPMI-FCS without antibiotic for the untreated wells. The plate was then incubated for another 1.5 hours. The cells were then washed twice with 2 ml PBS and then 500 µl of 0.05% Triton-X-100 in PBS was added. The plate was incubated at room temperature for 10 minutes before adding 500 µl of LB broth and resuspending the cells (Figure 2.6). The bacterial suspension from each well were diluted with PBS, spread plated onto LB agar and incubated overnight at 37°C. The percentage of invading bacteria was calculated as follows:

$$\text{Percent Invading Bacteria} = \frac{\text{No. of Bacteria recovered from gentamicin-treated cells}}{\text{No. of Bacteria recovered from untreated cells}} \times 100$$

$$\text{Percent Invading Bacteria} = \frac{\text{Intracellular cells}}{\text{Total cells}} \times 100$$

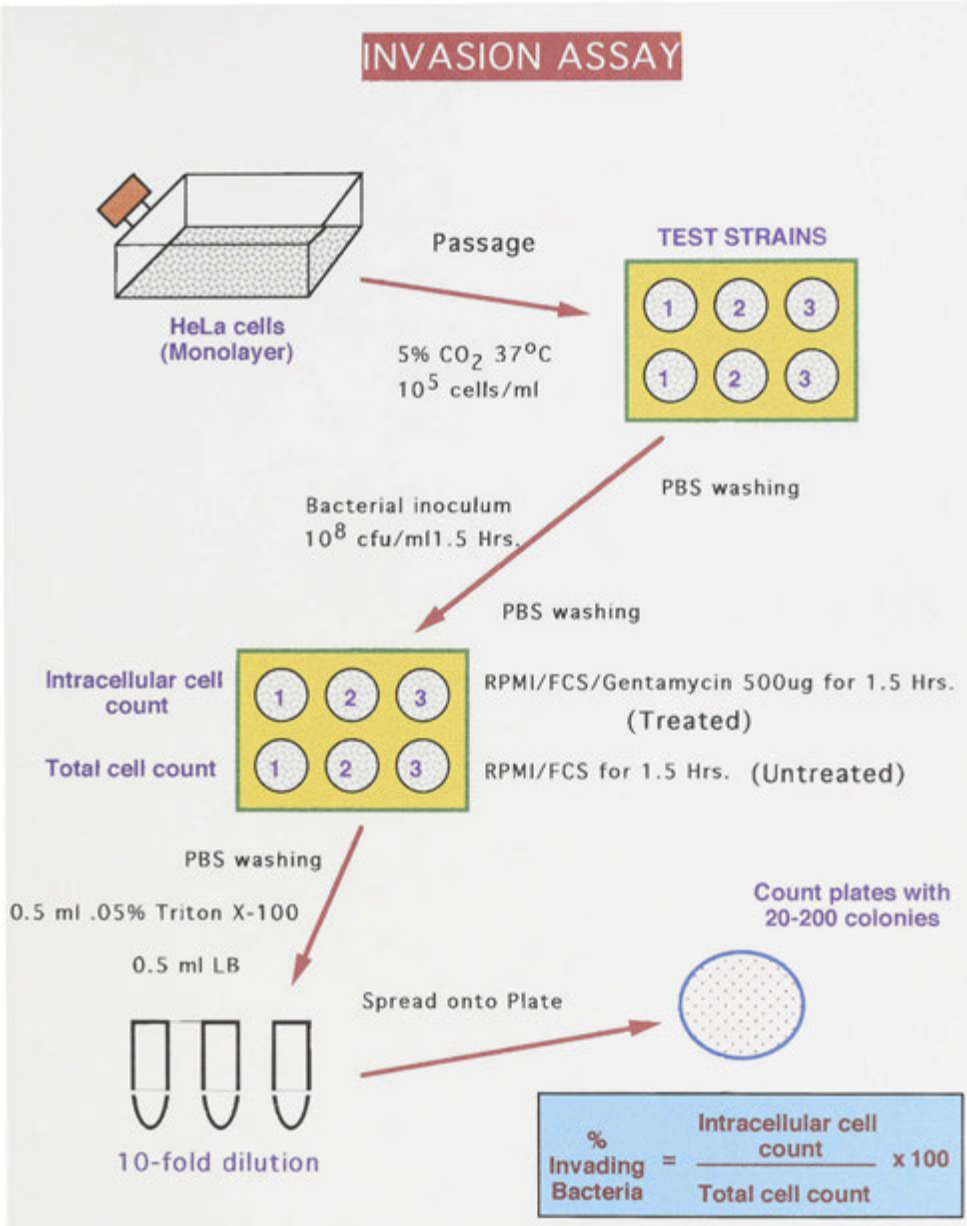


Figure 2.6. Flow diagram showing the invasion assay method

III RESULTS

CHAPTER 3

DNA Sequencing Strategy for the Serotype-Converting *Shigella flexneri* Bacteriophage SfV *Bam*HI Fragment A

3.1. Introduction

SfV is a temperate bacteriophage which carries genes important in mediating serotype-conversion of *S. flexneri* through O-antigenic modification [Huan, 1997A #118]. Phage SfV 0-antigen modification gene cluster, is composed of three genes designated as $gtrA_{(V)}$, $gtrB_{(V)}$, and *gtr(V)*. The function of these genes is to transfer glucosyl residues from the carrier molecule to a membrane-anchored lipid, move the lipid-linked glucose from the cytoplasmic to the periplasmic region and attach the glucosyl residue to the O-antigen sugar units, respectively [Allison, 2000 #199][Guan, 1999 #151]. The addition of the glucosyl group to the rhamnose II of the basic tetrasaccharide repeating units of the O-antigen changes *S. flexneri* serotype Y to serotype 5a [Simmons, 1987 #2][Huan, 1997A #118]. These genes are located immediately downstream of the integration and excision locus composed of the *att*P site preceded by the *xis* and *int* genes, and the glucosyltransferase gene *gtrV* is followed downstream by *orf2* and *orf3* [Huan, 1997A #118][Huan, 1997B #117]. This genetic organisation was also observed in the genome of *Salmonella typhimurium* phage P22 and other *S. flexneri* bacteriophages [Vander Byl, 2000 #351][Allison, 2000 #199].

Knowledge of the entire phage SfV genome is important in order to determine the influence of the phage genes on the traits and characteristics of its *S. flexneri* host. The usage of sequence information were described in other double-stranded bacteriophages whose functional organisation was derived after completion of its genomic sequence [Nakayama, 1999 #203][Kaneko, 1998 #204][Alonso, 1997 #205][Altermann, 1999 #206][Tremblay, 1999 #207]. The recent completion of bacteriophage P22 and prophage VT1-Sakai sequences have been reported [Vander Byl, 2000 #351][Katsushi, 2000 #505]. Incidentally in 1999, there were only about 30 completed phage genome sequences submitted in the databases [Whitman, 1998 #506][Hendrix, 1999 #166]. This represents a minute proportion of the many known bacteriophages, therefore, more completed phage genome should be discovered in order to perform an accurate evolutionary and comparative studies on vital phage gene modules. Sequences of prophages have also been reported during the sequencing of bacterial genomes such as those of *E. coli*, *H. influenzae*, and *Mycobacterium* species [Cole, 1998 #328][Kunst, 1997 #507][Blattner, 1997 #4].

Many strategies have been utilised in various sequencing projects. Most projects involving large-scale DNA sequencing of higher eukaryotic organisms employ the random shotgun approach as an efficient strategy to generate fast and reproducible sequence information [Messing, 1981 #231][Dos Santos, 1987 #250][Hunkapiller, 1991 #214][Fraser, 1997 #215]. In sequencing moderate-sized DNA regions containing

sequences greater than 500 bases in length, directed strategies are frequently used [Henikoff, 1987 #217]. The approach uses insertional plasmid cloning, primer walking and creating nested deletions. Several *in vivo* transposon-based sequencing strategies adapting the combined features of random and directed approaches have also been used [Adachi, 1987 #219][Kasai, 1992 #221][Berg, 1993 #220]. However, the *in vivo* approach presented requirements which limit its applicability such as the need for special host strains and traditional manipulation steps. Thus, the *in vitro* DNA transposition-based sequencing strategy was developed. This was shown to be efficient in sequencing repetitive DNA [Devine, 1997 #222].

In bacteriophage SfV, a large region of the genome has been sequenced, including the O-antigen modification and the site-specific integration region. Other portions of SfV genome sequence has been recently determined providing information on the essential early genes, the repressor locus, immunity and regulatory genes (Allison, *et al.*, submitted for publication). Sequencing of genomic portions has been based on a physical map composed of different *Bam*HI and *Eco*RI restriction fragments of the entire genome. In this study, we intend to proceed with the sequencing and characterisation of phage SfV genome focusing on the 13 kb SfV *Bam*HI restriction fragment A portion adjacent to the *pac* site.

3.2. Results

In our experiments, we used a directed approach to clone and sequence the initial 5.5 kb portion of *S. flexneri* bacteriophage SfV *Bam*HI fragment A. Southern hybridisation was then performed to map overlapping adjacent fragments. A 0.7 kb *Eco*RV-*Sac*I and a 0.6 kb *Hin*dIII DNA segment located near or at the end of the initial 5.5 kb Fragment A were used as probes for the hybridisation experiments. Detected segments were cloned and DNA regions beyond the hybridised segment were sequenced through primer walking using the whole SfV genome as template. A 10.1 kb sequence of the approximately 13 kb fragment A was derived. And as in the serotype conversion genes flanked by the SfV *Bam*HI fragment C which has been characterised, the sequencing of SfV *Bam*HI fragment A will contribute to the completion of the entire genomic sequence, an important preliminary step in extracting further information on the nature and properties of its protein products through sequence analysis and further works in proteomics.

3.2.1. *Bacteriophage SfV fingerprint for confirmation of its identity*

To generate and capture SfV *Bam*HI fragment A for cloning and sequencing, our SfV stock was propagated in attenuated *S. flexneri* serotype Y strain SFL124, recovered by polyethylene glycol (PEG) precipitation, and its genomic DNA purified by chloroform extraction and dialysis against Tris-EDTA buffer [Sambrook, 1989 #145]. Once isolated, the DNA was digested with *Bam*HI and *Eco*RI which produced DNA fragments identical

to the restriction band pattern visualised in an earlier SfV DNA digest (Figure 3.1A and B)[Huan, 1997C #230][Allison, 2000 #199].

3.2.2. *Characterisation of pNV728*

The 8.2 kb recombinant plasmid pNV728, is composed of pUC18 [Yanisch-Perron, 1985 #158], conferring ampicillin resistance, and the 5.5 kb phage SfV *Bam*HI-*Sac*I segment of *Bam*HI fragment A. pNV728 which was created by G. Allison, was transformed into JM109 and the resulting transformant designated as strain B823 (Figure 3.2). The insert corresponds to the initial segment of the 13 kb *Bam*HI fragment A adjacent to the D fragment. pNV728 was characterised by digesting it with different enzymes and observing which would generate fragments of appropriate sizes for subcloning and sequencing. The enzyme should have several sites in the insert and none in the vector except for the one in the polycloning site. *Bgl*II and *Cla*I sites were not present in pNV728 while *Sma*I cut once at the vector's polycloning site. These were shown by the presence of a large single high molecular weight band along the lane while *Kpn*I and *Pvu*II produced 6.1 and 2.1 kb fragments and 2.6, 2.3, 2.3, 1.0 kb fragments, respectively (Figure 3.3A). The double 2.3 kb band in *Pvu*II digest was detected based on band intensity comparison seen in subsequent digests (data not shown) and in consideration of the total size of the 8.2 kb pNV728 plasmid. *Eco*RV generated two bands and *Hin*dIII four bands of appropriate sizes which were ligated to pUC19 for cloning and for use as sequencing templates (Figure 3.3A and Figure 3.4).

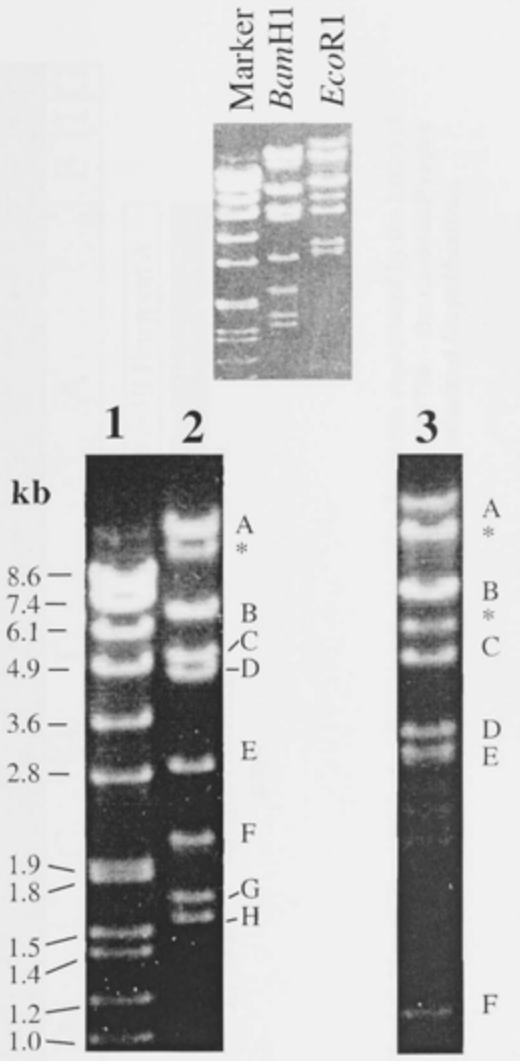


Figure 3.1. A, 0.7% Agarose gel electrophoresis of phage SfV whole genome digests. Lane 1, Marker phage SPP1 cut with *Eco*RI; Lane 2, *Bam*HI cleavage products: A-13 kb, B-6.7 kb, C-5.0 kb, D-4.6 kb, E-2.8 kb, F-2.1 kb, G-1.7 kb, H-1.5 kb; Lane 3, *Eco*R1 cleavage products: A-15 kb, B-7.4 kb, C-4.8 kb, D-3.3 kb, E-3.0 kb, F-1.2 kb. Asterisks represent submolar bands.

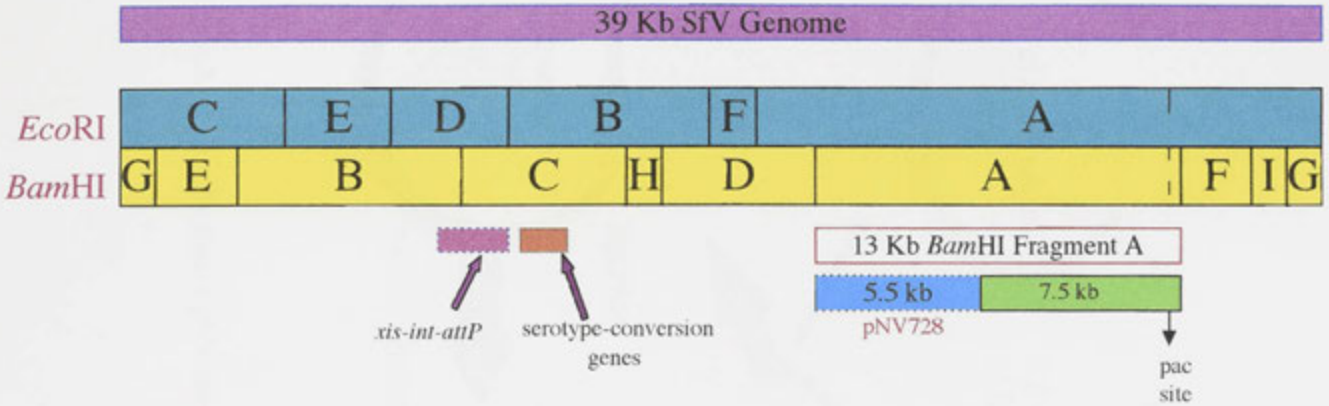


Figure 3.1. B, Physical Map of Bacteriophage SfV showing restriction fragments represented by the letters of the alphabet. Also indicated are the locations of *Bam*HI Fragment A, pNV728, the *xis-int-attP* region and the putative pac site (Adapted from Huan *et al.*, 1997 and Allison *et al.*, submitted for publication).

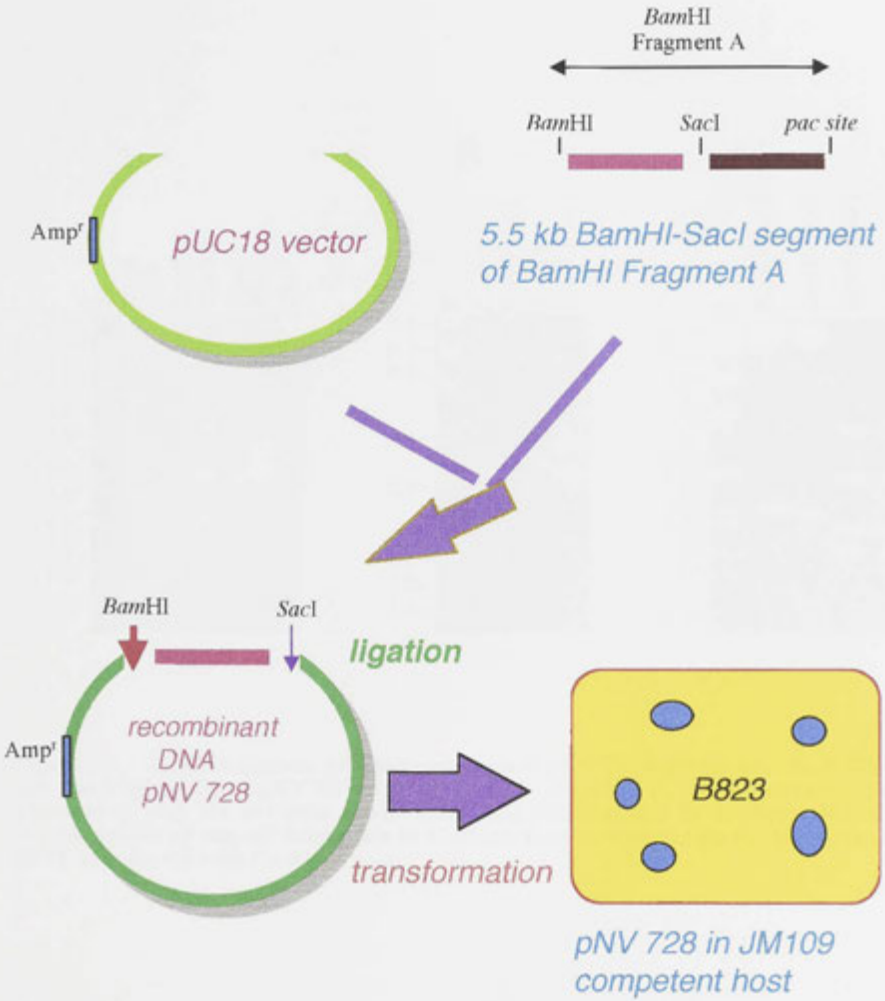


Figure 3.2. Construction of pNV728 and transformation into JM109

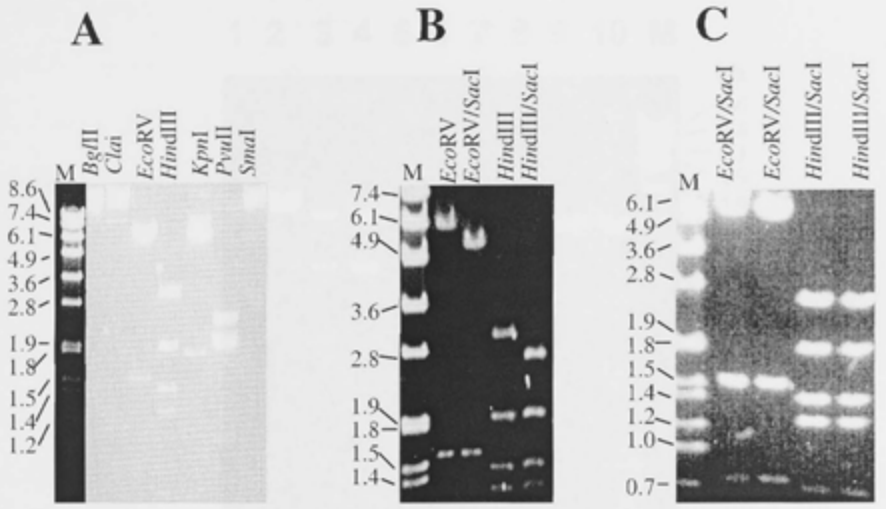


Figure 3.3. A, 0.7% agarose gel electrophoresis of pNV728 single digests. B, 0.7% agarose gel electrophoresis of pNV728 single and double digests. C, 0.8% agarose gel electrophoresis of pNV728 cut with *Eco*RV/*Sac*1 and *Hind*III/*Sac*1 in duplicate lanes. Note the resolution of run-off fragments in C which were not shown in B. M designates the SPP1 marker cut with *Eco*RI

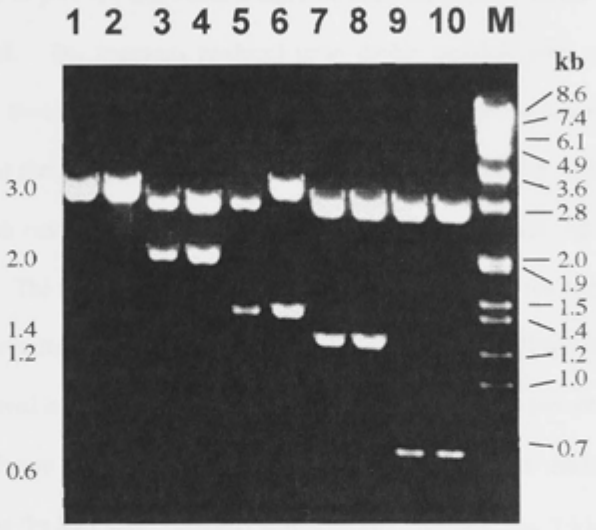


Figure 3.4. ***Hin*dIII digests of recombinant plasmids containing pUC19 vector plus the five *Hin*dIII fragments of pNV728: Lanes 1-2, pNV753 is self-ligated pUC19 with 0.3 kb portion from pNV728 insert ; Lanes 3-4, pNV751 with the 2.0 kb insert; Lanes 5-6, pNV749 with the 1.4 kb insert; Lanes 7-8, pNV750 with the 1.2 kb insert; and Lanes 9-10, pNV752 with the 0.6 kb insert. M, phage Spp1 marker cut with *Eco*RI.**

In order to sequence the SfV genome beyond the only *Sac*I site in *Bam*HI fragment A, pNV728 was mapped to locate fragments near the *Sac*I site which can be used as probes that would detect fragments flanking the *Sac*I site. This was performed by doing a single enzyme digestion of pNV728 and a double digestion using the enzyme which was used initially plus *Sac*I. The fragments produced upon double digestion were considered candidate probes since these were generated only in the presence of both *Sac*I and the initial enzyme and that these fragments must be bounded by the *Sac*I site. Figure 3.3B showed a ~0.7 kb run off band which was deduced when pNV728 was digested with *Eco*RV and *Sac*I. The size was estimated from the difference between the larger *Eco*RV fragment and its resultant size after digestion with *Eco*RV and *Sac*I. When the double digests were resolved in a higher agarose gel concentration of 0.8%, the run-off fragment was visualised (Figure 3.3C). This 0.7 kb *Eco*RV-*Sac*I fragment was used to probe fragments flanking the *Sac*I site. Similarly at this gel concentration, the 0.6 kb run-off fragment of the *Hin*dIII-*Sac*I double digest which was not seen in the single *Hin*dIII digest, was visualised (Figure 3.3B and Figure 3.3C). This 0.6 kb *Hin*dIII fragment was also utilised as probe.

3.2.3. *Sequencing of the 5.5 kb portion of BamHI Fragment A*

*Hin*dIII digestion of pNV728 yielded fragments of appropriate sizes for cloning and sequencing. These fragments ranging in size from 0.5-2 kb were chosen because the

integrity of sequencing signals in the sequencing reaction covers around the first 500 bases of the template past the primer. Therefore, if we could use a template with a 1 kb insert, the entire insert would be sequenced in a single sequencing run without the need for primer walking. One 3 kb pNV728 digest fragment containing the vector and a 0.3 kb DNA from the 5.5 kb fragment A was also purified from the gel, self-ligated and transformed into JM109 (Figure 3.4 lanes 1 and 2). The other *Hin*dIII digest products from pNV728 with molecular weight sizes of 2.0, 1.4, 1.2, 0.6 kb (Figure 3.4) were eluted from the gel, ligated with *Hin*dIII cut, gel-purified and dephosphorylated pUC19 vector, before transformation into rubidium chloride competent [Davis, 1990 #213] JM109 host. Five white colonies were picked from each ligation. These likely transformants represent strains which were not capable of α-complementation indicating carriage of the recombinant plasmid. After alkaline lysis minipreparations of the plasmids, uncut forms were visualised on agarose gel and those which appeared to have acquired the insert were selected. These were cut with *Hin*dIII and the insert sizes verified by agarose gel electrophoresis (Figure 3.4). The successful transformants were strains B860 containing pNV753 with self-ligated 2.7 kb pUC19 vector and 0.3 kb insert, B858 containing pNV751 with the 2 kb insert, B856 had pNV749 with the 1.4 kb insert, B857 had pNV750 with the 1.2 kb insert, and B859 had pNV752 with the 0.6 kb insert (Figure 3.4). These plasmids served as templates in the sequencing reactions of pNV728 inserts initially using the universal forward and reverse M13 primers. Amplified extension products were purified using the sodium acetate protocol in preparation for sequencing in the ABI Prism 377 DNA Sequencer. Primer walking using custom primers

(Life Technologies, Inc., USA) were used to fill in gaps and to sequence both strands of the fragment. At this stage, the initial 5.5 kb *Bam*HI-*Sac*I portion of phage SfV *Bam*HI fragment A in pNV728 was sequenced and a physical map is presented in Figure 3.5.

3.2.4 Restriction probes detected fragments flanking DNA immediately downstream of the initial 5.5 kb portion of BamHI Fragment A

To proceed with fragment A sequencing beyond the lone *Sac*I site of the 5.5 kb pNV728 insert, we identified and cloned fragments adjacent to and overlapping the *Sac*I site. Two probes were chosen based on the proximity of their location to the *Sac*I site of the 5.5 kb *Bam*HI-*SacI* portion of pNV728 insert deduced from restriction mapping (Figure 3.6). The first probe was the 0.7 kb *Eco*RV-*Sac*I that was not produced when pNV728 was cut with *Eco*RV alone (Figure 3.3A and 3.3B). When comparing *Eco*RV digests of pNV728 to *Eco*RV-*Sac*I double digests, a ~0.7kb reduction in the size of the largest fragment was noted suggesting that an *Eco*RV site was 0.7 kb apart from the *Sac*I site (Figure 3.3B and 3.3C). Hence this 0.7 kb fragment was used as a probe. Another probe was used that provided additional supportive information in the analysis of the Southern hybridisation results. This was the ~0.6 kb *Hin*dIII digest fragment of pNV728 (Figure 3.6). Assembly of the sequence of the 5.5 kb *Bam*HI-*Sac*I fragment indicated that the 0.6 kb *Hin*dIII fragment is adjacent to the 0.3 kb sequence of pNV753 (bounded by the *Sac*I site), making the 0.6 kb *Hin*dIII fragment another suitable probe to use (Figure 3.6). Southern hybridisation was performed using these ^{32}P-labelled probes

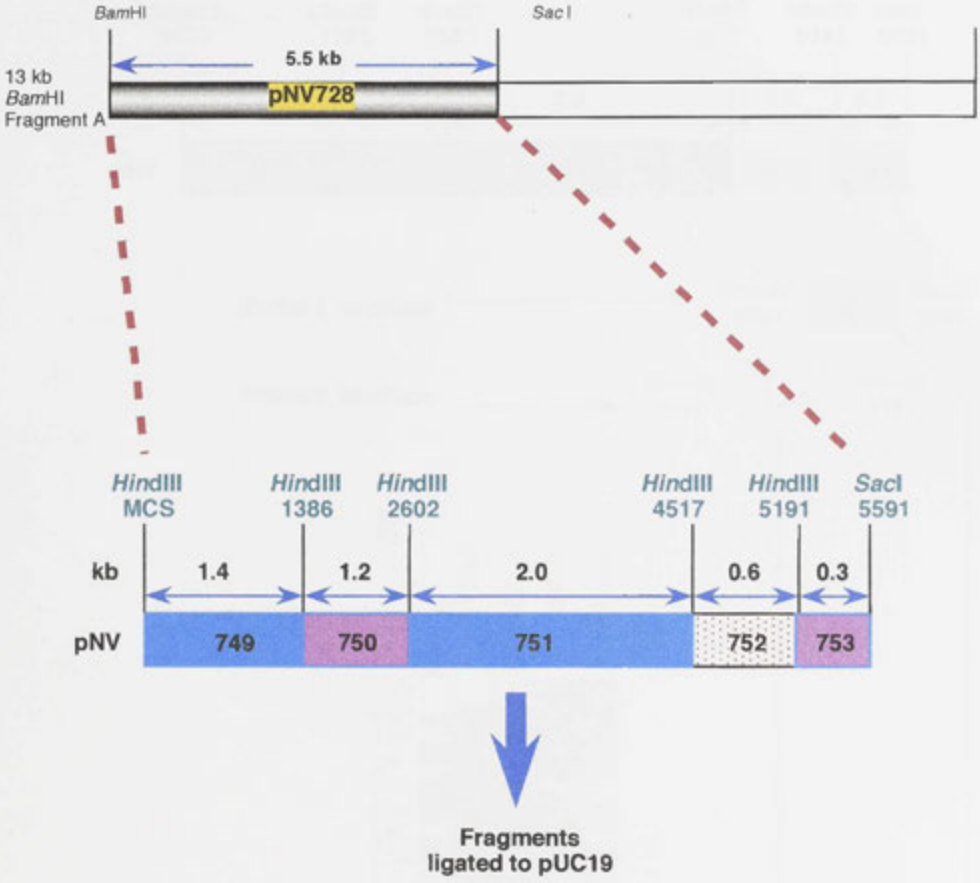


Figure 3.5. Genetic map of pNV728 insert containing the DNA of the initial 5.5 kb portion of phage SfV *Bam*HI fragment A. Relevant restriction sites are indicated. MCS is the multiple cloning site.

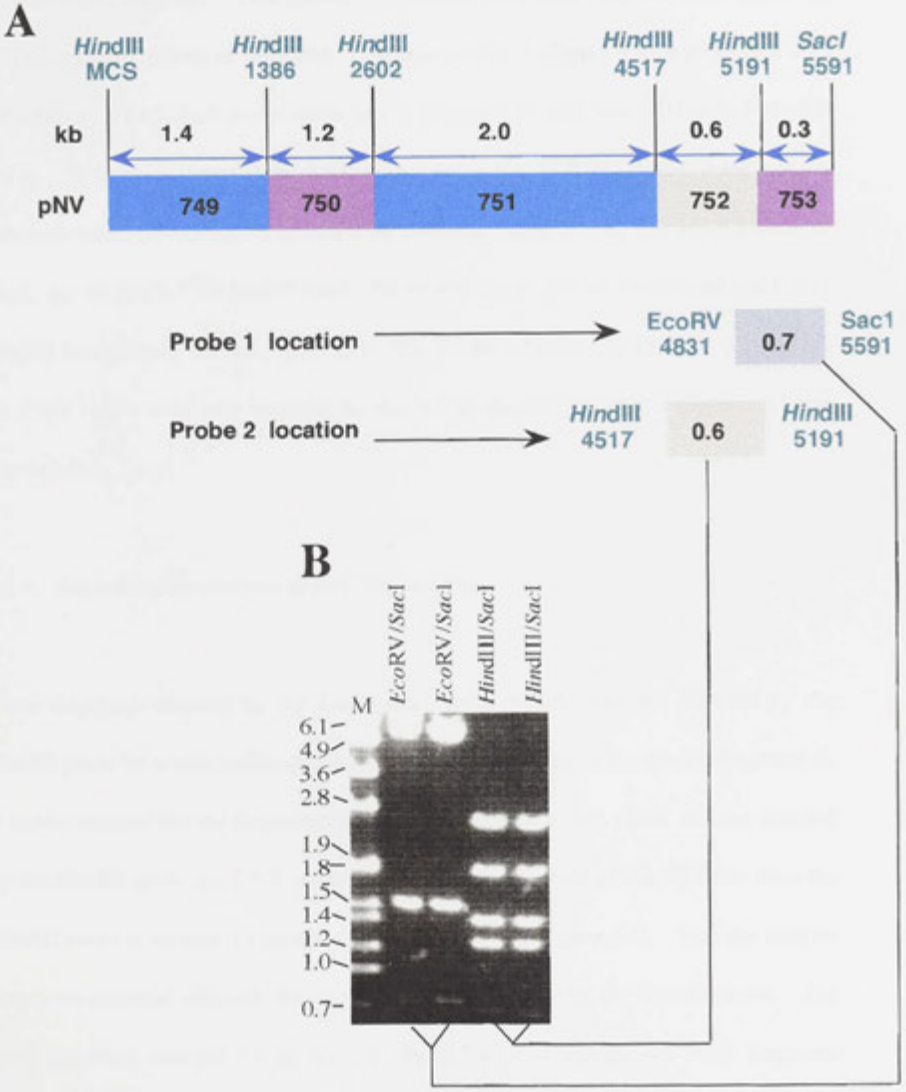


Figure 3.6. A, Location of the 0.7 kb *Eco*RV-*Sac*I probe and the 0.6 kb *Hin*dIII probe relative to the genetic map of pNV728 insert. B, 0.8% agarose gel electrophoresis of pNV728 double digests showing the 0.7 kb *Eco*RV-*Sac*I and the 0.6 kb *Hin*dIII fragments which were radio-labelled and used as probes for Southern hybridisation.

which hybridised with alkali-membrane blotted fragments of the SfV genome digested with different enzymes. Two identical membrane blots were prepared each containing various enzyme digests of SfV DNA. Membrane blot 1 (Figure 3.7) was probed with ^{32}P-labelled *Eco*RV-*Sac*I probe while blot 2 (Figure.3.8) with the ^{32}P-labelled *Hin*dIII probe. Fragments detected by both probes were the 13.0 kb *Bam*HI band and its submolar band; 18.0 kb *Eco*RI band and its submolar band; 20 kb, 6.5 and 5.1 kb *Kpn*I band; 6.1 kb and 5.3 kb *Eco*RV band; 7.0 kb *Pst*I band; 2.0 kb *Pvu*II band and 0.6 kb *Hin*dIII band (Figure. 3.7 and Figure 3.8). The 2.0 kb *Hin*dIII, 2.7 kb *Pst*I, and the 4.2 kb *Pvu*II bands were only detected by the 0.7 kb *Eco*RV-*Sac*I probe (Figure. 3.7 and Figure 3.8).

3.2.5. *Sequencing downstream of pNV728 SacI site*

Three fragments detected by the *Eco*RV-*Sac*I probe which were not detected by the *Hin*dIII probe were selected for cloning and further sequencing of the *Bam*HI Fragment A. It can be deduced that the fragments detected by the *Eco*RV-*Sac*I probe and not detected by the *Hin*dIII probe are DNA segments extending to the right of the *Sac*I site since the *Hin*dIII probe is situated 0.3 kb to the left of the *Sac*I site (Figure 3.6). Had the selected fragments extended leftward, they would have been detected by the *Hin*dIII probe. The three fragments were the 2.0 kb *Hin*dIII, the 2.7 kb *Pst*I and the 4.2 *Pvu*II fragments (Figure. 3.7, 3.8). The location of these probe-detected fragments relative to the *Bam*HI fragment A of phage SfV genome is shown in a schematic representation of the

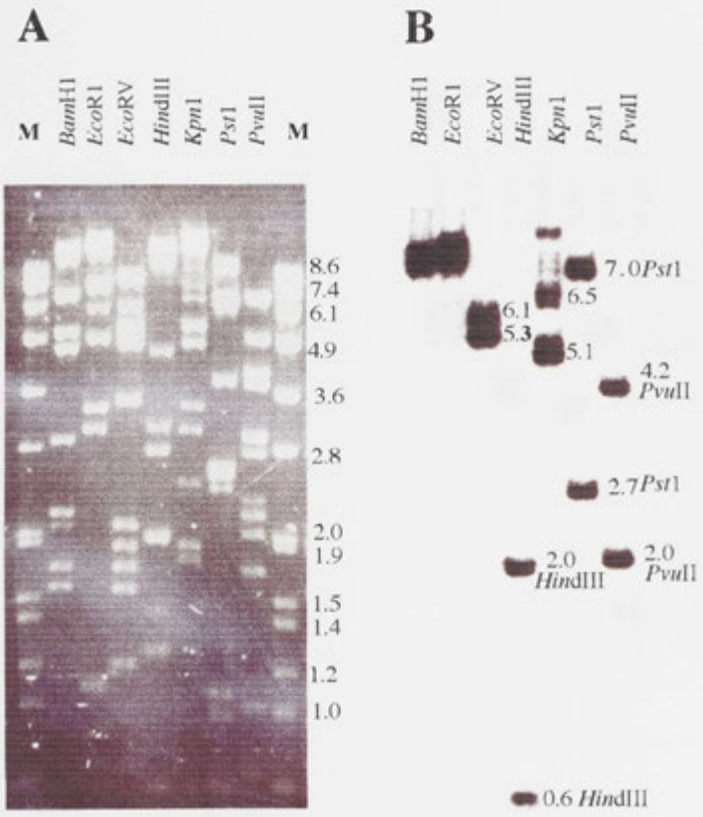


Figure 3.7. Southern hybridisation of the SfVgenome. A, 1.0% agarose gel electrophoresis of SfV DNA digest fragments. B, Autoradiograph of membrane blot 1 showing fragments from gel A that hybridised with a 0.7 kb *Eco*RV/*Sac*1 radioactive probe. (see also Figure 3.8).

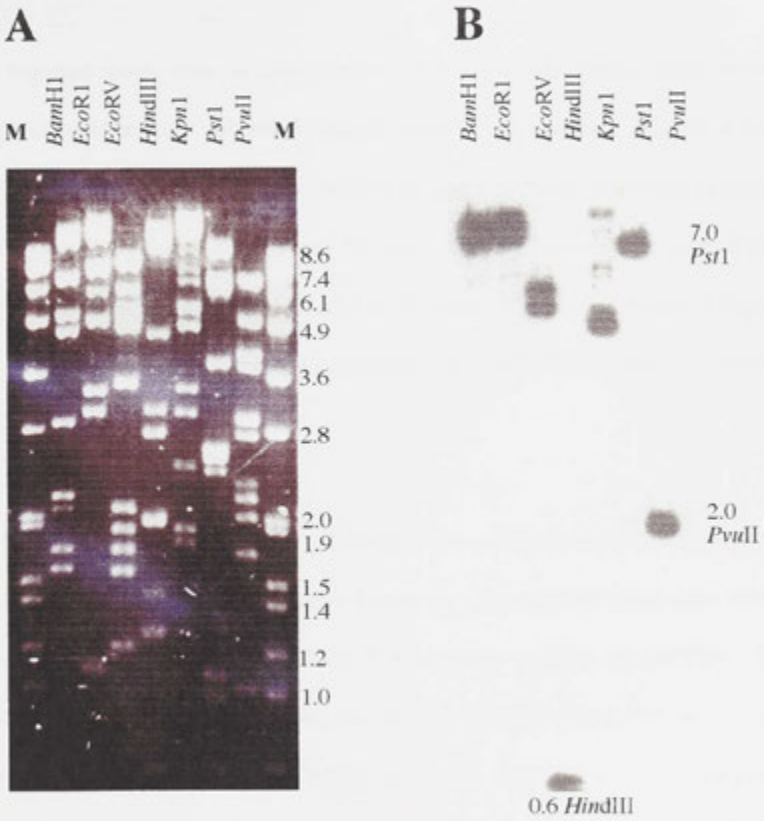


Figure 3.8. Southern hybridisation of the SfVgenome. A, 1.0% agarose gel electrophoresis of SfV DNA digest fragments. B, Autoradiograph of membrane blot 2 showing fragments from gel A that hybridised with a 0.6 kb *Hin*dIII radioactive probe.

combination of sequence and hybridisation results in Figure 3.9. Sequence information from the overlapping fragments provided a basis that ensured our sequencing through the right side past the *Sac*I site of the SfV *Bam*HI fragment A. Sequences beyond the 4.2 kb *Pvu*II fragment were read through primer walking.

All the sequence results were initially proofread and edited individually, based on the sequence information from the complementary strand of the genome, prior to a final assembly using 2D WebAngis fragment assembly program. A single contiguous sequence of 10109 bases was derived from a total 58 input sequences including 4 major edited contigs (editcons1 to 4), and utilised a total of 34 customised primers shown in Figure 3.10. The complete 10.1 kb sequence is presented in Figure 3.11 showing the relevant restriction sites.

To confirm the identity and the correct continuity of sequencing from pNV728 through to the *Sac*I site and the three probe-detected fragments, relevant restriction sites which were separately predicted by the WebAngis WAG mapsort program for pNV728, the three probe detected fragments containing the *Sac*I site [2.0 kb *Hin*dIII (pNV756), 2.7 kb *Pst*I (pNV755) and 4.2 kb *Pvu*II (pNV754) fragments], and the 10.1 kb contiguous sequence were compared. Four restriction sites in the order of *Hin*dIII, *Pst*I, *Pvu*II, and *Sac*I sites inherent in the pNV728 insert were also found in the same ordered location in pNV756, pNV755, pNV754 and the 10.1 kb fragment A sequence (Figure 3.9 and 3.11). The pNV756 *Hin*dIII fragment had the *Pst*I site at nucleotide position 18, *Pvu*II site at

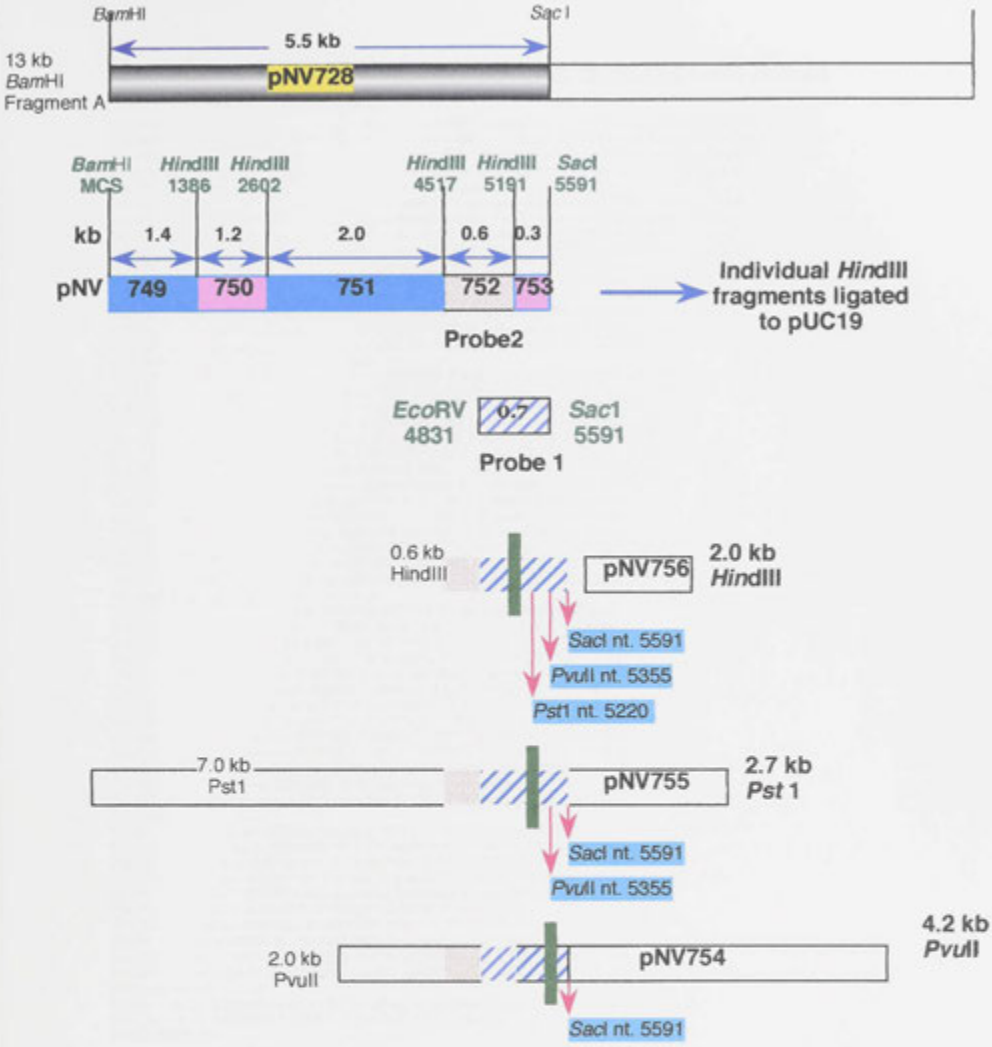


Figure 3.9. Schematic representation of the membrane blotted fragments detected by the 0.7 kb *Eco*RV-*Sac*I and 0.6 kb *Hin*dIII probes. Relevant *Hin*dIII, *Pst*I, *Pvu*II, and *Sac*I restriction sites on the 0.3 kb *Hin*dIII-*Sac*I pNV728 fragment with their corresponding nucleotide position in pNV756, pNV755, and pNV754 are indicated by red arrows. The nucleotide position of restriction sites were predicted by the Mapsort program of the WebAngis WAG package. The green solid block serves as boundary line distinguishing between 0.7 kb *Hin*dIII and 2.0 kb *Hin*dIII fragments; 7.0 kb *Pst*I and 2.7 kb *Pst*I; and the 2.0 kb *Pvu*II and 4.2 kb *Pvu*II fragments.

Figure 3.10. Gelassemble bigpicture of the single contig 10,109 kb bacteriophage SfV *Bam*HI fragment A showing the primers used with its location and template host. Arrows indicate sequence location, length and direction of strand replication. Total contig input was 58 and editcons 1-4 represent four major edited contigs.

Figure 3.11. The linear sequence of the 10109 bases of phage SfV *Bam*HI fragment A showing the relevant restriction sites used in mapping as predicted by the Mapsort program. the restriction sites areindicated in red text and the enzyme recognition sequences in italicised bold print.

BamHI

```
   1  GGATCCCCTG ACGCACCAGC ATCTGCATCA TGTTCTGGAA ATCAGCCGTT
  51  GTACCGGGTA GCTGGTTACC CAGGCCAATA GCCAGTTTAT TGATGTCCTG
 101  AAAGCTCTTT CCAACCTCGC CGTTCGCATC CATCATGGCG ACTTTCAGCC
 151  CGGTGGCGGC GTTTTCCTGA TCGGCATAAG ATTTCAGGGA AAGCGTCAGC
 201  CCCGCTGCCA GTCCGCCACC AAGCGCCAGC CCACCCTGTG ACGCTTCTTC
 251  CGCCTGGCGT TTAAATCCCC GGATTTTCTT TTGCATTTTC GACAGCGCGG
 301  GAGAAAGCCT GTCGACACCG GTGATCAACG CCTTAAGCTC AAATTCAGCC
 351  ATGTGTGCGT TTCTCCTGCT CTATCCTGTT TGCCTGACTG ACCAGTAAGG
 401  GAATTTCACT GATCGGCATA TTCAGCAATT CGAAAGGATT AATGCGCCAG
 451  TAGCTGGCGC AGTCAAAGAA GCGATCAGTG AGGTATTCAG CCGTCAGGCC
 501  TGGAGGAAAA AACCAGCCAC AAGCCACGCC GCTGCATTCA GGTCTGCCGG
 551  AGACATCTGG TCGACAGAGC TTTGCGGCAC TTTCGCCAGC CGCACAATGT
 601  ATTTCGACAC CACATGCGCC AGAAGTCTGA CTGACTCATC CTGATTCATC
 651  TGGTAGGGAT ACCCCAGCTC GCGGACATCC TTCCCGCTGG GTTCATCAAA
 701  CTCCAGTACG GAGAGTGTCT CGCCATGAGC AGTAATCGGT TTCTTTAACT
 751  CAAGCTCTTT CATTACTGGT AATCCCCTTC TTCACCGTGG AACTCAAGAT
 801  CAACCGTACC TTCTTCGGCA TTATGGTTCG CTTCGCCGTG CAGCCAGGCT
 851  GACGACAATA CATAGACCTG ACCGTTCGCC AGCTCGGCAG TGATGGTCAT
 901  CTCATCAGAC GAGGTGATTT TGCTCACCGG AAAATTCTTC GGCACCTTGA
 951  AGGTCCCTTT GACATAAGGC GCACGGTGAG TTTCCTTGCG GTCCACTGAA
1001  CCGTCCAGGC CGATGATGTC ATCATTGACC GTCCTGTTCA TGGGCACCTC
1051  AATGCCGCCG GTCAGCGATA GCTGCTGACC GTCAATTTTG AAATAACAGG
1101  TTCCCCCGAT ACGGGCCATT ATGCAGACTC CTCTGAATAC TGAAGACGGA
1151  ACTGGTTAAC CACGGCAAAA ACACGCAACT GGTTAACATA GTCAGGCGGG
1201  AACAGCGTGT TCAGGCGGTT CGGAACGCTG GCATCACGCT CCACAACCAG
1251  GTACTGCTTA AACAGTTCGT AGTTTTCCAC GATCCCCGCA CGCTCAAGCT
1301  GACGGTAGGT TGCCAGCAGT TCCCCTTTGA TTACCGCCGG GGTGACAATC
1351  GCCTGACCGG GACCAAAGCG GGTACCGTCG CTGGCAAGCT TGTGACGCCC  HindIII1386
1401  GTACTTACTG GTAATGACGG ATTTCAGTTT GCGCAGTACA TACGCACTGG
1451  TATGCAGCGT CTCGCTGTCG AGGTAGCTGT TATCCGCAAC CCCGTAAGCA
1501  TTTTTCCTGT ACGTGGTGAC ATCACGCTGA ATGCGCAGCA CCCCGCTTTC
1551  GACATACGCC GTTGCCACGC CATGAGACAG CAGGGTCTGC TGCTCGGTCA
1601  TCGTGAACCG TTTCCCCTTC GGCGCAGGCA GCATACCCAC CAGCTCACCG
1651  GTCTGCGTGG GACGTGCCGG ATCGTTGCGG ATAAACACCG CTGCACGGGC
1701  GGTACGGCTT GCCGCCAGCT CGTCGGCAGG CGTCTGGGTC TCTTTTCGT
1751  ACCCCGCCAG GGTGATGTGC TGCTGGTTAA ACTGGTCACC TGCGTTCACC
1801  AGTTCTGACA GTGTGCCGGT CTTTGCCGTA TACACATGAC CATACAGCTG
1851  ACGCGCATAG CTCCAGCGAC CGCTGGTATC GTTCATCTCG GTCACCAGCG
1901  TGTTAACGGA GGCCGTGTCG TTGAACGGCA GACCGATATA ATCAAACGGC
1951  TCATCCGCCA TTGCAGCCAC CGCGCCGGTG AGAACCGGAG CGCCCGTTCC
2001  GGCGGTCCCC GTCGCCACGG CAATCTGTAC GCCCGCTGGC AGCACTTCGC
2051  CCCCACCGAA GCCGTAGTAA TTGAGGCTGA CAGGAATTTC ATTCCCGCAA
2101  AGCCCCTTAT GACGCGCGGT CAGCGTGACA ACACCAGCCG AAGATGAAGC
2151  TGTAAACGGC AGAGTCGGAA CGGCATTGAT GGCATCCTGG ATACTGCTGG
2201  CAATCGTCGT GACGTTATCG CCGTTGGTCA CCGGAGCCTG CACGCGGGTA
2251  CGTCCCACAT AGACATTCAC CGTGCCGCTT TCGGTTGCTT CCCCGGTCAC
2301  CGTCAGCGTA ACCGTTGCCG CCGCGCCTGT GGCTTCCGGA ACGGCAATCA
2351  CATACAGCTC ACCAAACGGG TCGGTCTGGC GATAAGCCTC GACCATACGC
```

```
2401  GCCAGCTGAC TTCCCGCACC ACAAATCTGG CGTGCATAGT CTGCCGACGG
2451  CATCAACACC AGACTGTTGG CAACAATCTC TGCACCGTTA TTGGCGTGAC
2501  CAATCAGCAA CGATGCCCCG CTGTCCTGTG CAGTATTCGC CGCCTGGTTA
2551  TCCATTTCCG CATAAAACAG CGGAACCAGC GTATTCGACG GAATGGTGTT
2601  AAAGCTTATC GTCATCGGTG TTCACCTTTT TATTAACGCG CCGGATATCA  HindIII2602
2651  CCCGCTGCTT CACGGCGCAG CCAGTAGTTG TTCTCGTCAA CATTTCGCCC
2701  TTCGGCGGGC AAAAGGTCAC CGCGGGCAGG GTCAGGCACT GACCGCCCTT
2751  TAACAGGTTT CACAAACATG AAGATTCTCA GGAAGGAAGG GTTATTTCGG
2801  TGTGATGTTC GATATCGCCG TCAGGCCCGT TACCGGGATC GAGATAATCA
2851  ACATCAATCG CCAGCGTTCG CAGTTCATCC AGACTGTTCA GCTCATCCTG
2901  CTGGCGGGTA TCGTCTTCGG TCAGCTCGCT GATGACCGAA AAATCGAACT
2951  GATAAATCAG CTCATGACGA TTCAGATCCA GCAGCGTGCC GCCGTCATAG
3001  GTAATCGGGT TACCGCACGC TTCCGGGTTC CAGCCCAGCA GGGCCTTAAA
3051  GAGCATCTGC CGGACATCGT CCACCACATC ATACGAAGCA AACTGACCGC
3101  GCTCATCACG CCCGTTACTC AGTATGACAA CCACGGAGAA GCCCTCTTTC
3151  AGCTCCTGCC AGTAGTCGGT CTGGCTTTTG TTTTCTCCCG GAGAGTCATC
3201  ACCCGGTACC ACATACGCCG CCGGGAGTCT CAGCTTTCCG ACCTCCGGCA
3251  GATTTTTGAA CTGTGCCGCG CCTGCCACCC GGTTTTCAAA ATACGGGCAG
3301  CGGGCACGCA GCGCAGCAAT AACAGGCGTC AGTTTCATCT GCGTCGTCGC
3351  TCCGGCTTCA GTGATTTACG TAATTCCCGC GCCAGAAAAT AGCGTGTCCA
3401  GCTGCGGTTC TTTTCAAGCG TTTCCACCAT AAAGTTATTA CGTGGAGCCA
3451  GCCGCCAGCC GCTGCCACCG GATGCACCAC GATGATGACT ACGACGACGT
3501  TTTGCTCCTC CCCGGACGCC ATAGAACAAA AAAGCCGGAT AAAAATCACC
3551  GGTGATGCGG CGGTTTCCCT CACCATTACG CTGGTTAGGG GCTATACGTG
3601  CCATAAAACC GGGGCGATGT TTACTGGCTC TGGGTACCAT GTAACCAATC
3651  GAACGAGCCA GGCGTCCGGT CTGATAACCG GGGTTTTCAC CCGGTGCCGA
3701  CCGCGCATGG CGCATCACCA GCCGACGGGC ATCACGCATA TGACGCTGAC
3751  CAATCGTGAC AAACGCCCGC CGGACACGGG CGCGGTTAAA GCGCATCTCC
3801  GCGGGCTGCT GAAAATCAAC GTGCAAAAAG GAAGTCGTCA TTGTTGCCTC
3851  CGTGACTCTG CCTACATTCG CCCAGCTCCG TACACTCCAG CAGCAGAAAA
3901  CGCCGCGCCC CGTTCAGATC GCGCTGACGT TTCACCCGGT ACACACTGTC
3951  ACCGCAGACC ACCTCATAAT CAGTGGTGAT CCCCCGGCGG TAACGAATGG
4001  TGATGTAATG GGTGATGGCG TCCCCGGTCT GCGCGGTTTC CTGCCAGGTG
4051  GTGGCACTGG TCTGGATAAC CTTCGCCCAT GTCCGGAACG TAACCGGGTA
4101  TTGAGGCTCC ACGCCAAAGT TATCCGCGGG CATATCCACC CGCAGGCGGA
4151  TCAGGACGCG TTTATTCAGT TCGCCGGGGT CCGGCAGAAT GTAGGTTGCG
4201  CTGGTCTGTG CCTGACGAAT TTTCATAGTG GTATAAGGCG ATAAGGAGCA
4251  ACCAACCAGT TAAAACTCAT TGGCAACTCC ATTTTCTCAA CGTCTGTAAC
4301  CGTTGAGCGG TTTTCGTAGA AATGGCTGAC AAGTAGCAGA AGTGCCAGCT
4351  TCACATCATC AGATATCACA AGCCCATCAG GATCATCCGC AGGCCTGTCA
4401  TCTGCGGTTG CATACAACTT ACGGTTAAGG AAGTTTTCCG TTCGACTCTG
4451  AGCGGCCTTA CCAAGCAGTT CAAGCAACTC ATCTTCATCA GAGAAATCAT
4501  CATCCAGACG GAGCTGAAGC TTAATCTCTT CCATTTTTAA CAGCATAAAA  HindIII4517
4551  CCTCCTGTGC CCGCCAGAAC GCGGGCACAA AAAAACCGCA TTACGCGGCG
4601  TGCTGTATTA CGTAAAAAGA CTAATCAACC ACCAACGCTA CCTTTCCCCA
4651  CCAGCGCTTT AATGGCAGAG GTGTCTTCCA GGATACAGTC AAAACGATGG
4701  AAGGCCAGAA AACCGGTCTG ATCATATTCC GCGTAACGCT CAACCAGACG
4751  TTTAAGAATC ATGTATCGCA CACGACGGAT AATGAAGCGA TCAAAGTCAC
4801  CACAGAACAT GAATTTTTTA CCCGCCCCGA TATCATCAAT TTCCTGATCA  EcoRV4831
4851  ATGACATACG GTACATTCAA CACTGAAGCA GGTGCCACAC CAACAATATC
4901  CGGCAACCAT AAAGGGCGTC CCTGACCGTC TTCCATCTCA CTGATCAGTT
4951  TCAGCGTATT ATCGTTAAAC GCCAGGCGGA ATTTCGGTCC GCGACGATAT
5001  GCAGGATCAA TGCTGTGTTT CAGAGCCAGA ATTTCCTGCC ACTTCACCGC
5051  ATTTGCCGCG GCAGTCTGTG TTGTGCCGGT CACTGATGCT GCCAGCCCTT
5101  TGGGTTGTTT AGGCGTACCA GCACCCGTCC CCTGAATCAG ATAACGGGCT
5151  TCACCACGAC CAATACGTTC AGCAATGCGA CGGGCAAGAT AAGCTTCCAT  HindIII5191
```

```
5201  ATCGATCGCG CTGTCCTGCA GCAACTCATT AGACACACGA ATGATTTTCG PstI5220
5251  ATGTCATTTT GAGCGCCCCA AGACTTCCCA TACCGAAATC GGTGTCTTCT
5301  TCACCGGCTT CTTCATTTTC GCCCAGCAGA ACACCAACTT CGGAAGTACC
5351  ATCAGCTGTT GCCCACTCCA TGGTGCGACC GTCAGAAGTG GTCAGAATCT PvuII5355
5401  GCGCCACACT GGCGATGCCA CCGTAGGATT TCATCTTCTC AACAACTTTC
5451  GCCAGGAATG TTTCTGGTAC GGTATATCCG CCCTTTTCAT CCTGAGCTAC
5501  ACCCTGGGCA CGAAGTTCAC GCAACGCCTT TCGCTCTTCT GATGTCAGCT
5551  CACTGGCACC GTGACGCATC CACTTATCAA AAACCTGAGC TCGTTTCTCA SacI5591
5601  TCCTGTTGCG AATTGTTTTC CGGATCAAGA TTCTGACGCT GCTCTTCCTC
5651  ATTGCTTTCA ATGTACGCCT GATCCTGACG ACGCAGTTCT TCTTCGCGTG
5701  CAATTCGTTC ATCAAGCGCT TCCAGTTCGG ATTTTGCTTT GTTCCACTCC
5751  GTGCGCTGCT CTTCCGTCCA TGCGTTATCA CCAATTTTTT CATTCAGGGC
5801  GCGCATGTCA GTTGCGATAG TATTACGTTT CTGTTTCAGT TCATGCAGTT
5851  TCATGATGTT TCCTTTACGC GTTAAGAAGG GTCAGGACGC GTTCACGCGC
5901  CATACGTTGA TTAATGGCTT TCTGTAGCGC GCCGCTGTTG CGCGCCTCCT
5951  GCCATGCTTT CATGGAGCGA ACAGCCGAGT CAGCCTCCTG ATAGGCAGGA
6001  TATGTCACAG GACTGACATC CAGCAGACGG GAAAAGCGGG TTATCTCGCG
6051  AATAACAACC CCGTCCTCAT CCTGATACCA CTCCTCACCG TCACGGGCGA
6101  CACGGAAAGC GAAAGATGAC TGGTTAATAT CTCCACGTTG CATCGGGGCC
6151  AGCACCAGAT CACGAATGGT CTGTGTCTCC GGAGCCTGGA TGTCATAGCG
6201  TAATCCGCGC TCATCAACTG AGAGATTCAG CGTGCCTGCT GCACTACGCC
6251  CAAGAATAAA ATTAGGATCG TGGTTAAACA GTGCGCGTAC ATCATCACCA
6301  AGCACATCGT CAAAAGCGCC GGGCCGGATG ATTTCGCGGA ATGAACCGAA
6351  TATCAGCTCA GAACGACAGT CAAACACCGA TCCATAACCG ATAATGTGCG
6401  CCGGGTTATC GTCATGCCTC TCAGCACGCA CCTCACCGCT GTAACAACGG
6451  ATTTCACGGT CATTCATTGG TTTTTCCCTC ATCGTTTTTT GGGGGCTTAA
6501  AATCTCCTGC CGGGTTAGCA GCATTCACGC TTACCAGCAT CTCGTCCAGC
6551  CCTTCAACCG GATTCATATC CTCGAATGCG CGGGCCTCAT TACGGCTCAT
6601  CCATCCATCG GTAATAGCGA AGTGATAGAA TTGCGCGCGC TCCTGCGGAG
6651  TTCCGCGTAA AAGCCCCGTC AGATTGAACC TGACGTAATA CCCGGCGGCT
6701  AACTCAGCGC GGGTAAACAA GCGACGGTTA AGCTCCTGCT CCCAGTTCGT
6751  CACCCACGGC ATCATCGTGT AGCGGACAAA CTGAATCGCC TGCGCAGAAA
6801  TATTGGAGAA GGTGGCTTTT TCGAGGTCAT TAATCATGTG CGCAGGAATA
6851  TTGAAAATAC CGGCGATCAT TGAACGGTTC AGCTTCATCA TGTCAATGAT
6901  CTGAGCGTCA ACTGGCGACA CAGTCAGTGC CTTGTAATCC AGATCGGCTG
6951  GCAGCAGCAT GGTTTTGTTT TCCTGGCGGC GTAACGCCTG CGATGCCTTC
7001  TGCCACTGAT CTTTAAGCCA GCCCCAGCTT TCCTTATTGA GTCCGCTTTT
7051  AACGGATACT ATCCCCGCCG GACGGGCATT ACCGCTGAAG AAGCTTTCTG HindIII7091
7101  TGTACTTCTG ACCGCTCATC CCCATGCCTA TTGTTTCGGC ATGTTGCATA
7151  ATCGGACTCA GCCCCATCTT CTGATTATTA CCCAGCGCAC GGATGTGGAT
7201  CATATCGTCC GGACTGATCG CAAACGCCCC ATATTCGTTG TACAAACCGT
7251  AGGTATATCG GCCACCAGTA TTCATCAGCG TCGTTTCCCA CGGCATACAG
7301  CAATCCAGGG ATATGACTTC ACCGCGACGA TTACGTTTCA CCCAGGTATA
7351  CCCATTCCCC CAGCCAAGGA TGTGACGTTG TTTCAGTTCG CGCCATTTGT
7401  AGCTGGTTTG CCAGGTATTG GGCTCATCAT GAACCAGATA AAACGCCGGA
7451  TGATCGCGTG CGGGTTCAAC CTTCCCCTTG TGCCTGCGCA TAACATGCAA
7501  CGGCATCTGG GCAAGGCTGG AAGACAGGAC ATAGATACAG GAATACACCG
7551  CAGCCAGTTT CATCGCAGTC TCAGGACTGA CATAAACGTC TGCCCGGAAC
7601  AGCCCATCAG TATCAACGGC ATCACCGGTT ATCGGGGTGG AAGGATTCTC
7651  CAGTGATTTA CTTCTGAACA GAGCATCAAG CAGCACGCGT CCCCCTTCTG
7701  GCCATAGCCA GTGCGCCCAC CAGCAGTAAA GCACCGGACA AAATCAGAGC
7751  CGGAGCCATA CCAAACTGCA GGTAAACCCC GCACGTAAGC AGGCCAAAAC PstI7770
7801  CAGCCAGCCC GATAACATCA GCAATTAGTG ATTTCATAGA ATTAAGAGAT
7851  CATCGTCCGG ATCAAGAGAT GAGAGGAAAT CGTCAGGTTC TTTGAGCATT
7901  GCCCGACCGA TCGTCATAAT CAGTGCAACC GCACCATCGA TTTTGTTTTC
7951  CGCCTGCTCC TTGACGGGCT TCACTAAATC ATCGTTACCT GGCATGTTTT
```

```
 8001  TGCCGACCAC ATTGCCGATA CACCAGGTCA TGATGGGATT GCCGTCATGA
 8051  TGAAAGCGTC CCGATTCAAT CGCTGCTTCC AGCTCTTTCA TCGGATCGGA
 8101  CATATTGGCG AAGTTCTGGA CGATAGTGAC GGGATTCAGA TCTTCATCAG
 8151  CAAGGTCATG TGACAGCCCG GTCGCCCCGA AGGGGTCGAT GGGTGACTCA
 8201  CTGACCGGGC TGATTTTGTT CGCCGCTTTG GCCTCCTCGA GGATGTAGCG
 8251  ATAATCCACC TCAGCACCAT CGGTAACGGT CAGGACGCCC ATTTCCACCC
 8301  ATTTCTGAAA GCGTTCGGCT GTCCGTCTAT CTTCATTTTT CTCGACGCTG
 8351  TACACCGTGT CATACGGTAC CCAGAAGCGC GGGGCCACAC TGTAGTAATG
 8401  CGTTTTACCG TCAATCTCGC GGGTATAAAG TCGCGCCATG CTGTTCATAT
 8451  CCAGCTTACG CGCCAGGTCA AAGGCCAGAA TGCACGGCTG CCCCTCGAAC
 8501  TGCTCAAGGG TCAGTGATTT ATCCTCGCAG CTCTGCCAGC TCACCAGGTT
 8551  GAAATACGCC GAACGCGCCG ACACCCAGAT ATTGAGGTGT TTTGTTTTAA
 8601  AGACGTTTGC CAGACGGGCG TTATTTTTCG CACGCTGCTG CTGACTTAAC
 8651  AAAAATTCGC GATAAACCGA CACGCCAATA TTTGGATTGG CTTTTTCCAG
 8701  CACCTGCGGG TCGGTCCAGT CGTCACCTTC ATCAACGGTA TAGATGATCC
 8751  CGAACAGTTC ATCGTTAGGC ACCGAGCCGT TGAGCATCTC GATGACTTCC
 8801  CGCCGCTTGT CGTAGCACGG CCCCTCAATG TTGTACCCGG CGGTGGTGAT
 8851  GGCCCACATC AGTGGCTGAC GTCGCGCCCC CATCCCGGTA AGCATTGTGG
 8901  TATAAAGCGC ATCGGTGGCA TGCTCGTGAT ATTCATCAAC CACGGCACAG
 8951  TGGGGTGATG AACCATCACC GGGGTTGCCG ATCAGCGGTT CAAACCGCGC
 9001  GCCATCCTCC GGACGGTTCA TGTTTGAGGC GTTAACCTCA ATCCCGAACG
 9051  CTTCCGTCAG CATGGGTGTG CGTTTACACA TCAGTCGCGC CGGGCGAAAG
 9101  ACTTCCCACG CCTGTTTCTC TGTCGTGGCA CCGGAATACA CTTCCGCGCC
 9151  GAACTCGTTA TCACAGGCAA AACAATACAG GGCAACACCG GCAGAGATTG
 9201  CCGATTTGCC GTTCTTACGG GGGATTTCGG TATACACCTC CCTGAAGCGG
 9251  CGCAGCCGGG AGCCTTTATT GACCCAGCCA AACGCACAGC AGATCACAAA
 9301  GAGCTGCCAC GGTTCCAGCG TGATGGGCAT CCTCTTGAAT GCCCACTCCC
 9351  CCTTGGTGTG TGGCAACAGC TGAATAAATT TCGCGACCCG TTCAGCCAGG  PvuII9369
 9401  TCCTTGTCGA AGCGGTAACG AAACGACTTA CTTTTTCCG CCATCAGGTC
 9451  ATCAAGATGG CGCTGGCAGG CCTGAATCAC AAACTGGCAG GCCACAATCT
 9501  TTCCGCGCAC GACATCACGG GCATACTGAT TGGCAGCATT TACGTTGGGG
 9551  TAAGATTTCC GGCTCATGAT TCGATGATTT TCAGAAACGG GTTAGTGGCT
 9601  TTCTTCTTCC CCGCCAGGCC AATCAGACGC TGGCGGCTGC TGGGGTCGAG
 9651  TCCGAGCATT GCCCCCGTAC TGCTCATCTC GGACTCCTGT TCTTTTTTGG
 9701  CGGTCAGCTC CGGATTTTTG ACCATACCGC CCATTGCACC GGTGATGGTG
 9751  TTGCCCTGTC TGGCAATATT TTTCACGGCA CGCCGCCAGA ACTCGTAGGC
 9801  CACGCACCAC CGCTCAAGCA CCGCGAGGTC AGTCACGCAC AGCAGGCCCT
 9851  GACCGCAGAG TTCTTTAGTT GTCAGTTGCC ACATGATCGT AGCGAGAGGG
 9901  AGATCTTCTT CAGCGAACCA CTCCGGTGGC TCAACACCTT TGATGGGCGT
 9951  AAAAACAGGT TCATCTTTAT TCAGGGCTCG CTTGCCGGGG TTTCCGGCCA
10001  GCGCCTTGCG CGCCGTTGGC TTGGGGCGAC GCCCGGAACG CCCCGCCGTT
10051  CCAGCCATAT GCGGCACTCC TGGTTAAATT TCATTTTTCG CGGGTATAAA
10101  AAACGATAA
```

nt. 153, and *Sac*I site at nt. 389. The pNV755 *Pst*I fragment had the *Pvu*II site at nt. 142 and *Sac*I site at nt. 378. And the pNV754 containing the *Pvu*II fragment had the *Sac*I site at position nt. 264. In the 10.1 fragment A contiguous sequence, these sites had corresponding positions at *Hin*dIII 5191, *Pst*I 5220, *Pvu*II 5355, and *Sac*I 5591. These similar enzyme site location and orientation ensured the identity of the DNA segments which were cloned and sequenced to attain the contiguous 10.1 *Bam*HI fragment A portion.

3.3. Discussion

Fragment A of the SfV genomic DNA is approximately 13 kb long and is situated between fragments D and F according to the *Bam*HI map of the complete genome (Figure 3.1B). *Eco*RI and *Bam*HI were used to generate distinct DNA digest fingerprints which were used to construct a physical map of the SfV genome (Figure. 3.1A) [Huan, 1997C #230][Allison *et al.*, submitted for publication]. This map served as a guide in establishing the identity of our phage stock and locating fragment A during the cloning process. Identical digest patterns were visualised on a 0.6% agarose gel electrophoresis run of our purified Proteinase K-treated SfV DNA (Figure. 3.1A).

A restriction map of the fragment A section of the phage genome was produced (Figure 3.9). Due to the moderate size of the portion to be sequenced, a more direct sequencing approach was utilised. This involved fragment cloning and subcloning to produce

templates with inserts of 1-2 kb length, the optimum insert size range for sequencing template. Universal forward and reverse primers were initially used to minimise the need for customised primers. Created templates were characterised for insert inclusion prior to PCR amplification and automated sequencing (ABI Prism) which employs the dideoxy method developed by Sanger *et al.*, [Sanger, 1977 #233]. The vector sequence was subsequently removed during the contig assembly process.

In mapping the 10.1 kb fragment A portion, the cloning experiments provided a valuable exercise in the analysis of enzyme restriction products. For example, we have deduced that *Pvu*II generated two 2.3 kb pNV728 fragments which had similar electrophoretic mobility, an observation confirmed by the appearance of a more intense fluorescence at the 2.3 kb band position and by summation of the various sizes of the *Pvu*II digest fragments relative to the total size (Figure. 3.3A). More useful were the side by side electrophoretic patterns produced by single and double enzyme digests on pNV728 (Figure. 3.3B and C). The highest molecular weight fragment which were likely vector carriers, showed clear reduction in size when restricted with an additional enzyme. The *Eco*RV cut pNV728, for example, had an additional 0.7 kb run off fragment deduced through size calculation after an observed size reduction of the higher molecular weight vector band when simultaneously digested with *Sac*I (Figure. 3.3B and C). Another 0.6 kb fragment appeared with the *Hin*dIII-*Sac*I digest of pNV728 which was not evident when digested with *Hin*dIII alone (Figure. 3.3B and C). In both instances, it was obvious that the fragments removed in combination with *Sac*I were cleaved from the larger

fragment containing the vector and it was bounded at one end by the lone *Sac*I site of the *Bam*HI fragment A. This made the fragments ideal hybridisation probes for detecting DNA segments containing the *Sac*I site and DNAs past the *Sac*I site.

In conclusion, this study has demonstrated a strategy for sequencing the complementary strands of a moderate-sized nucleotide fragment through direct sequencing approach using DNA cloning and primer walking. The use of DNA hybridisation in sequence targetting and determining the order of overlapping sequences from clones have also been employed in our experiments. Having determined the 10.1 kb of phage SfV *Bam*HI fragment A, subsequent investigations would involve DNA and protein sequence analysis, and homology searches in order to gain a more in-depth understanding of the protein products encoded in the phage genome. This will be instrumental in designing experimental protocols for the molecular characterisation of the putative protein products.

IV RESULTS

CHAPTER 4

Sequence Analysis of *Shigella flexneri* Bacteriophage SfV *Bam*HI Fragment A which Encodes for the Structural Proteins of the Late Region

4.1. Introduction

Temperate bacteriophage SfV is one of several *Shigella flexneri* bacteriophages involved in mediating host serotype-conversion through O-antigenic variation of the LPS [Petrovskaya, 1982 #48]. *S. flexneri* serotype Y has the basic O-antigen structure consisting of repeating units of tetrasaccharides which comprise the LPS together with the lipid A and the core polysaccharide [Brahmbhatt, 1992 #51]. Upon lysogeny, bacteriophage SfV confers type V O-antigen modification by attaching a glucosyl group to the rhamnose II sugar of the tetrasaccharide repeat unit through an α1,3 linkage [Simmons, 1987 #2]. The O-antigen modification genes $gtrA_{(V)}$, $gtrB_{(V)}$, $gtrV$, are located in the *Bam*HI fragment C of the SfV physical map immediately downstream of the site-specific integration locus composed of the *attP* site, *int* and *xis* genes [Huan, 1997A #118][Huan, 1997B #117](Figure 4.1).

Recently, 15 kb of sequence upstream of the *xis* gene was determined (Allison et al., submitted for publication). This region to the left of the *xis* gene in the physical map was described to contain early region genes necessary for lysis, regulation, DNA

modification, recombination and replication, immunity and regulation (Allison *et al.*, submitted for publication). Analysis of the 15 kb sequence and functional studies revealed many features that are similar to phage λ. Phage SfV was described as utilising a lambda-like repression sytem, multiple superinfection immunity systems, and a P4-like transcription termination mechanism (Allison et al., submitted for publication) (Figure 4.1). Like the coliphage [Enquist, 1984 #305], bacteriophage SfV also inserts its DNA through site-specific recombination and requires recombinase enzymes belonging to the integrase family [Argos, 1986 #306]. The genetic recombination between phage (*att*P) and bacterial genome (*att*B) occurs when short homologous common core sequences overlap, producing the prophage attachment sites *att*L and *att*R that flank the phage genome [Campbell, 1992 #307][Campbell, 1962 #308]. The phage SfV *att*P, *xis* and *int* genes located in a 2.2 kb sequence cluster adjacent to the serotype conversion genes have been sequenced and characterised [Huan, 1997A #118]. These three genes have a conserved order which is not unique to phage SfV. *Salmonella* phage P22, *S. flexneri* phage SfX and SfII also exhibited homologous cluster arrangement wherein the *att*P site is immediately adjacent to the serotype conversion region [Clark, 1991 #119][Leong, 1985 #198][Poteete, 1988 #310][Leong, 1986 #311][Guan, 1999 #151][Verma, 1993 #116].

This chapter was focused on the analysis of the 10.1 kb *Bam*HI fragment A sequence adjacent to the left side of the *pac* site. Based on the extensive homologies observed between phage λ and phage SfV genome, database homology search and comparative analysis between fragment A open reading frames and similar regions from other dsDNA

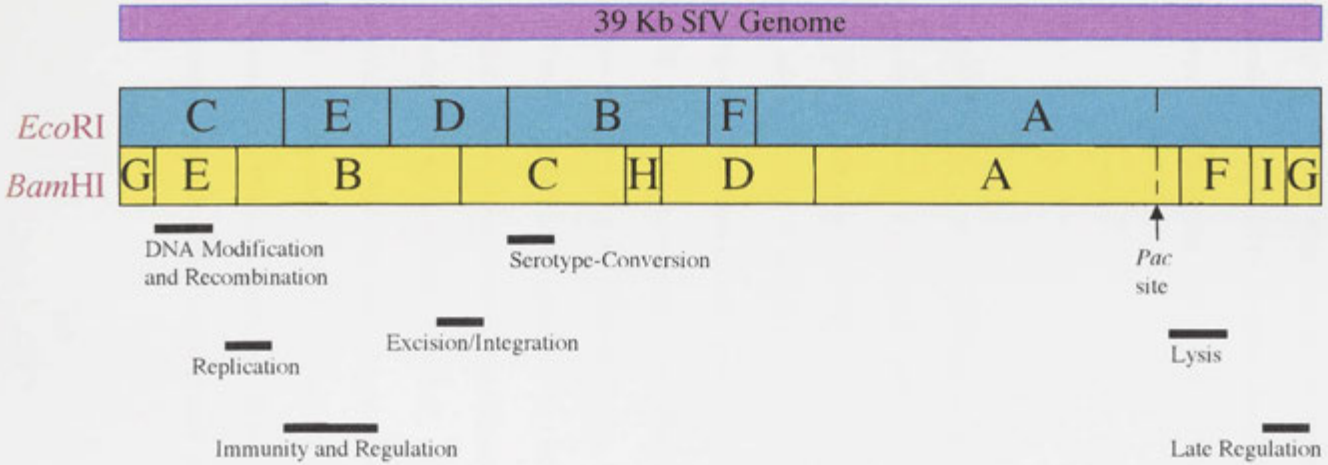


Figure 4.1. Functional Map of Bacteriophage SfV showing restriction fragments represented by the letters of the alphabet. Line bars indicate the location of identified functional regions of the SfV genome. Functional designations are written below the line bars and the *pac* site is indicated by an arrow. (Modified and adapted from Huan *et al.*, 1997and Allison *et al.*, submitted for publication).

bacteriophages were also conducted to establish likely gene product functions and to gain an understanding of the evolutionary link between various groups of bacteriophages. This region contained genes and products which shares homology to essential major and minor structural proteins of bacteriophages, primarily of the family Siphoviridae.

4.2. Results

4.2.1. *Analysis for the presence of ORFs in the 10.1 kb Sequence*

The contiguous sequence of 10.1 kb portion of the *Bam*HI fragment A had an overall G+C content of 53.9%. Analysis of the 10.1 kb fragment using the National Center for Biotechnology Information ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and the Frames program available through the Australian National Genomic Information Service (ANGIS) (http://www.angis.org.au), identified 1 incomplete ORF and 13 complete putative ORFs based on the following criteria: 1) A minimum of 50 codons per ORF; 2) Presence of start codons AUG (ATG), UUG (TTG), or GUG (GTG); and 3) Start codons preceded by a potential Shine-Dalgarno sequence complementary to the 3' end of the 16 S rRNA of *S. flexneri* (Table 4.1). The core consensus sequence of the ribosome binding sites, AGGAG, was conserved in most cases for the 14 putative ORFs identified. The stop codons UAA (TAA), UAG (TAG), and UGA (TGA) were found in 4, 1, and 9 of the ORFs, respectively. All putative ORFs were coded on the same coding strand and transcribed in the same leftward direction (Figure. 4.2 and 4.3).

Table 4.1. Genetic features of bacteriophage SfV *Bam* HI fragment A ORFs.

ORF*	START	STOP	INTERGENIC DISTANCE	ISOELECTRIC POINT	LENGTH (nt)	MOLECULAR WEIGHT (kDa)	PUTATIVE RBS AND START CODON
164	10058 (ATG)	9564 (TGA)	0	10.36	495	17.9	gaaatttaaccAGGAGTGccgcatATG
577	9567 (ATG)	7834 (TGA)	-3	5.33	1734	65.3	GGGGAAGAAGAAAGAAAGccactaacccgtttctGAAAAtcatcGAAtcATG
367	7563 (ATG)	6460 (TGA)	271	9.45	1104	41.9	GtcAGtcctGAGActGCGATG
200	6467 (ATG)	5865 (TAA)	-8	4.92	603	22.7	GAGGGAAAAAccAATG
409	5854 (ATG)	4625 (TGA)	11	5.08	1230	45.8	ttAAcGcGtAAAGGAAAcatcATG
107	4546 (ATG)	4223 (TGA)	34	4.23	324	12.4	ttctGGcGGGcAcAGGAGGttttATG
104	4130 (ATG)	3816 (TGA)	93	7.26	315	11.5	tgatccgcctGcGGGtGGAtATG
168	3841 (ATG)	3335 (TGA)	-25	12.56	507	20.0	GtAGGcAGAGtcacGGAGGcaacaATG
186	3338 (ATG)	2778 (TGA)	-3	4.20	561	21.0	cactGAAGccGGAGcGAcGcAGATG
56	2769 (ATG)	2599 (TAA)	9	10.11	171	06.4	ttccttcctGAGAAtcttcATG
498	2615 (ATG)	1119 (TAA)	-16	5.54	1497	54.8	cgcgttaatAAAAAGGtGAAcAccGATG
118	1119 (ATG)	763 (TAA)	0	4.66	357	12.9	ttcagtattcAGAGGAGtctgcataATG
89	763 (ATG)	494 (TGA)	0	5.15	270	10.0	tccAcGGtGAAGAAGGGGAttaccagtaATG
116	352 (ATG)	incomplete	142	9.87	351	12.5	AAAcAGGAtAGAGcAGGAGAAAcgcacacATG

* ORF designation is based on the number of amino acids

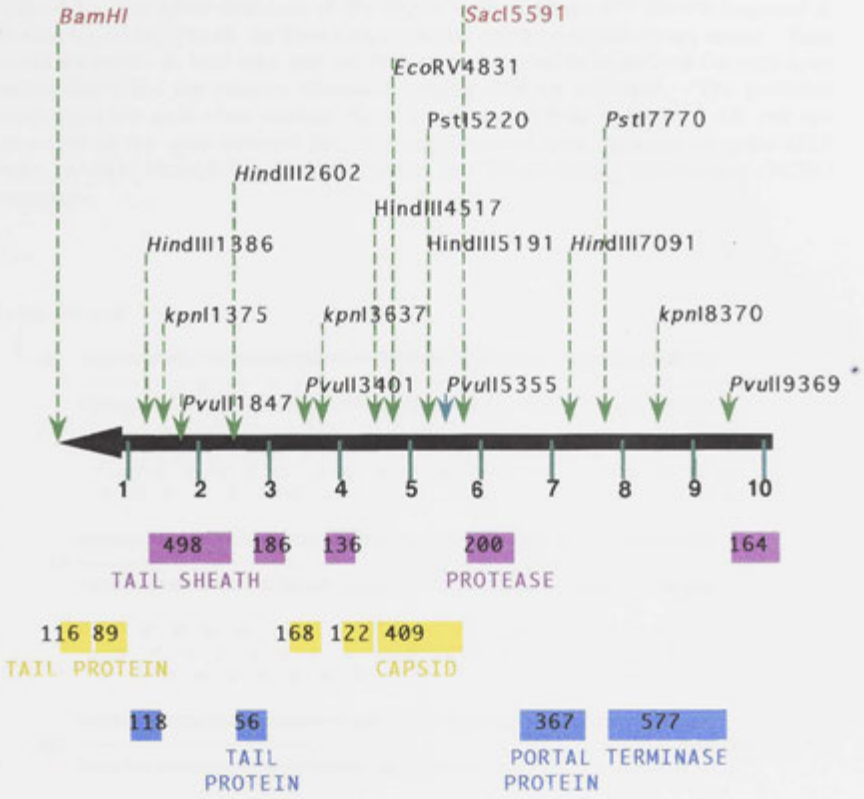
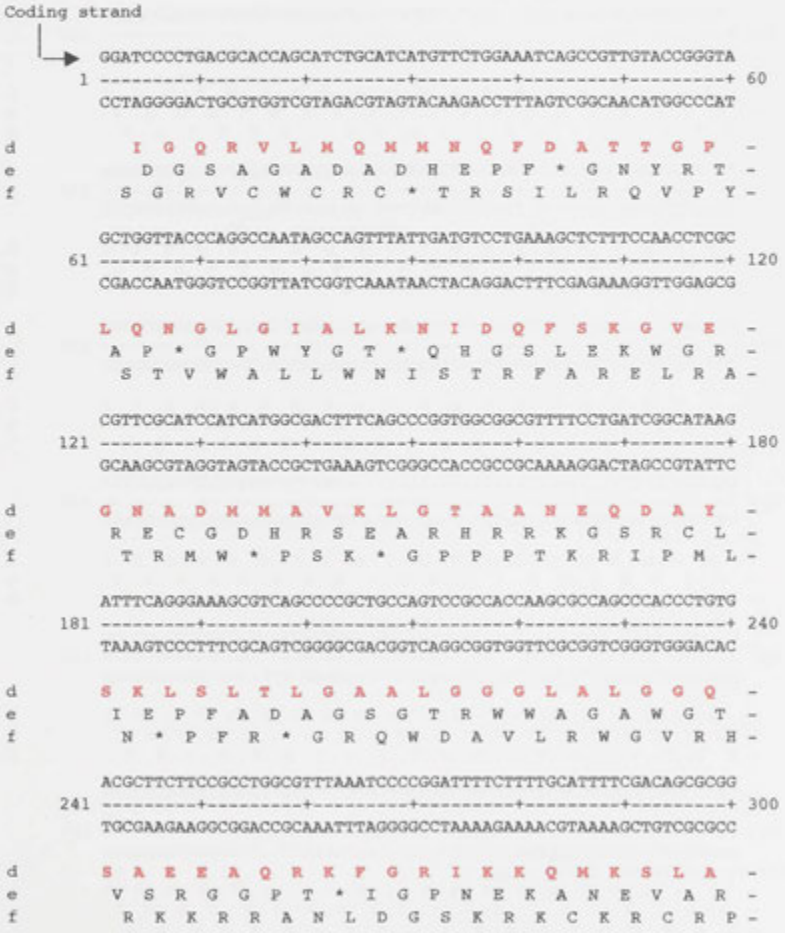
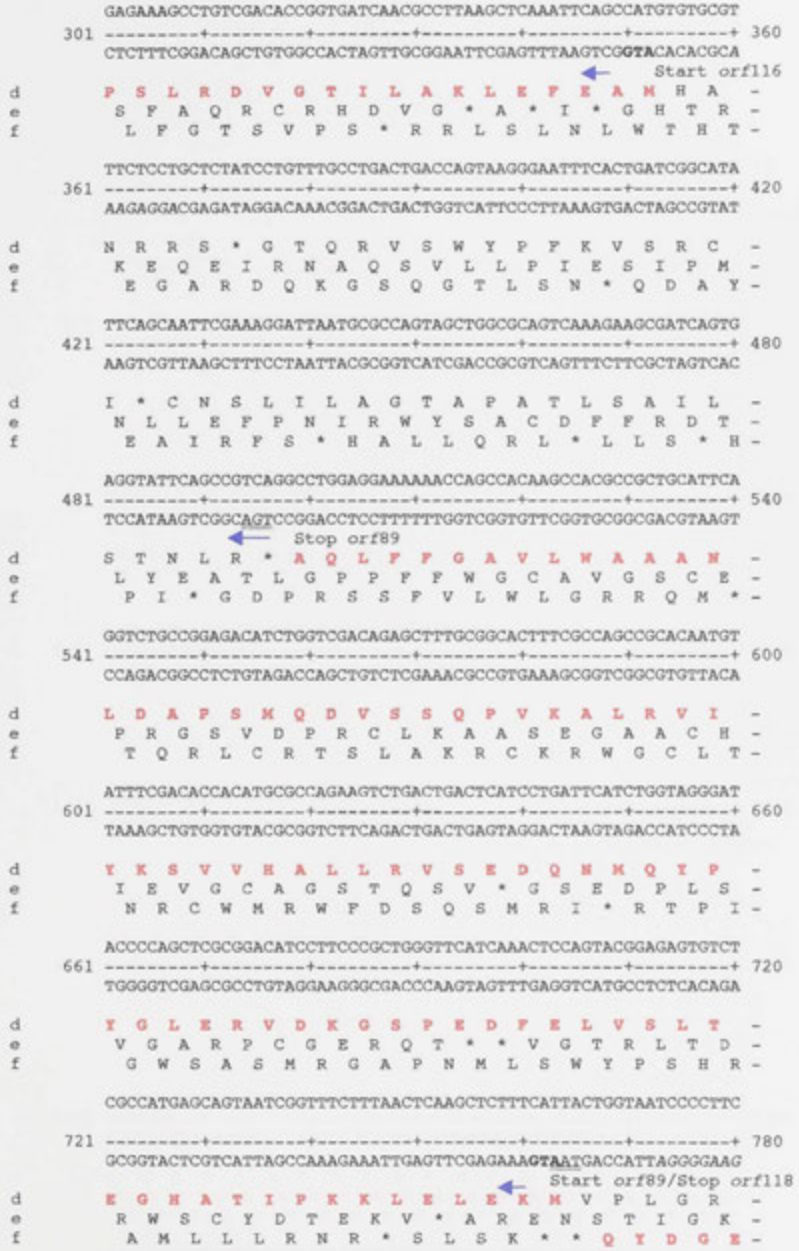
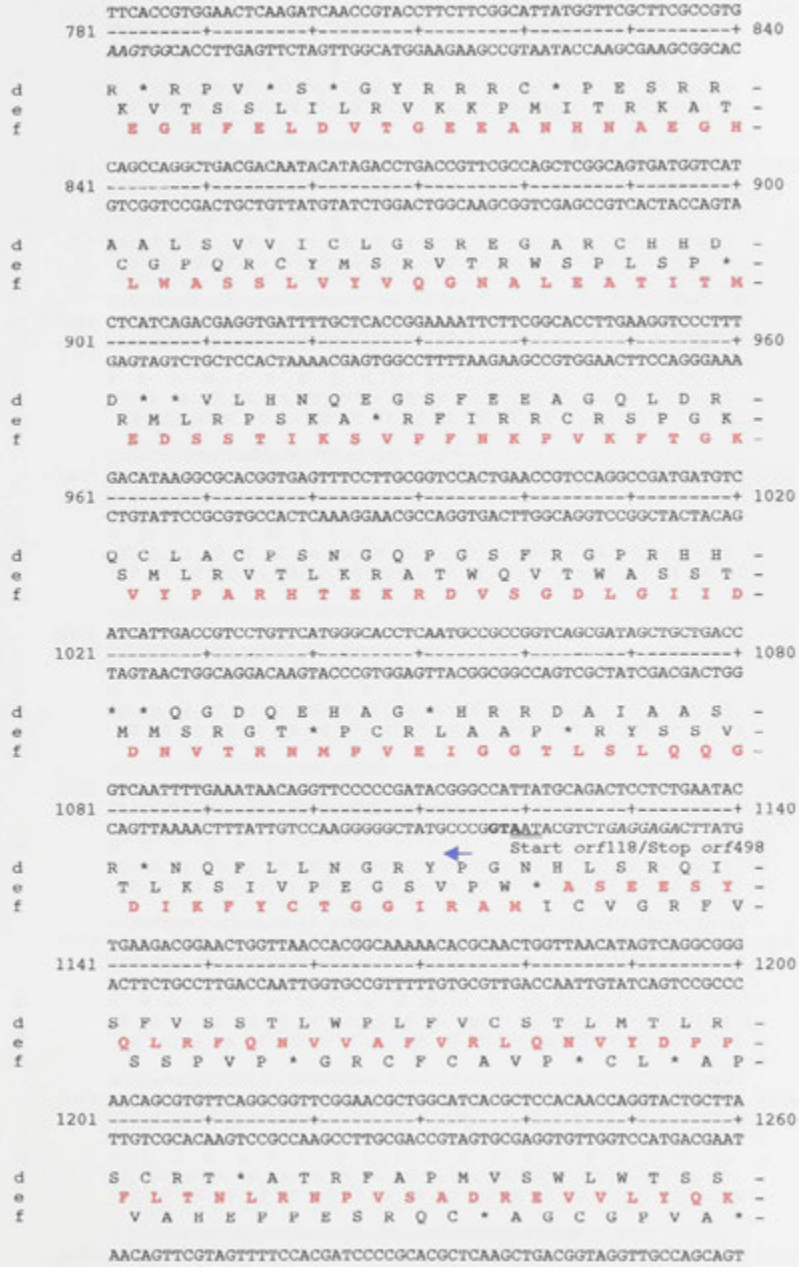


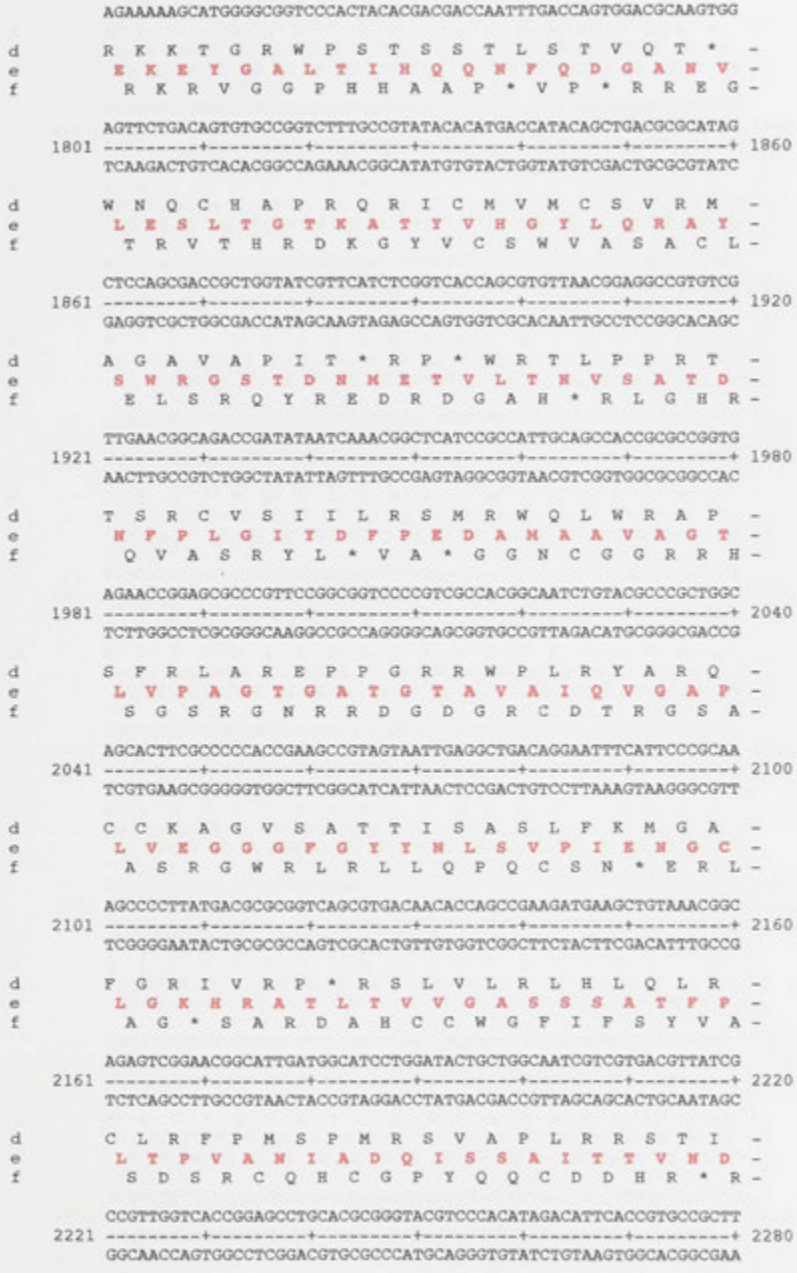
Figure 4.2. The 10.1 kb phage SfV *Bam*HI fragment A is represented by the solid line with a terminal arrowhead specifying the 5'-3' transcriptional direction of the *orfs*. Relevant restriction sites with the nucleotide position are indicated by dashed arrowlines above the mapline. The incomplete *orf116* and thirteen putative *orfs* are shown below the mapline as coloured blocks plotted according to their approximate length and relative location: Pink blocks represent *orfs* in reading frame D, yellow blocks represent *orfs* in reading frame F and blue represent *orfs* in reading frame E. The orfs are designated in Arabic numbers based on the number of their amino acids and predicted functions based on database homologies are shown below the block.

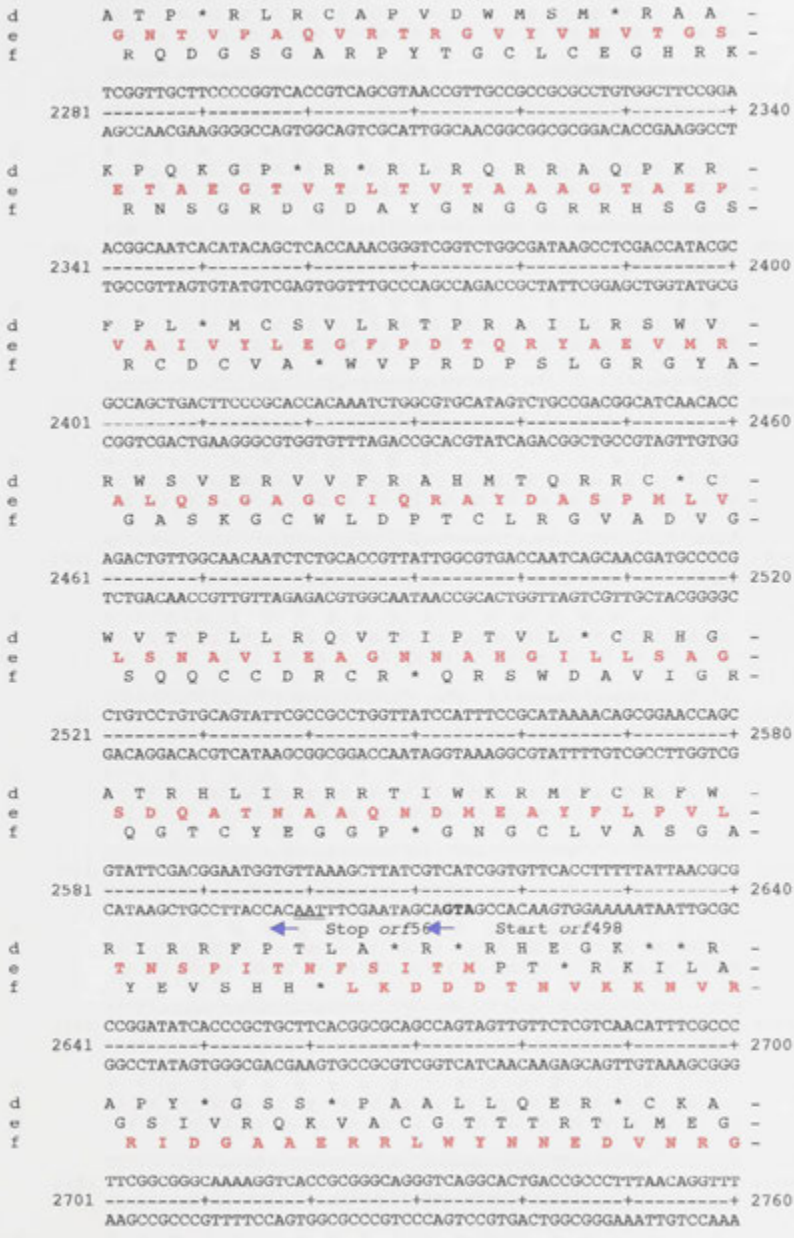
Figure 4.3. The linear sequence of the 10109 bases of phage SfV *Bam*HI fragment A showing the coding strand, its three reading frames and the complementary strand. Start codons are shown in bold print and the stop codons are double underlined for each open reading frame and the putative ribosomal binding sites are italicised. The predicted amino acids for each open reading frame are shown in bold red text. All *orfs* are transcribed in the same leftward direction and these *orfs* were detected using the ORF finder program through the National Center for Biotechnology Information (NCBI) homepage.

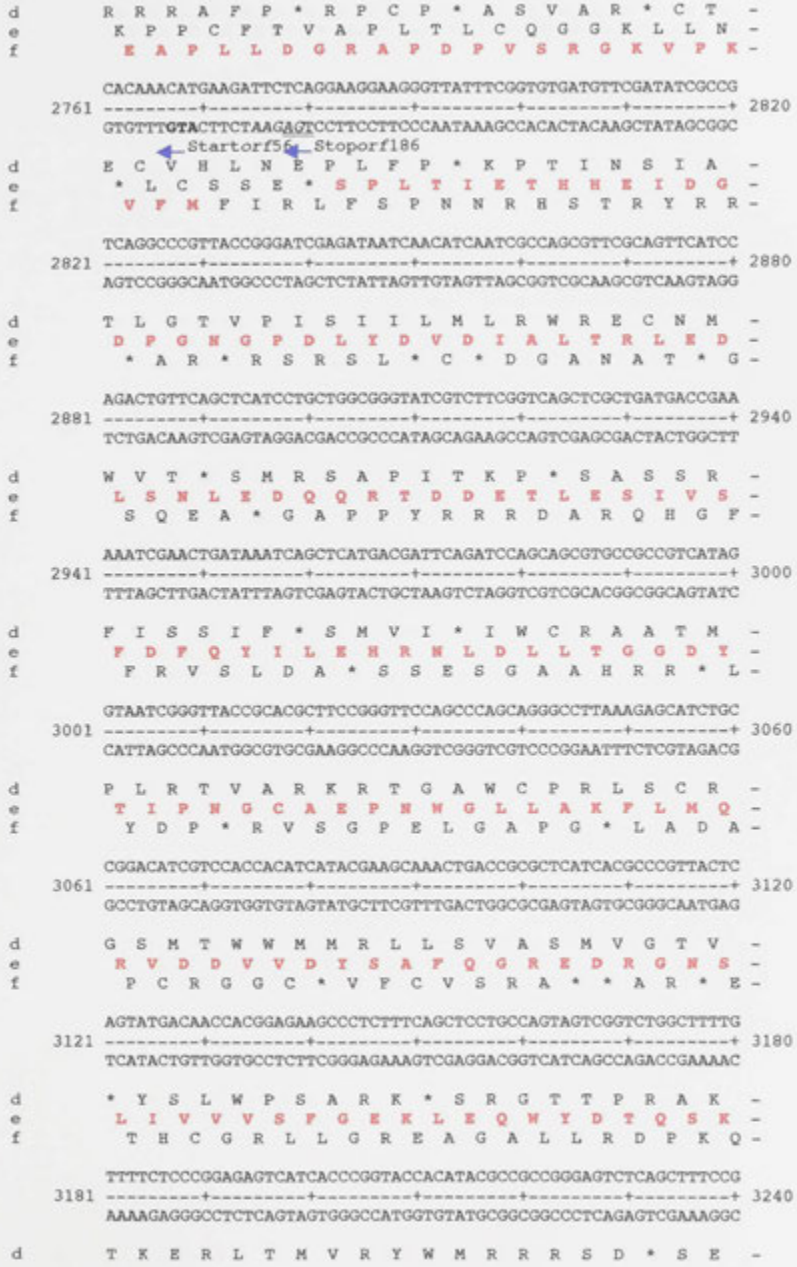


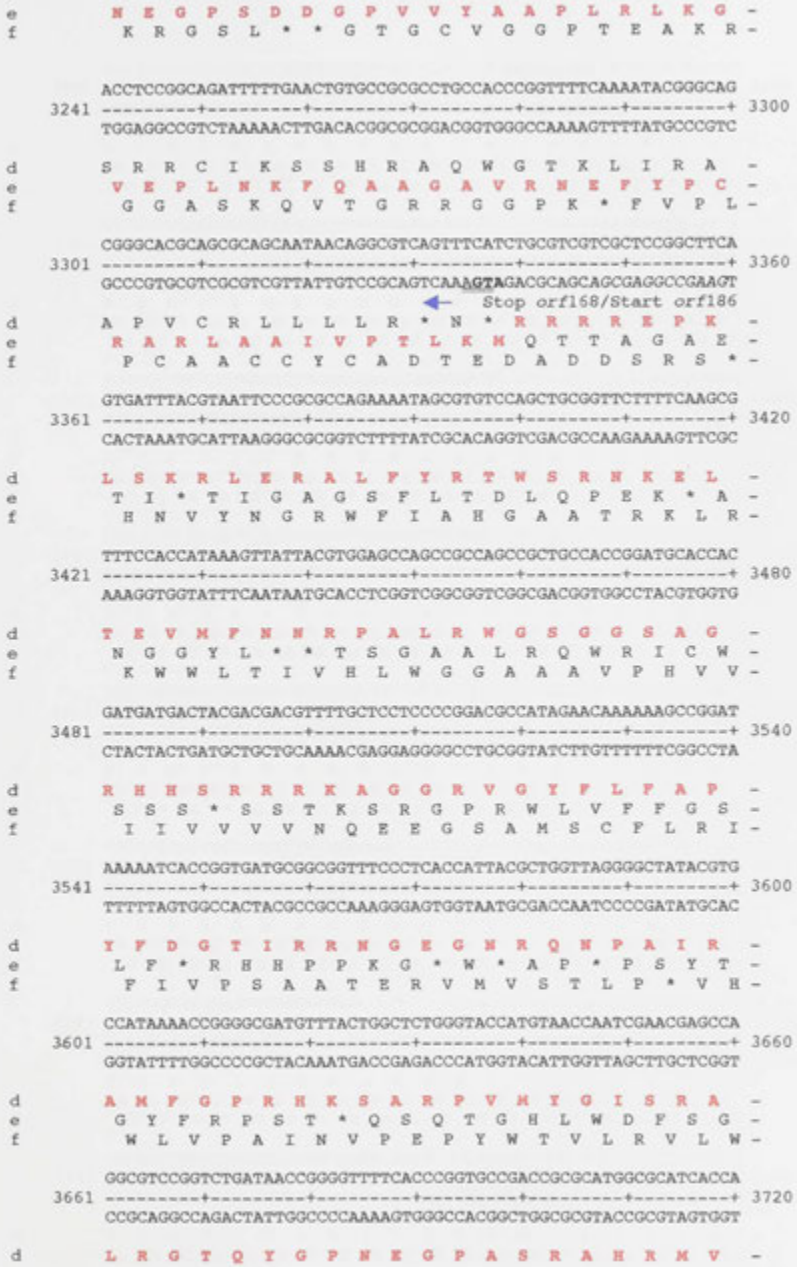

Start orf116
Stop orf89
Start orf89/Stop orf118

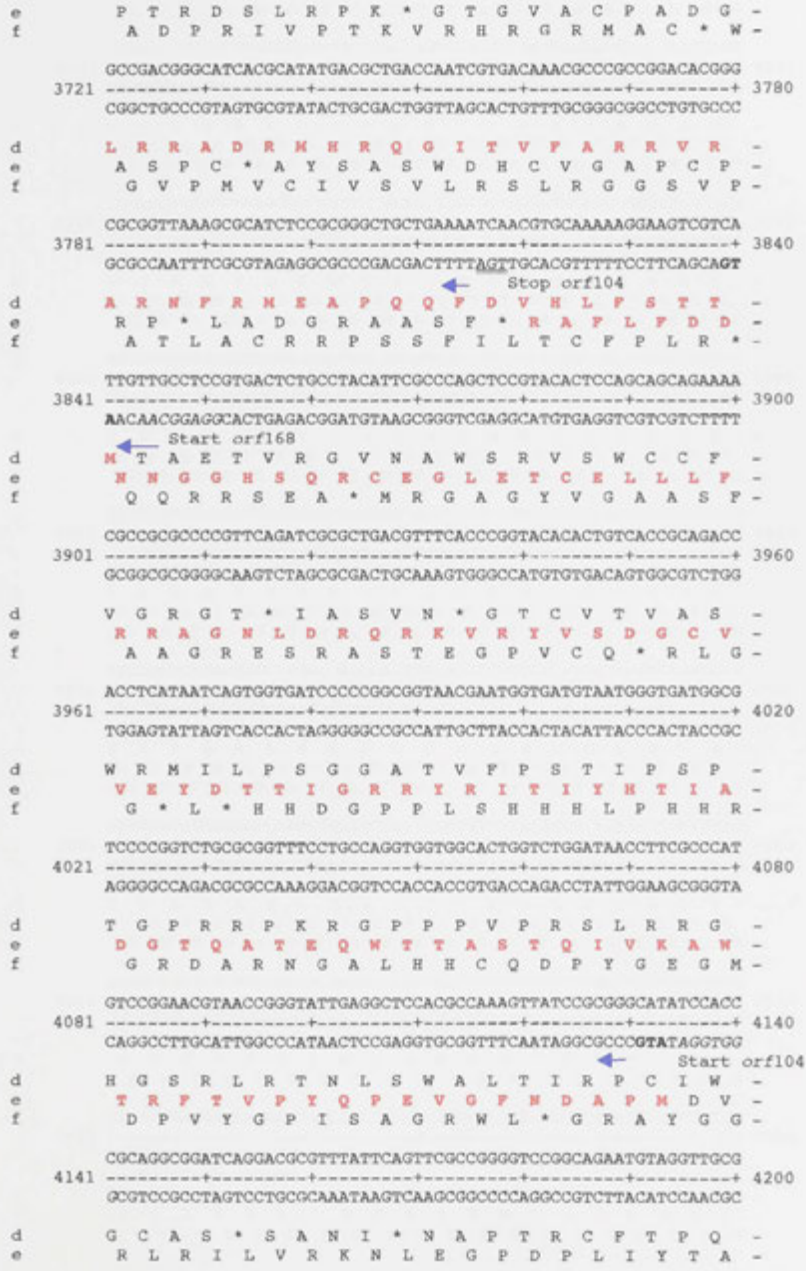

Start orf118/Stop orf498

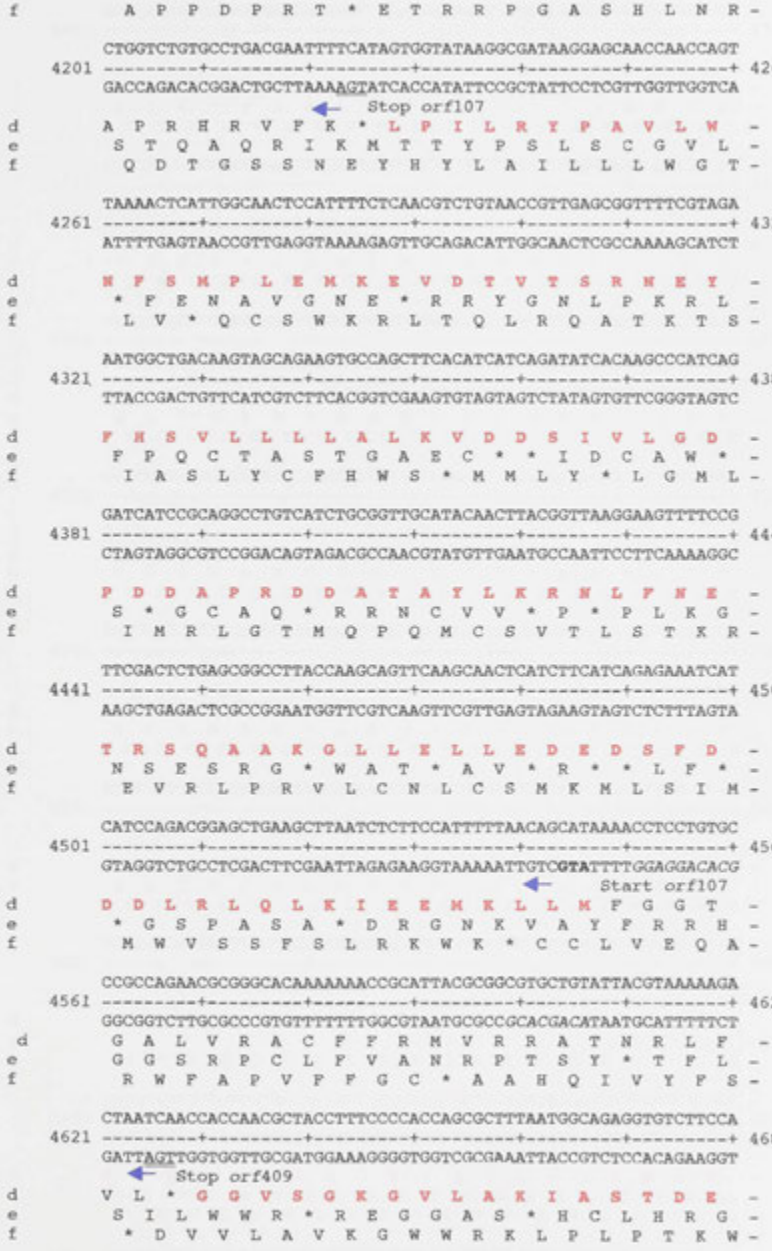

AGAAAAAGCATGGGGCGGTCCCACTACACGACGACCAATTTGACCAGTGGACGCAAGTGG
d R K K T G R W P S T S S T L S T V Q T * -
e E K E I G A L T I H Q Q N F Q D G A N V -
f R K R V G G P H H A A P * V P * R R E G -
AGTTCTGACAGTGTGCCGGTCTTTGCCGTATACACATGACCATACAGCTGACGCGCATAG
1801 ---------+---------+---------+---------+---------+---------+ 1860
TCAAGACTGTCACACGGCCAGAAACGGCATATGTGTACTGGTATGTCGACTGCGCGTATC
d W N Q C H A P R Q R I C M V M C S V R M -
e L E S L T G T K A T Y V H G Y L Q R A Y -
f T R V T H R D K G Y V C S W V A S A C L -
CTCCAGCGACCGCTGGTATCGTTCATCTCGGTCACCAGCGTGTTAACGGAGGCCGTGTCG
1861 ---------+---------+---------+---------+---------+---------+ 1920
GAGGTCGCTGGCGACCATAGCAAGTAGAGCCAGTGGTCGCACAATTGCCTCCGGCACAGC
d A G A V A P I T * R P * W R T L P P R T -
e S W R G S T D N M E T V L T N V S A T D -
f E L S R Q Y R E D R D G A H * R L G H R -
TTGAACGGCAGACCGATATAATCAAACGGCTCATCCGCCATTGCAGCCACCGCGCCGGTG
1921 ---------+---------+---------+---------+---------+---------+ 1980
AACTTGCCGTCTGGCTATATTAGTTTGCCGAGTAGGCGGTAACGTCGGTGGCGCGGCCAC
d T S R C V S I I L R S M R W Q L W R A P -
e N F P L G I Y D F P E D A M A A V A G T -
f Q V A S R Y L * V A * G G N C G G R R H -
AGAACCGGAGCGCCCGTTCCGGCGGTCCCCGTCGCCACGGCAATCTGTACGCCCGCTGGC
1981 ---------+---------+---------+---------+---------+---------+ 2040
TCTTGGCCTCGCGGGCAAGGCCGCCAGGGGCAGCGGTGCCGTTAGACATGCGGGCGACCG
d S F R L A R E P P G R R W P L R Y A R Q -
e L V P A G T G A T G T A V A I Q V G A P -
f S G S R G N R R D G D G R C D T R G S A -
AGCACTTCGCCCCCACCGAAGCCGTAGTAATTGAGGCTGACAGGAATTTCATTCCCGCAA
2041 ---------+---------+---------+---------+---------+---------+ 2100
TCGTGAAGCGGGGGTGGCTTCGGCATCATTAACTCCGACTGTCCTTAAAGTAAGGGCGTT
d C C K A G V S A T T I S A S L F K M G A -
e L V E G G G F G Y Y N L S V P I E N G C -
f A S R G W R L R L L Q P Q C S N * E R L -
AGCCCCTTATGACGCGCGGTCAGCGTGACAACACCAGCCGAAGATGAAGCTGTAAACGGC
2101 ---------+---------+---------+---------+---------+---------+ 2160
TCGGGGAATACTGCGCGCCAGTCGCACTGTTGTGGTCGGCTTCTACTTCGACATTTGCCG
d F G R I V R P * R S L V L R L H L Q L R -
e L G K H R A T L T V V G A S S S A T F P -
f A G * S A R D A H C C W G F I F S Y V A -
AGAGTCGGAACGGCATTGATGGCATCCTGGATACTGCTGGCAATCGTCGTGACGTTATCG
2161 ---------+---------+---------+---------+---------+---------+ 2220
TCTCAGCCTTGCCGTAACTACCGTAGGACCTATGACGACCGTTAGCAGCACTGCAATAGC
d C L R F P M S P M R S V A P L R R S T I -
e L T P V A N I A D Q I S S A I T T V N D -
f S D S R C Q H C G P Y Q Q C D D H R * R -
CCGTTGGTCACCGGAGCCTGCACGCGGGTACGTCCCACATAGACATTCACCGTGCCGCTT
2221 ---------+---------+---------+---------+---------+---------+ 2280
GGCAACCAGTGGCCTCGGACGTGCGCCCATGCAGGGTGTATCTGTAAGTGGCACGGCGAA

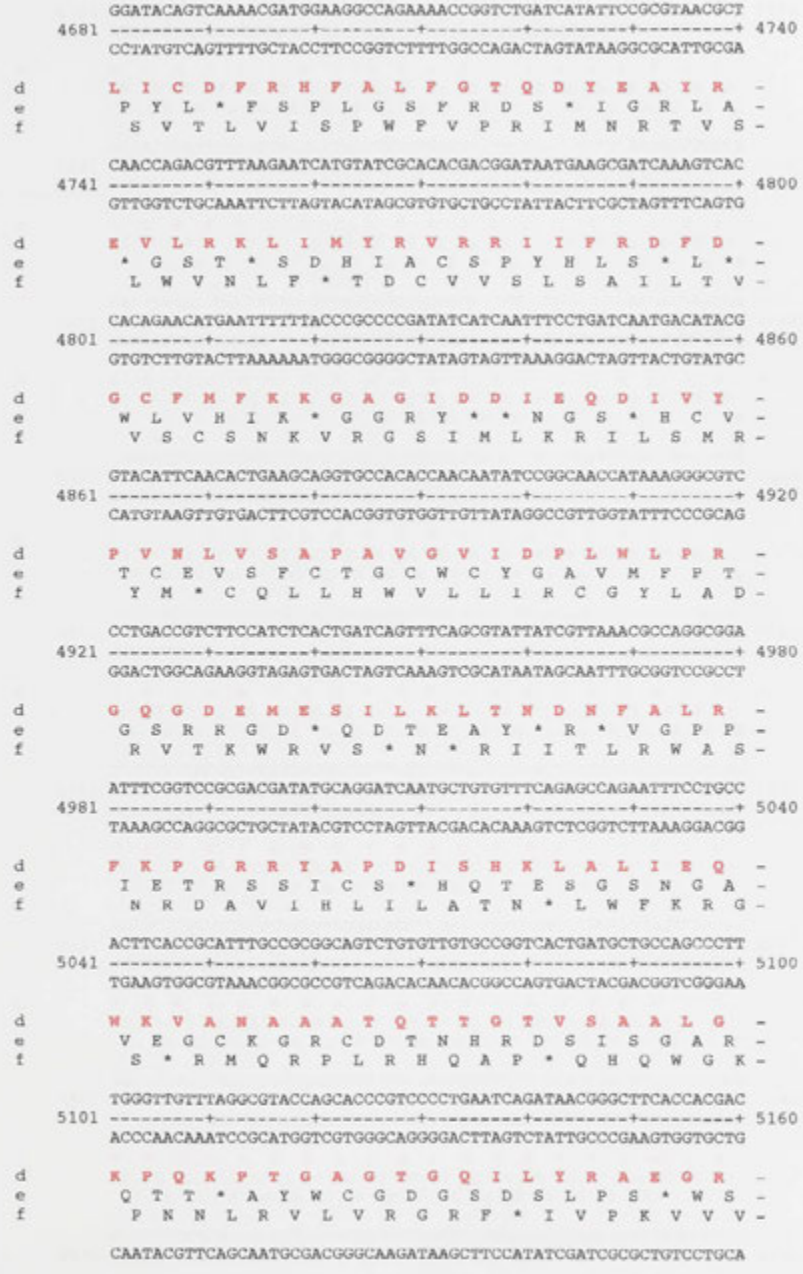

d A T P * R L R C A P V D W M S M * R A A -
e G N T V P A Q V R T R G V Y V M V T G S -
f R Q D G S G A R P Y T G C L C E G H R K -
TCGGTTGCTTCCCCGGTCACCGTCAGCGTAACCGTTGCCGCCGCGCCTGTGGCTTCCGGA
2281 + 2340
AGCCAACGAAGGGGCCAGTGGCAGTCGCATTGGCAACGGCGGCGCGGACACCGAAGGCCT
d K P Q K G P * R * R L R Q R R A Q P K R -
e E T A E G T V T L T V T A A A G T A E P -
f R N S G R D G D A Y G N G G R R H S G S -
ACGGCAATCACATACAGCTCACCAAACGGGTCGGTCTGGCGATAAGCCTCGACCATACGC
2341 + 2400
TGCCGTTAGTGTATGTCGAGTGGTTTGCCCAGCCAGACCGCTATTCGGAGCTGGTATGCG
d F P L * M C S V L R T P R A I L R S W V -
e V A I V Y L E G F P D T Q R Y A E V M R -
f R C D C V A * W V P R D P S L G R G Y A -
GCCAGCTGACTTCCCGCACCACAAATCTGGCGTGCATAGTCTGCCGACGGCATCAACACC
2401 + 2460
CGGTCGACTGAAGGGCGTGGTGTTTAGACCGCACGTATCAGACGGCTGCCGTAGTTGTGG
d R W S V E R V V F R A H M T Q R R C * C -
e A L Q S G A G C I Q R A Y D A S P M L V -
f G A S K G C W L D P T C L R G V A D V G -
AGACTGTTGGCAACAATCTCTGCACCGTTATTGGCGTGACCAATCAGCAACGATGCCCCG
2461 + 2520
TCTGACAACCGTTGTTAGAGACGTGGCAATAACCGCACTGGTTAGTCGTTGCTACGGGGC
d W V T P L L R Q V T I P T V L * C R H G -
e L S N A V I E A G N N A H G I L L S A G -
f S Q Q C C D R C R * Q R S W D A V I G R -
CTGTCCTGTGCAGTATTCGCCGCCTGGTTATCCATTTCCGCATAAAACAGCGGAACCAGC
2521 + 2580
GACAGGACACGTCATAAGCGGCGGACCAATAGGTAAAGGCGTATTTTGTCGCCTTGGTCG
d A T R H L I R R R T I W K R M F C R F W -
e S D Q A T N A A Q N D M E A Y F L P V L -
f Q G T C Y E G G P * G N G C L V A S G A -
GTATTCGACGGAATGGTGTTAAAGCTTATCGTCATCGGTGTTCACCTTTTTATTAACGCG
2581 + 2640
CATAAGCTGCCTTACCACAATTTCGAATAGCAGTAGCCACAAGTGGAAAAATAATTGCGC
Stop orf568
Start orf498
d R I R R F P T L A * R * R H E G K * * R -
e T N S P I T N F S I T M P T * R K I L A -
f Y E V S H H * L K D D D T N V K K N V R -
CCGGATATCACCCGCTGCTTCACGGCGCAGCCAGTAGTTGTTCTCGTCAACATTTCGCCC
2641 + 2700
GGCCTATAGTGGGCGACGAAGTGCCGCGTCGGTCATCAACAAGAGCAGTTGTAAAGCGGG
d A P Y * G S S * P A A L L Q E R * C K A -
e G S I V R Q K V A C G T T T R T L M E G -
f R I D G A A E R R L W Y N N E D V N R G -
TTCGGCGGGCAAAAGGTCACCGCGGGCAGGGTCAGGCACTGACCGCCCTTTAACAGGTTT
2701 + 2760
AAGCCGCCCGTTTTCCAGTGGCGCCCGTCCCAGTCCGTGACTGGCGGGAAATTGTCCAAA

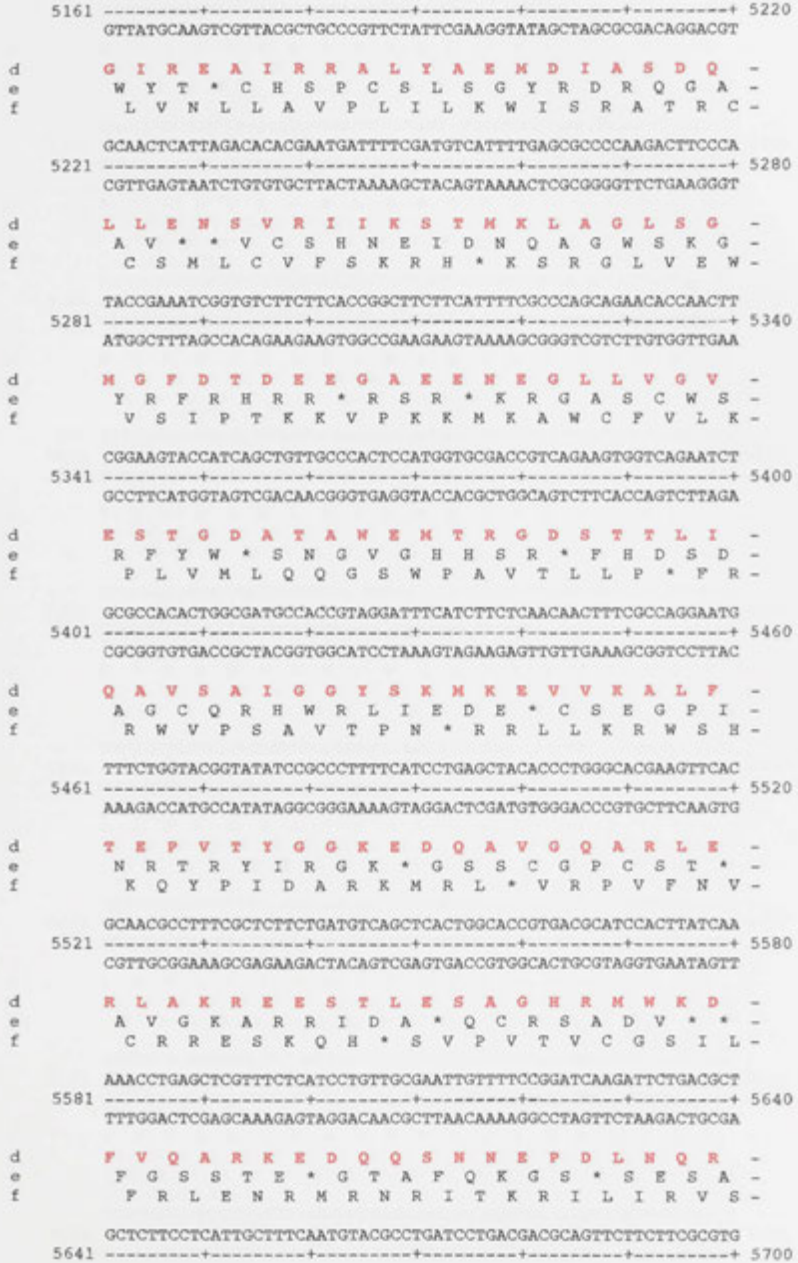

d R R R A F P * R P C P * A S V A R * C T -
e K P P C F T V A P L T L C Q G G K L L N -
f E A P L L D G R A P D P V S R G K V P K -
CACAAACATGAAGATTCTCAGGAAGGAAGGGTTATTTCGGTGTGATGTTCGATATCGCCG
2761 ---------+---------+---------+---------+---------+---------+ 2820
GTGTTTGTACTTCTAAGAGTCCTTCCTTCCCAATAAAGCCACACTACAAGCTATAGCGGC
Startorf56
Stoporf186
d E C V H L N E P L F P * K P T I N S I A -
e * L C S S E * S P L T I E T H H E I D G -
f V F M F I R L F S P N N R H S T R Y R R -
TCAGGCCCGTTACCGGGATCGAGATAATCAACATCAATCGCCAGCGTTCGCAGTTCATCC
2821 ---------+---------+---------+---------+---------+---------+ 2880
AGTCCGGGCAATGGCCCTAGCTCTATTAGTTGTAGTTAGCGGTCGCAAGCGTCAAGTAGG
d T L G T V P I S I I L M L R W R E C N M -
e D P G N G P D L Y D V D I A L T R L E D -
f * A R * R S R S L * C * D G A N A T * G -
AGACTGTTCAGCTCATCCTGCTGGCGGGTATCGTCTTCGGTCAGCTCGCTGATGACCGAA
2881 ---------+---------+---------+---------+---------+---------+ 2940
TCTGACAAGTCGAGTAGGACGACCGCCCATAGCAGAAGCCAGTCGAGCGACTACTGGCTT
d W V T * S M R S A P I T K P * S A S S R -
e L S N L E D Q Q R T D D E T L E S I V S -
f S Q E A * G A P P Y R R R D A R Q H G F -
AAATCGAACTGATAAATCAGCTCATGACGATTCAGATCCAGCAGCGTGCCGCCGTCATAG
2941 ---------+---------+---------+---------+---------+---------+ 3000
TTTAGCTTGACTATTTAGTCGAGTACTGCTAAGTCTAGGTCGTCGCACGGCGGCAGTATC
d F I S S I F * S M V I * I W C R A A T M -
e F D F Q Y I L E H R N L D L L T G G D Y -
f F R V S L D A * S S E S G A A H R R * L -
GTAATCGGGTTACCGCACGCTTCCGGGTTCCAGCCCAGCAGGGCCTTAAAGAGCATCTGC
3001 ---------+---------+---------+---------+---------+---------+ 3060
CATTAGCCCAATGGCGTGCGAAGGCCCAAGGTCGGGTCGTCCCGGAATTTCTCGTAGACG
d P L R T V A R K R T G A W C P R L S C R -
e T I P N G C A E P N W G L L A K F L M Q -
f Y D P * R V S G P E L G A P G * L A D A -
CGGACATCGTCCACCACATCATACGAAGCAAACTGACCGCGCTCATCACGCCCGTTACTC
3061 ---------+---------+---------+---------+---------+---------+ 3120
GCCTGTAGCAGGTGGTGTAGTATGCTTCGTTTGACTGGCGCGAGTAGTGCGGGCAATGAG
d G S M T W W M M R L L S V A S M V G T V -
e R V D D V V D Y S A F Q G R E D R G N S -
f P C R G G C * V F C V S R A * * A R * E -
AGTATGACAACCACGGAGAAGCCCTCTTTCAGCTCCTGCCAGTAGTCGGTCTGGCTTTTG
3121 ---------+---------+---------+---------+---------+---------+ 3180
TCATACTGTTGGTGCCTCTTCGGGAGAAAGTCGAGGACGGTCATCAGCCAGACCGAAAAC
d * Y S L W P S A R K * S R G T T P R A K -
e L I V V V S F G E K L E Q W Y D T Q S K -
f T H C G R L L G R E A G A L L R D P K Q -
TTTTCTCCCGGAGAGTCATCACCCGGTACCACATACGCCGCCGGGAGTCTCAGCTTTCCG
3181 ---------+---------+---------+---------+---------+---------+ 3240
AAAAGAGGGCCTCTCAGTAGTGGGCCATGGTGTATGCGGCGGCCCTCAGAGTCGAAAGGC
d T K E R L T M V R Y W M R R R S D * S E -

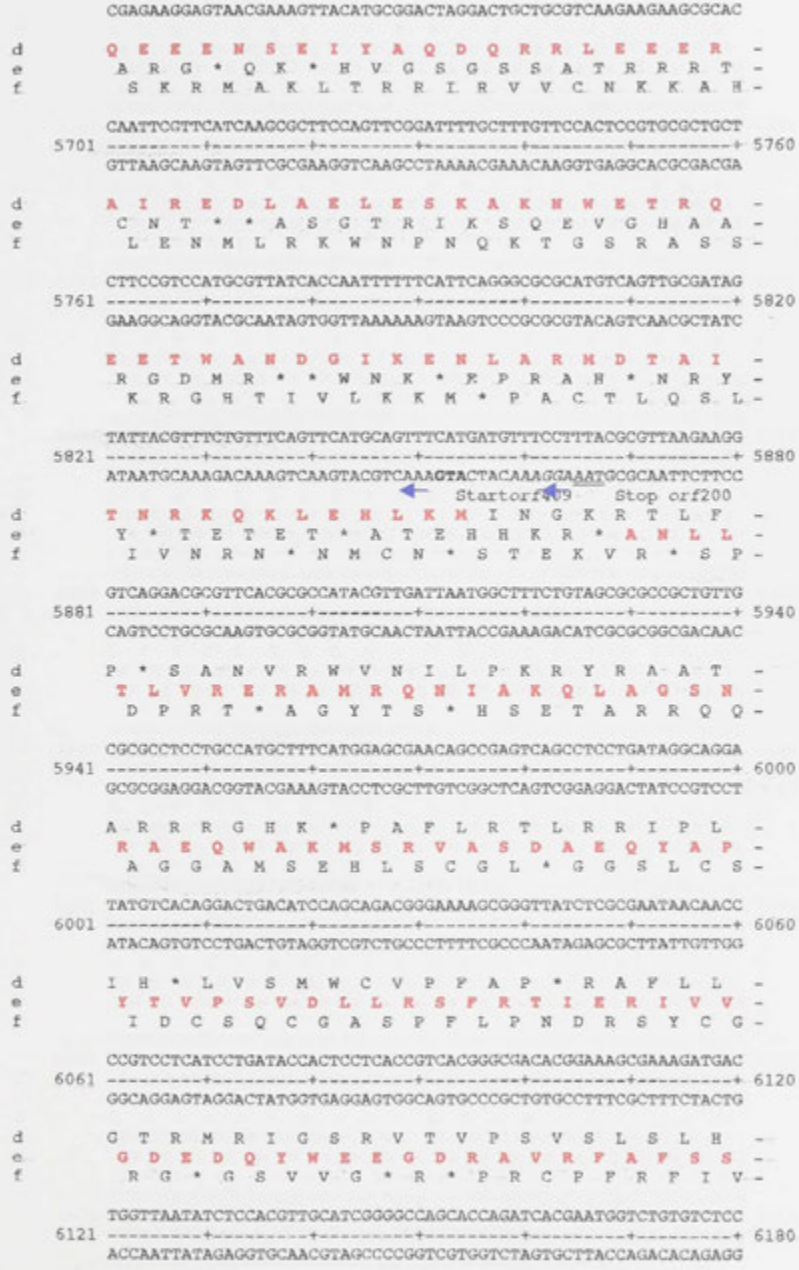

Stop orf168/Start orf186

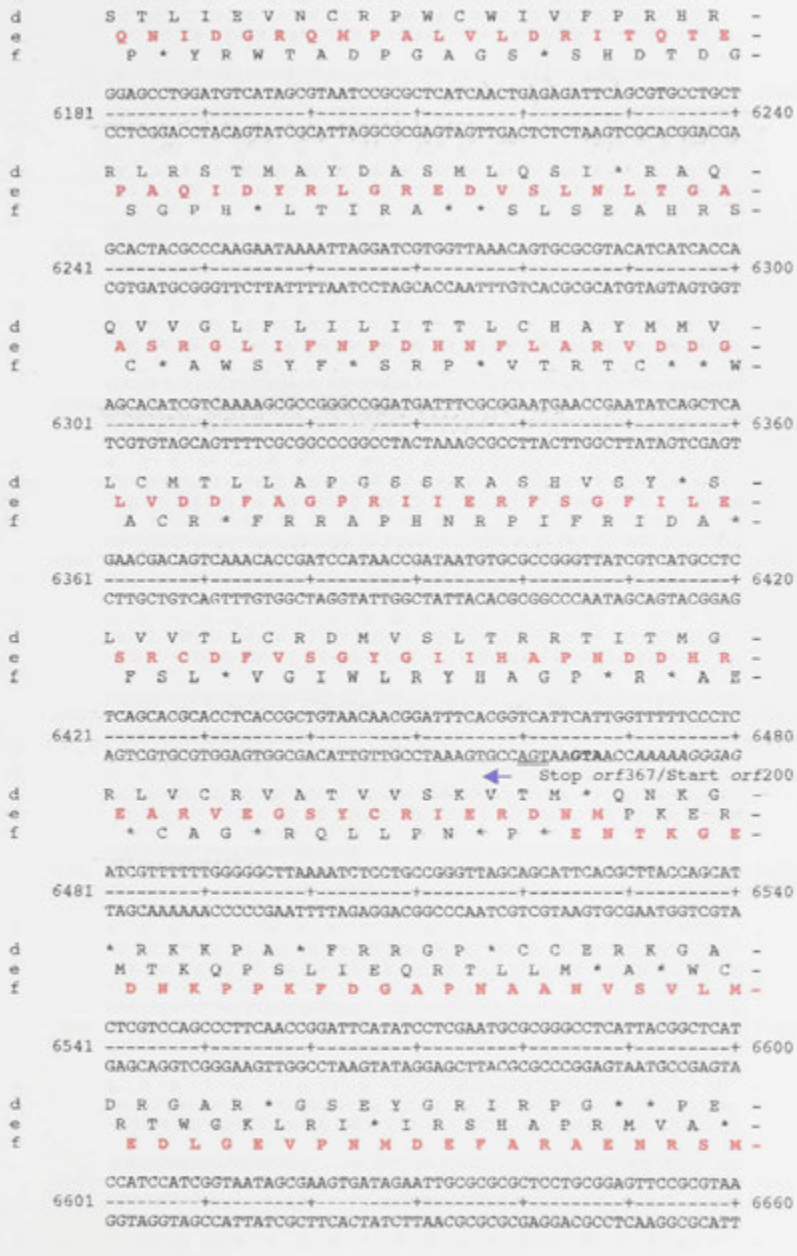

Stop orf104
Start orf168
Start orf104

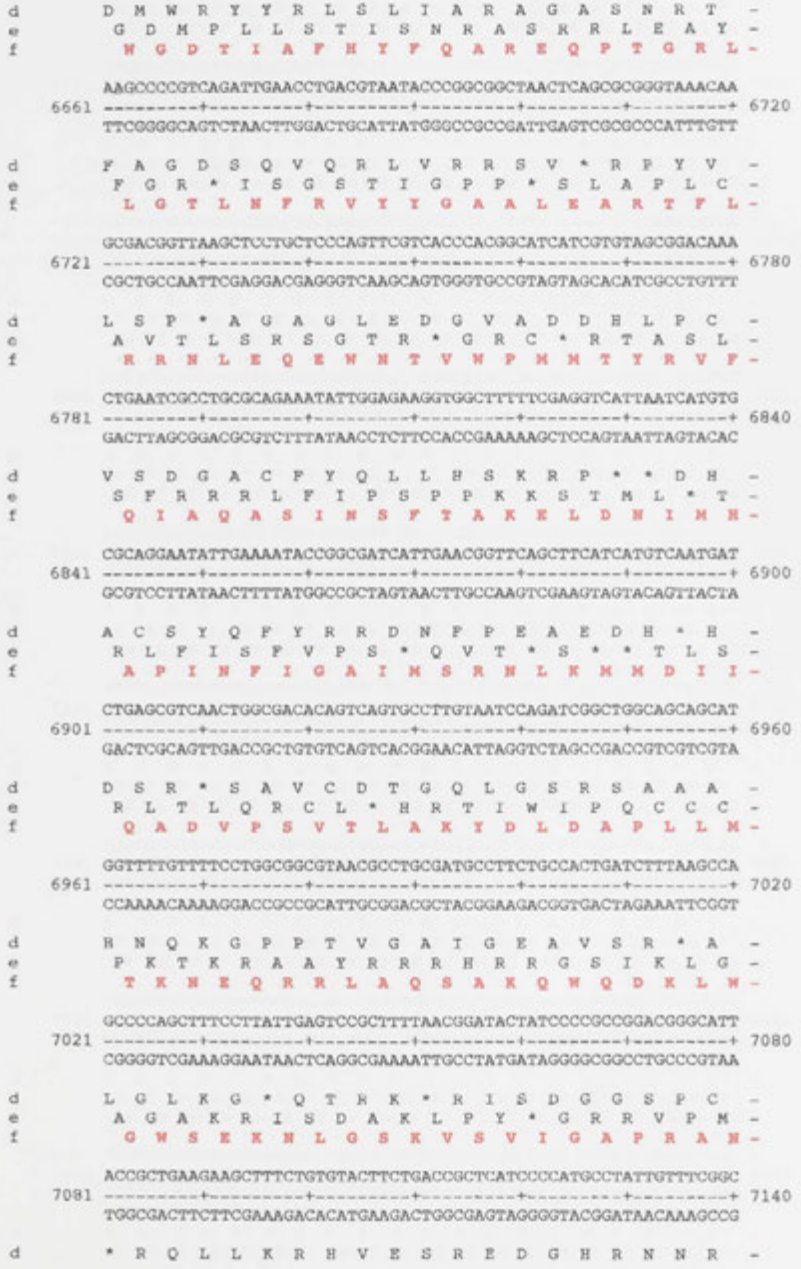

Stop orf107
Start orf107
Stop orf409

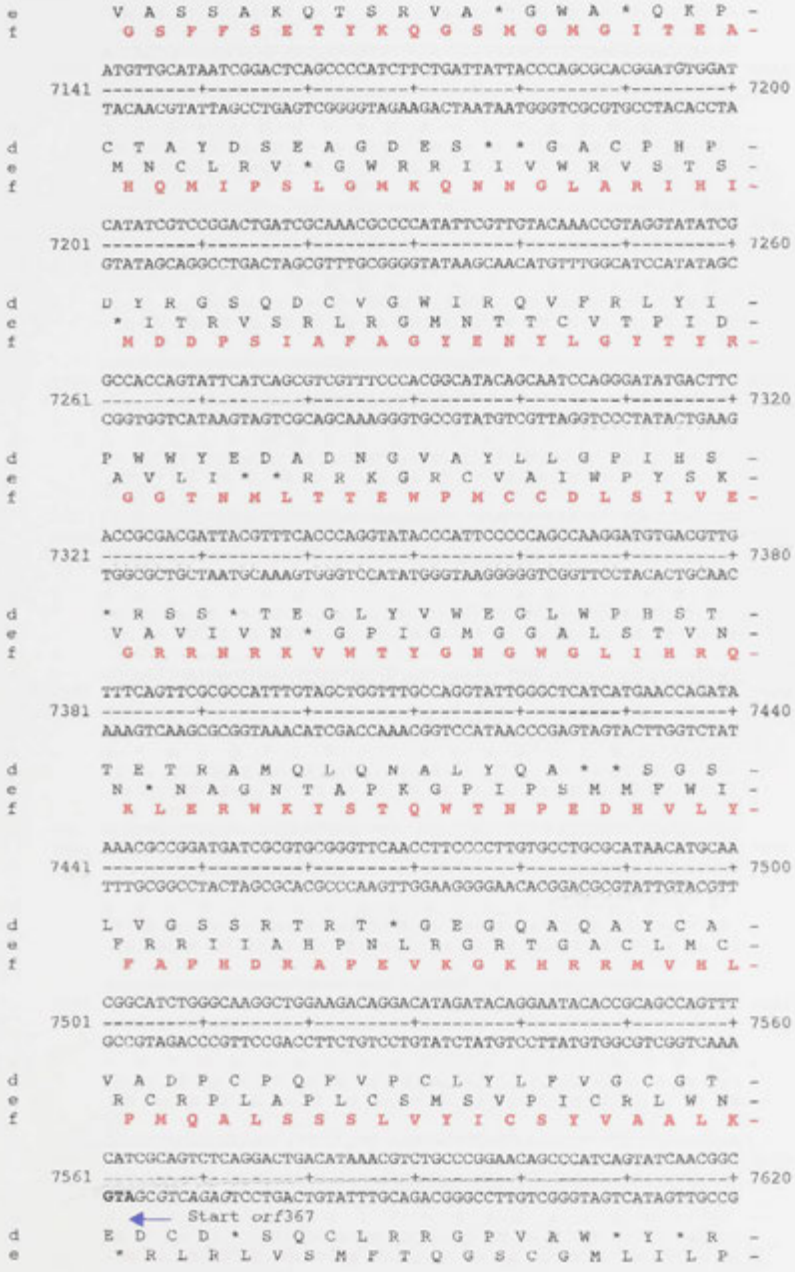


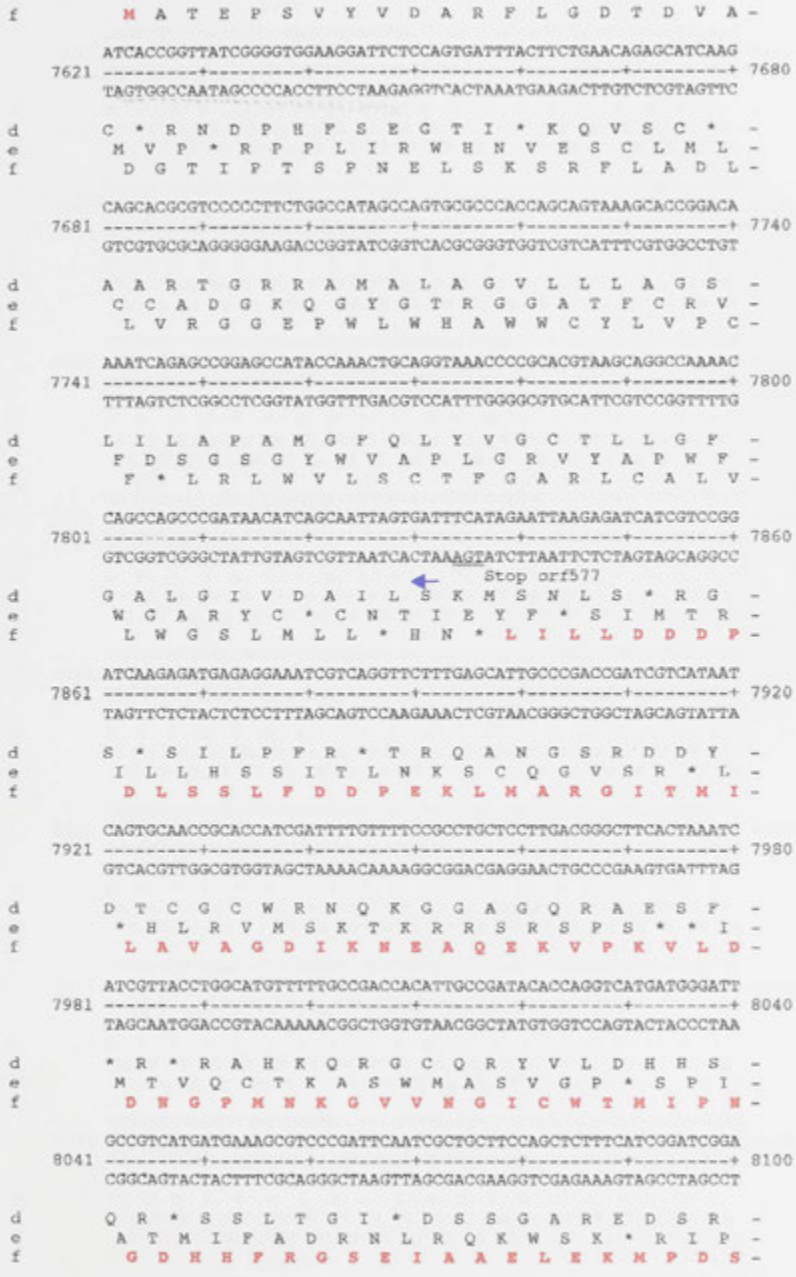


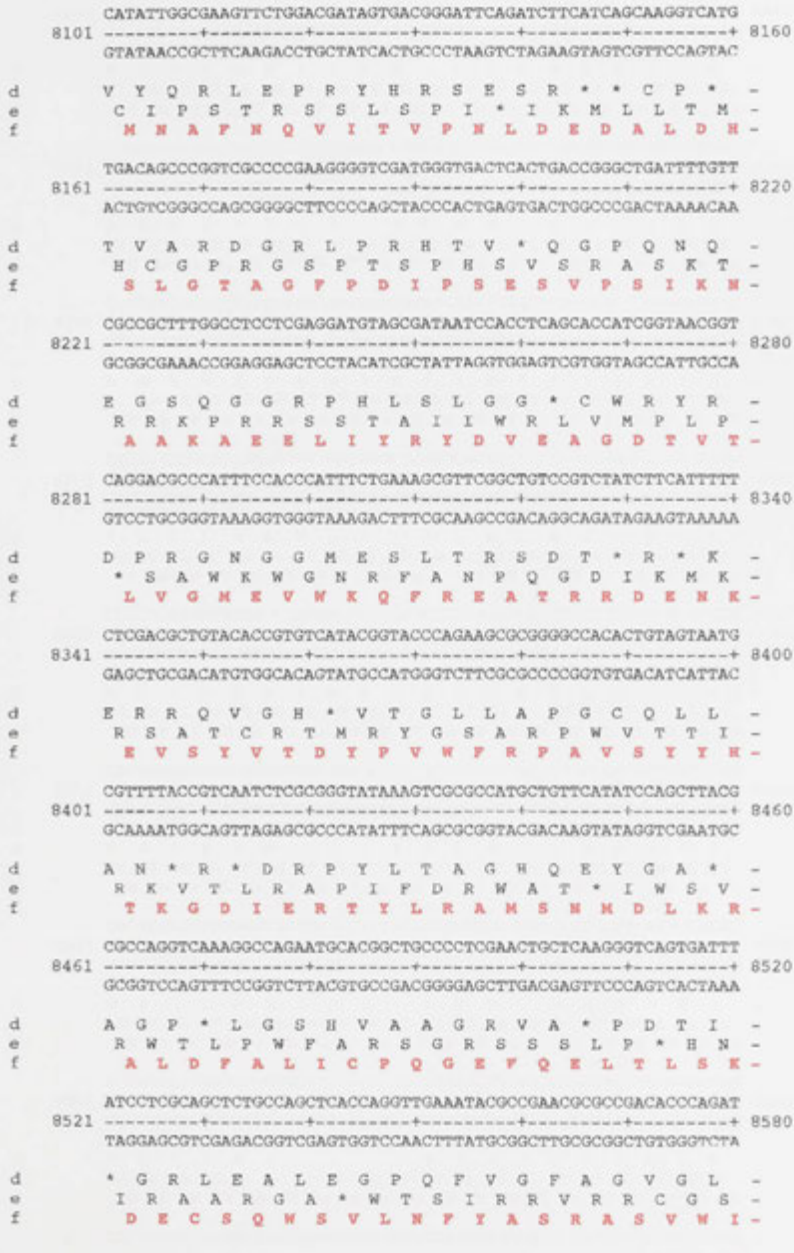

CGAGAAGGAGTAACGAAAGTTACATGCGGACTAGGACTGCTGCGTCAAGAAGAAGCGCAC
d Q E E E N S E I Y A Q D Q R R L E E E R -
e A R G * Q K * H V G S G S S A T R R R T -
f S K R M A K L T R R I R V V C N K K A H -
CAATTCGTTCATCAAGCGCTTCCAGTTCCGATTTTGCTTTGTTCCACTCCGTGCGCTGCT
5701 ---------+---------+---------+---------+---------+---------+ 5760
GTTAAGCAAGTAGTTCGCGAAGGTCAAGCCTAAAACGAAACAAGGTGAGGCACGCGACGA
d A I R E D L A E L E S K A K N W E T R Q -
e C N T * * A S G T R I K S Q E V G H A A -
f L E N M L R K W N P N Q K T G S R A S S -
CTTCCGTCCATGCGTTATCACCAATTTTTTCATTCAGGGCGCGCATGTCAGTTGCGATAG
5761 ---------+---------+---------+---------+---------+---------+ 5820
GAAGGCAGGTACGCAATAGTGGTTAAAAAAGTAAGTCCCGCGCGTACAGTCAACGCTATC
d E E T W A N D G I K E N L A R M D T A I -
e R G D M R * * W N K * K P R A H * N R Y -
f K R G H T I V L K K M * P A C T L Q S L -
TATTACGTTTCTGTTTCAGTTCATGCAGTTTCATGATGTTTCCTTTACGCGTTAAGAAGG
5821 ---------+---------+---------+---------+---------+---------+ 5880
ATAATGCAAAGACAAAGTCAAGTACGTCAAAGTACTACAAAGGAAATGCGCAATTCTTCC
Startorf209 Stop orf200
d T N R K Q K L E H L K M I N G K R T L F -
e Y * T E T E T * A T E H H K R * A N L L -
f I V N R N * N M C N * S T E K V R * S P -
GTCAGGACGCGTTCACGCGCCATACGTTGATTAATGGCTTTCTGTAGCGCGCCGCTGTTG
5881 ---------+---------+---------+---------+---------+---------+ 5940
CAGTCCTGCGCAAGTGCGCGGTATGCAACTAATTACCGAAAGACATCGCGCGGCGACAAC
d P * S A N V R W V N I L P K R Y R A A T -
e T L V R E R A M R Q N I A K Q L A G S N -
f D P R T * A G Y T S * H S E T A R R Q Q -
CGCGCCTCCTGCCATGCTTTCATGGAGCGAACAGCCGAGTCAGCCTCCTGATAGGCAGGA
5941 ---------+---------+---------+---------+---------+---------+ 6000
GCGCGGAGGACGGTACGAAAGTACCTCGCTTGTCGGCTCAGTCGGAGGACTATCCGTCCT
d A R R R G H K * P A F L R T L R R I P L -
e R A E Q W A K M S R V A S D A E Q Y A P -
f A G G A M S E H L S C G L * G G S L C S -
TATGTCACAGGACTGACATCCAGCAGACGGGAAAAGCGGGTTATCTCGCGAATAACAACC
6001 ---------+---------+---------+---------+---------+---------+ 6060
ATACAGTGTCCTGACTGTAGGTCGTCTGCCCTTTTCGCCCAATAGAGCGCTTATTGTTGG
d I H * L V S M W C V P F A P * R A F L L -
e Y T V P S V D L L R S F R T I E R I V V -
f I D C S Q C G A S P F L P N D R S Y C G -
CCGTCCTCATCCTGATACCACTCCTCACCGTCACGGGCGACACGGAAAGCGAAAGATGAC
6061 ---------+---------+---------+---------+---------+---------+ 6120
GGCAGGAGTAGGACTATGGTGAGGAGTGGCAGTGCCCGCTGTGCCTTTCGCTTTCTACTG
d G T R M R I G S R V T V P S V S L S L H -
e G D E D Q Y W E E G D R A V R F A F S S -
f R G * G S V V G * R * P R C P F R F I V -
TGGTTAATATCTCCACGTTGCATCGGGGCCAGCACCAGATCACGAATGGTCTGTGTCTCC
6121 ---------+---------+---------+---------+---------+---------+ 6180
ACCAATTATAGAGGTGCAACGTAGCCCCGGTCGTGGTCTAGTGCTTACCAGACACAGAGG

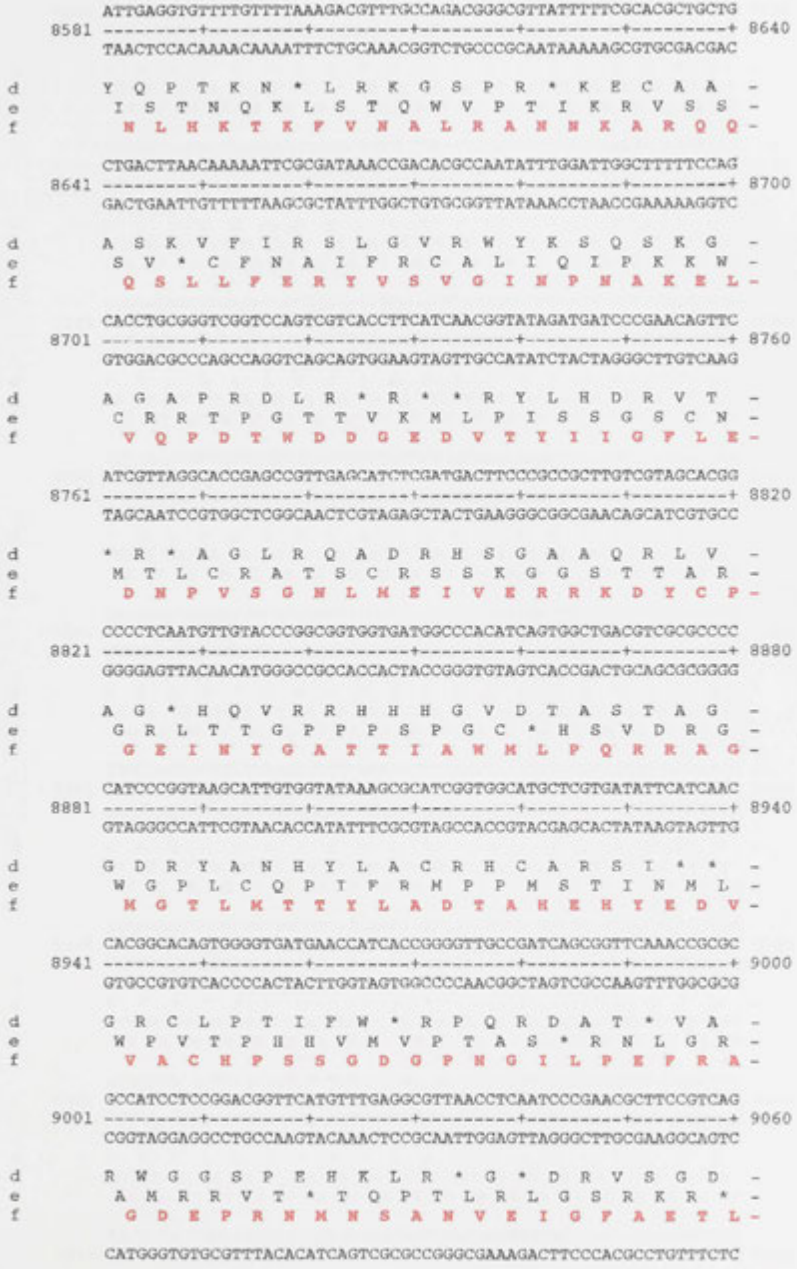

Stop orf367/Start orf200

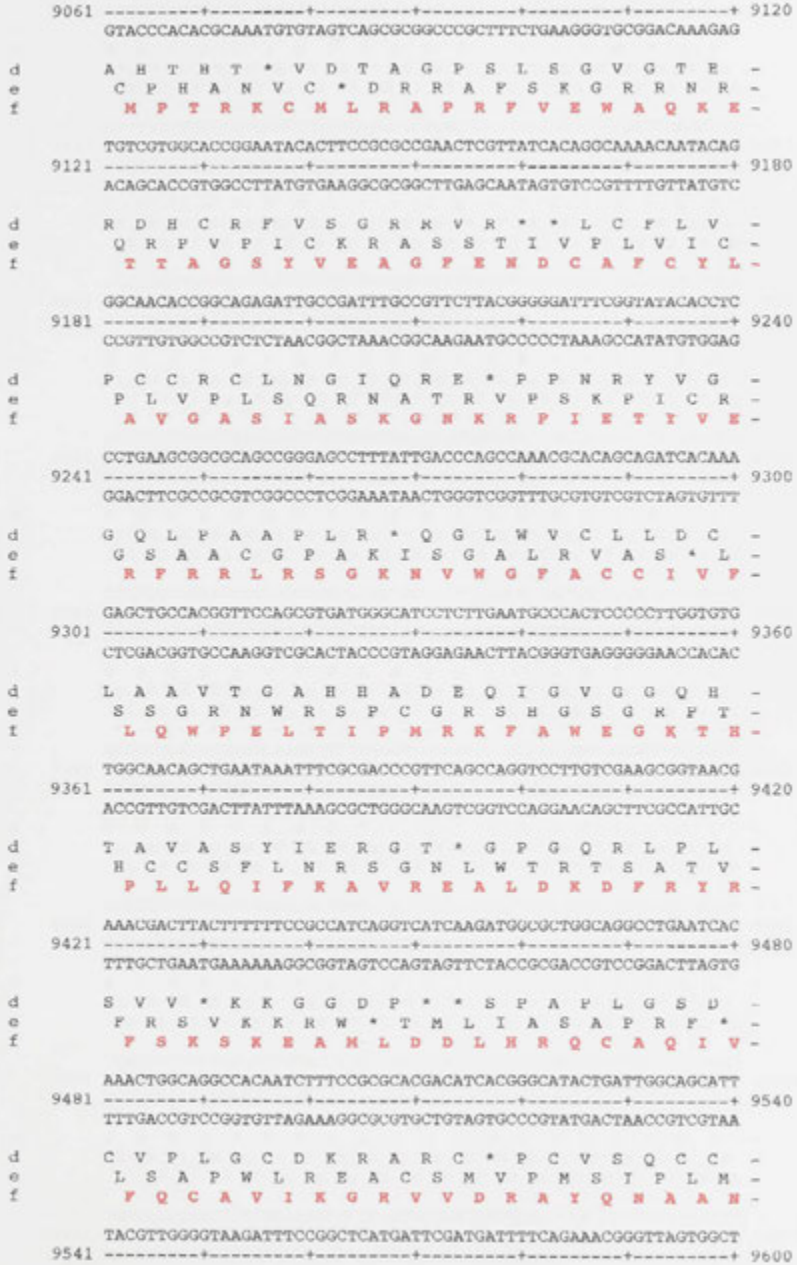

d D M W R Y Y R L S L I A R A G A S N R T -
e G D M P L L S T I S N R A S R R L E A Y -
f W G D T I A F H Y F Q A R E Q P T G R L -
AAGCCCCGTCAGATTGAACCTGACGTAATACCCGGCGGCTAACTCAGCGCGGGTAAACAA
6661 ----------+----------+----------+----------+----------+----------+ 6720
TTCGGGGCAGTCTAACTTGGACTGCATTATGGGCCGCCGATTGAGTCGCGCCCATTTGTT
d F A G D S Q V Q R L V R R S V * R P Y V -
e F G R * I S G S T I G P P * S L A P L C -
f L G T L N F R V Y Y G A A L E A R T F L -
GCGACGGTTAAGCTCCTGCTCCCAGTTCGTCACCCACGGCATCATCGTGTAGCGGACAAA
6721 ----------+----------+----------+----------+----------+----------+ 6780
CGCTGCCAATTCGAGGACGAGGGTCAAGCAGTGGGTGCCGTAGTAGCACATCGCCTGTTT
d L S P * A G A G L E D G V A D D H L P C -
e A V T L S R S G T R * G R C * R T A S L -
f R R N L E Q E W N T V W P M M T Y R V F -
CTGAATCGCCTGCGCAGAAATATTGGAGAAGGTGGCTTTTTCGAGGTCATTAATCATGTG
6781 ----------+----------+----------+----------+----------+----------+ 6840
GACTTAGCGGACGCGTCTTTATAACCTCTTCCACCGAAAAAGCTCCAGTAATTAGTACAC
d V S D G A C F Y Q L L H S K R P * * D H -
e S F R R R L F I P S P P K K S T M L * T -
f Q I A Q A S I N S F T A K E L D N I M H -
CGCAGGAATATTGAAAATACCGGCGATCATTGAACGGTTCAGCTTCATCATGTCAATGAT
6841 ----------+----------+----------+----------+----------+----------+ 6900
GCGTCCTTATAACTTTTATGGCCGCTAGTAACTTGCCAAGTCGAAGTAGTACAGTTACTA
d A C S Y Q F Y R R D N F P E A E D H * H -
e R L F I S F V P S * Q V T * S * * T L S -
f A P I N F I G A I M S R N L K M M D I I -
CTGAGCGTCAACTGGCGACACAGTCAGTGCCTTGTAATCCAGATCGGCTGGCAGCAGCAT
6901 ----------+----------+----------+----------+----------+----------+ 6960
GACTCGCAGTTGACCGCTGTGTCAGTCACGGAACATTAGGTCTAGCCGACCGTCGTCGTA
d D S R * S A V C D T G Q L G S R S A A A -
e R L T L Q R C L * H R T I W I P Q C C C -
f Q A D V P S V T L A K Y D L D A P L L M -
GGTTTTGTTTTCCTGGCGGCGTAACGCCTGCGATGCCTTCTGCCACTGATCTTTAAGCCA
6961 ----------+----------+----------+----------+----------+----------+ 7020
CCAAAACAAAAGGACCGCCGCATTGCGGACGCTACGGAAGACGGTGACTAGAAATTCGGT
d R N Q K G P P T V G A T G E A V S R * A -
e P K T K R A A Y R R R H R R G S I K L G -
f T K N E Q R R L A Q S A K Q W Q D K L W -
GCCCCAGCTTTCCTTATTGAGTCCGCTTTTAACGGATACTATCCCCGCCGGACGGGCATT
7021 ----------+----------+----------+----------+----------+----------+ 7080
CGGGGTCGAAAGGAATAACTCAGGCGAAAATTGCCTATGATAGGGGCGGCCTGCCCGTAA
d L G L K G * Q T R K * R I S D G G S P C -
e A G A K R I S D A K L P Y * G R R V P M -
f G W S E K N L G S K V S V I G A P R A N -
ACCGCTGAAGAAGCTTTCTGTGTACTTCTGACCGCTCATCCCCATGCCTATTGTTTCGGC
7081 ----------+----------+----------+----------+----------+----------+ 7140
TGGCGACTTCTTCGAAAGACACATGAAGACTGGCGAGTAGGGGTACGGATAACAAAGCCG
d * R Q L L K R H V E S R E D G H R N N R -

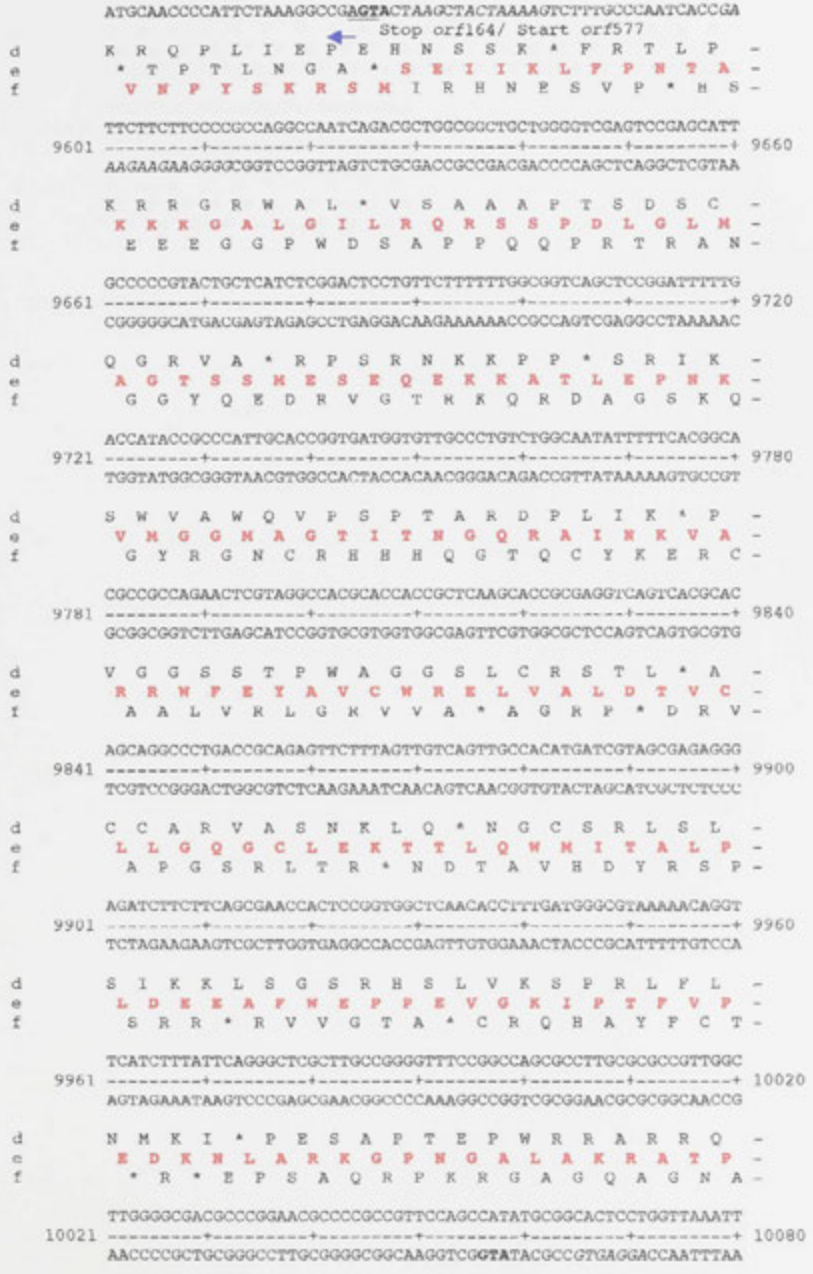

e V A S S A K Q T S R V A * G W A * Q K P -
f G S F F S E T Y K Q G S M G M G I T E A -
ATGTTGCATAATCGGACTCAGCCCCATCTTCTGATTATTACCCAGCGCACGGATGTGGAT
7141 ---------+---------+---------+---------+---------+---------+ 7200
TACAACGTATTAGCCTGAGTCGGGGTAGAAGACTAATAATGGGTCGCGTGCCTACACCTA
d C T A Y D S E A G D E S * * G A C P H P -
e M N C L R V * G W R R I I V W R V S T S -
f H Q M I P S L G M K Q N N G L A R I H I -
CATATCGTCCGGACTGATCGCAAACGCCCCATATTCGTTGTACAAACCGTAGGTATATCG
7201 ---------+---------+---------+---------+---------+---------+ 7260
GTATAGCAGGCCTGACTAGCGTTTGCGGGGTATAAGCAACATGTTTGGCATCCATATAGC
d D Y R G S Q D C V G W I R Q V F R L Y I -
e * I T R V S R L R G M N T T C V T P I D -
f M D D P S I A F A G Y E N Y L G Y T Y R -
GCCACCAGTATTCATCAGCGTCGTTTCCCACGGCATACAGCAATCCAGGGATATGACTTC
7261 ---------+---------+---------+---------+---------+---------+ 7320
CGGTGGTCATAAGTAGTCGCAGCAAAGGGTGCCGTATGTCGTTAGGTCCCTATACTGAAG
d P W W Y E D A D N G V A Y L L G P I H S -
e A V L I * * R R K G R C V A I W P Y S K -
f G G T N M L T T E W P M C C D L S I V E -
ACCGCGACGATTACGTTTCACCCAGGTATACCCATTCCCCCAGCCAAGGATGTGACGTTG
7321 ---------+---------+---------+---------+---------+---------+ 7380
TGGCGCTGCTAATGCAAAGTGGGTCCATATGGGTAAGGGGGTCGGTTCCTACACTGCAAC
d * R S S * T E G L Y V W E G L W P H S T -
e V A V I V N * G P I G M G G A L S T V N -
f G R R N R K V W T Y G N G W G L I H R Q -
TTTCAGTTCGCGCCATTTGTAGCTGGTTTGCCAGGTATTGGGCTCATCATGAACCAGATA
7381 ---------+---------+---------+---------+---------+---------+ 7440
AAAGTCAAGCGCGGTAAACATCGACCAAACGGTCCATAACCCGAGTAGTACTTGGTCTAT
d T E T R A M Q L Q N A L Y Q A * * S G S -
e N * N A G N T A P K G P I P S M M F W I -
f K L E R W K Y S T Q W T N P E D H V L Y -
AAACGCCGGATGATCGCGTGCGGGTTCAACCTTCCCCTTGTGCCTGCGCATAACATGCAA
7441 ---------+---------+---------+---------+---------+---------+ 7500
TTTGCGGCCTACTAGCGCACGCCCAAGTTGGAAGGGGAACACGGACGCGTATTGTACGTT
d L V G S S R T R T * G E G Q A Q A Y C A -
e F R R I I A H P N L R G R T G A C L M C -
f F A P H D R A P E V K G K H R R M V H L -
CGGCATCTGGGCAAGGCTGGAAGACAGGACATAGATACAGGAATACACCGCAGCCAGTTT
7501 ---------+---------+---------+---------+---------+---------+ 7560
GCCGTAGACCCGTTCCGACCTTCTGTCCTGTATCTATGTCCTTATGTGGCGTCGGTCAAA
d V A D P C P Q F V P C L Y L F V G C G T -
e R C R P L A P L C S M S V P I C R L W N -
f P M Q A L S S S L V Y I C S Y V A A L K -
CATCGCAGTCTCAGGACTGACATAAACGTCTGCCCGGAACAGCCCATCAGTATCAACGGC
7561 ---------+---------+---------+---------+---------+---------+ 7620
GTAGCGTCAGAGTCCTGACTGTATTTGCAGACGGGCCTTGTCGGGTAGTCATAGTTGCCG
Start orf367
d E D C D * S Q C L R R G P V A W * Y * R -
e * R L R L V S M F T Q G S C G M L I L P -

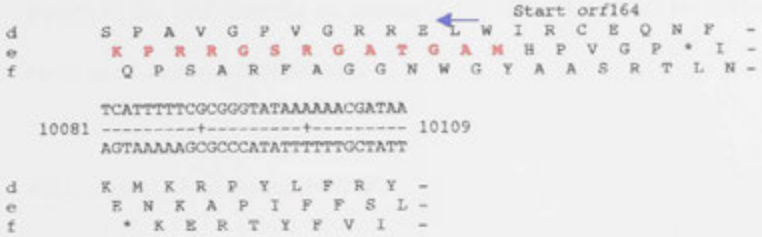

f M A T E P S V Y V D A R F L G D T D V A -
ATCACCGGTTATCGGGGTGGAAGGATTCTCCAGTGATTTACTTCTGAACAGAGCATCAAG
7621 + 7680
TAGTGGCCAATAGCCCCACCTTCCTAAGAGGTCACTAAATGAAGACTTGTCTCGTAGTTC
d C * R N D P H F S E G T I * K Q V S C * -
e M V P * R P P L I R W H N V E S C L M L -
f D G T I P T S P N E L S K S R F L A D L -
CAGCACGCGTCCCCCTTCTGGCCATAGCCAGTGCGCCCACCAGCAGTAAAGCACCGGACA
7681 + 7740
GTCGTGCGCAGGGGGAAGACCGGTATCGGTCACGCGGGTGGTCGTCATTTCGTGGCCTGT
d A A R T G R R A M A L A G V L L L A G S -
e C C A D G K Q G Y G T R G G A T F C R V -
f L V R G G E P W L W H A W W C Y L V P C -
AAATCAGAGCCGGAGCCATACCAAACTGCAGGTAAACCCCGCACGTAAGCAGGCCAAAAC
7741 + 7800
TTTAGTCTCGGCCTCGGTATGGTTTGACGTCCATTTGGGGCGTGCATTCGTCCGGTTTTG
d L I L A P A M G F Q L Y V G C T L L G F -
e F D S G S G Y W V A P L G R V Y A P W F -
f F * L R L W V L S C T F G A R L C A L V -
CAGCCAGCCCGATAACATCAGCAATTAGTGATTTCATAGAATTAAGAGATCATCGTCCGG
7801 + 7860
GTCGGTCGGGCTATTGTAGTCGTTAATCACTAAAGTATCTTAATTCTCTAGTAGCAGGCC
Stop orf577
d G A L G I V D A I L S K M S N L S * R G -
e W C A R Y C * C N T I E Y F * S I M T R -
f L W G S L M L L * H N * L I L L D D D P -
ATCAAGAGATGAGAGGAAATCGTCAGGTTCTTTGAGCATTGCCCGACCGATCGTCATAAT
7861 + 7920
TAGTTCTCTACTCTCCTTTAGCAGTCCAAGAAACTCGTAACGGGCTGGCTAGCAGTATTA
d S * S I L P F R * T R Q A N G S R D D Y -
e I L L H S S I T L N K S C Q G V S R * L -
f D L S S L F D D P E K L M A R G I T M I -
CAGTGCAACCGCACCATCGATTTTGTTTTCCGCCTGCTCCTTGACGGGCTTCACTAAATC
7921 + 7980
GTCACGTTGGCGTGGTAGCTAAAACAAAAGGCGGACGAGGAACTGCCCGAAGTGATTTAG
d D T C G C W R N Q K G G A G Q R A E S F -
e * H L R V M S K T K R R S R S P S * * I -
f L A V A G D I K N E A Q E K V P K V L D -
ATCGTTACCTGGCATGTTTTTGCCGACCACATTGCCGATACACCAGGTCATGATGGGATT
7981 + 8040
TAGCAATGGACCGTACAAAAACGGCTGGTGTAACGGCTATGTGGTCCAGTACTACCCTAA
d * R * R A H K Q R G C Q R Y V L D H H S -
e M T V Q C T K A S W M A S V G P * S P I -
f D N G P M N K G V V N G I C W T M I P N -
GCCGTCATGATGAAAGCGTCCCGATTCAATCGCTGCTTCCAGCTCTTTCATCGGATCGGA
8041 + 8100
CGGCAGTACTACTTTCGCAGGGCTAAGTTAGCGACGAAGGTCGAGAAAGTAGCCTAGCCT
d Q R * S S L T G I * D S S G A R E D S R -
e A T M I F A D R N L R Q K W S K * R I P -
f G D H H F R G S E I A A E L E K M P D S -


CATATTGGCGAAGTTCTGGACGATAGTGACGGGATTCAGATCTTCATCAGCAAGGTCATG
8101
GTATAACCGCTTCAAGACCTGCTATCACTGCCCTAAGTCTAGAAGTAGTCGTTCCAGTAC
8160
d V Y Q R L E P R Y H R S E S R * * C P * -
e C I P S T R S S L S P I * I K M L L T M -
f M N A F N Q V I T V P N L D E D A L D H -
TGACAGCCCGGTCGCCCCGAAGGGGTCGATGGGTGACTCACTGACCGGGCTGATTTTGTT
8161
ACTGTCGGGCCAGCGGGGCTTCCCCAGCTACCCACTGAGTGACTGGCCCGACTAAAACAA
8220
d T V A R D G R L P R H T V * Q G P Q N Q -
e H C G P R G S P T S P H S V S R A S K T -
f S L G T A G F P D I P S E S V P S I K N -
CGCCGCTTTGGCCTCCTCGAGGATGTAGCGATAATCCACCTCAGCACCATCGGTAACGGT
8221
GCGGCGAAACCGGAGGAGCTCCTACATCGCTATTAGGTGGAGTCGTGGTAGCCATTGCCA
8280
d E G S Q G G R P H L S L G G * C W R Y R -
e R R K P R R S S T A I I W R L V M P L P -
f A A K A E E L I Y R Y D V E A G D T V T -
CAGGACGCCCATTTCCACCCATTTCTGAAAGCGTTCGGCTGTCCGTCTATCTTCATTTTT
8281
GTCCTGCGGGTAAAGGTGGGTAAAGACTTTCGCAAGCCGACAGGCAGATAGAAGTAAAAA
8340
d D P R G N G G M E S L T R S D T * R * K -
e * S A W K W G N R F A N P Q G D I K M K -
f L V G M E V W K Q F R E A T R R D E N K -
CTCGACGCTGTACACCGTGTCATACGGTACCCAGAAGCGCGGGGGCCACACTGTAGTAATG
8341
GAGCTGCGACATGTGGCACAGTATGCCATGGGTCTTCGCGCCCCGGTGTGACATCATTAC
8400
d E R R Q V G H * V T G L L A P G C Q L L -
e R S A T C R T M R Y G S A R P W V T T I -
f E V S Y V T D I P V W F R P A V S Y I H -
CGTTTTACCGTCAATCTCGCGGGTATAAAGTCGCGCCATGCTGTTCATATCCAGCTTACG
8401
GCAAAATGGCAGTTAGAGCGCCCATATTTCAGCGCGGTACGACAAGTATAGGTCGAATGC
8460
d A N * R * D R P Y L T A G H Q E Y G A * -
e R K V T L R A P I F D R W A T * I W S V -
f T K G D I E R T Y L R A M S N M D L K R -
CGCCAGGTCAAAGGCCAGAATGCACGGCTGCCCCTCGAACTGCTCAAGGGTCAGTGATTT
8461
GCGGTCCAGTTTCCGGTCTTACGTGCCGACGGGGAGCTTGACGAGTTCCCAGTCACTAAA
8520
d A G P * L G S H V A A G R V A * P D T I -
e R W T L P W F A R S G R S S S L P * H N -
f A L D F A L I C P Q G E F Q E L T L S K -
ATCCTCGCAGCTCTGCCAGCTCACCAGGTTGAAATACGCCGAACGCGCCGACACCCAGAT
8521
TAGGAGCGTCGAGACGGTCGAGTGGTCCAACTTTATGCGGCTTGCGCGGCTGTGGGTCTA
8580
d * G R L E A L E G P Q F V G F A G V G L -
e I R A A R G A * W T S I R R V R R C G S -
f D E C S Q W S V L N F Y A S R A S V W I -






Stop orf164/ Start orf577


Start orf164
d S P A V G P V G R R E L W I R C E Q N F -
e K P R R G S R G A T G A M H P V G P * I -
f Q P S A R F A G G N W G Y A A S R T L N -
TCATTTTTCGCGGGTATAAAAAACGATAA
10081 ---------+---------+--------- 10109
AGTAAAAAGCGCCCATATTTTTTGCTATT
d K M K R P Y L F R Y -
e E N K A P I F F S L -
f * K E R T Y F V I -

Results of the ORF analysis are summarised in Table 4.1, with the ORFs designated based on the number of amino acids.

4.2.2. *Putative Functions of the ORFs*

Predicted amino acids of the ORFs were derived and compared to DNA and protein sequences in the databases using BLASTP [Altschul, 1997 #313], FASTA and FASTX programs [Pearson, 1988 #314]. Significant protein similarities were detected between those in the database and the translation products of ORFs encoding for late phage genes such as the terminase, protease enzymes and phage structural proteins like the tail sheath, portal protein, and capsid. The most significant similarities are reported in Table 4.2. Eight of the 14 putative ORFs showed similarity to other phage protein homologues. *orf164* had similarity to hypothetical proteins of unknown function expressed by bacterial genes of cryptic prophage origin. It also showed similarity to unknown proteins of bacteriophage GMSE-1 and *S. thermophilus* bacteriophage Sfi19. *Orfs107, 104, 168, 186* and *118* did not pick up significant homology to any database sequence. The remaining ORFs with relevant phage protein functions are presented in the succeeding sections.

4.2.3. *ORF577*

The putative gene product (gp) of phage SfV ORF577, has similarity to the large terminase subunits and putative terminases of several bacteriophages which infect diverse

Table 4.2. Table of similarity between phage SfV fragment A predicted gene products & proteins from the GenBank and Swissprot databases

ORF	Organism Matched	p	Percent Identity	Protein ID Locus	Accession Number	Predicted Function
164	*E.coli* 0157:H7 prophage CP-933C	2.0e-10	54% in 90 aa	AAG55950	AE005328	unknown
	Bacteriophage GMSE-1	2.0e-10	46% in 141 aa	AAG50266	AF311659	unknown
	H. influenzae Genetic Island 1	1.0e-09	36% in 90 aa	AAF27358	AF198256	unknown
	X. nematophilus proviral orf8	5.7e-06	27.8% in 158 aa	CAB58451 XNE133022	AJ133022	hypothetical protein
	M. tuberculosis H37Rv segment 70/162	5.7e-05	27.7% in 155 aa	CAB09059 MTCY336	Z95586	unknown
	M. tuberculosis H37Rv segment 118/162	9.8e-05	29.6% in 162 aa	CAB02357 MTCY441	Z80225	unknown
	S. thermophilus sfi19	2.0e-04	29.0% in 92 aa	AAD44055	AF115102	unknown
577	*E. coli* INTE-PIN region	2.4e-152	96.5% in 367 aa	YMFN_ECOLI	P75978	hypothetical 50.9 kD protein
	P. aeruginosa phage D3	2.6e-83	46.3% in 570 aa	AAD38954	AF165214	terminase
	R. capsulatus strain SB1003	3.6e-51	31.1%in 559 aa	AAC16226	AF010496	hypothetical protein
	H. influenzae Genetic Island 1	5.0e-61	29% in 552 aa	AAF27357	AF198256	phage D3 terminase-like protein
	S. aureus phage phi PVL	1.4e-29	24% in 567 aa	BAA31875	AB009866	unknown
	B. subtilis phage phi 105	6.8e-28	29% in 415 aa	BAA36628	AB016282	unknown
	L. casei phage A2 gp3	4.0e-25	27% in 585 aa	BPHA3GP3	X97563	unknown
	Streptomyces phage phi C31	1.2e-13	25.3% in 522aa	CAA07103	AJ006589	putative large terminase subunit
	S. thermophilus sfi19	9.7e-12	21.9% in 562 aa	AAD44056	AF115102	putative large terminase subunit
	S. thermophilus sfi21	8.3e-11	21.1% in 567 aa	AAD41029	AF112470	terminase large subunit
	S. thermophilus phage DT1	4.7e-09	33.1% in 121 aa	AAD21879	AF085222	putative terminase small subunit
367	*B. subtilis* phage phi 105	4.2e-38	29.9% in 355 aa	BAA36631	AB016282	hypothetical protein 25
	H. influenzae Genetic Island 1	8.0e-38	31% in 352 aa	AAF27362	AF198256	phi-105 ORF25-like protein
	Coliphage HK97	7.5e-27	25.7% in 428 aa	VP3_BPHK7	P49859	portal protein
	P. aeruginosa phage D3	1.4e-24	27.6% in 359 aa	AAD38955	AF165214	portal protein
	Streptomyces phage phi C31	2.0e-20	27.0% in 413 aa	CAA07104	AJ006589	putative portal protein
	R. capsulatus strain SB1003	1.0e-19	26.6% in 368 aa	AAC16225	AF010496	hypothetical protein
	S. aureus phage phi PVL	6.9e-11	22.6% in 349 aa	BAA31877	AB009866	portal protein

Table 4.2. Table of similarity between phage SfV fragment A predicted gene products & proteins from the GenBank and Swissprot databases (continued)

ORF	Organism Matched	p	Percent Identity	Protein ID Locus	Accession Number	Predicted Function
200	*Lactobacillus casei phage A2*	4.0e-24	44% in 171 aa	CAB63684	AJ251790	putative protease
	S. aureus phage phi PVL	9.0e-20	35% in 193 aa	BAA31878	AB009866	unknown
	Streptomyces phage phi C31	5.7e-15	38% in 171 aa	CAA07105	AJ006589	putative protease
	B. subtilis phage phi 105	4.2e-12	36.3% in 173 aa	BAA36632	AB016282	unknown
	H. influenzae Genetic Island 1	3.0e-14	34% in 171 aa	AAF27363	AF198256	phi-C31 protease-like protein
	Coliphage HK97	7.9e-03	29.9% in 157 aa	VP4_BPHK7	P49860	prohead protease
	M.tuberculosis seg. 118/162	3.4e-02	32% in 150 aa	CAB02356	Z80225	similar to VP4_BPHK7
409	*Streptomyces* phage phi C31	1.1e-09	23% in 397 aa	CAA07106	AJ006589	major capsid protein
	H. influenzae Genetic Island 1	2.0e-05	20% in 405 aa	AAF27364	AF198256	phage phi-C31 like capsid protein
	S. thermophilus phage Sfi21	3.0e-05	20% in 374 aa	AAD41033	AF112470	major head protein precursor
	S. thermophilus phage DT1	1.0e-04	22% in 304 aa	AAD21884	AF085222	major head protein
122	No significant homology					unknown
136	No significant homology					unknown
168	No significant homology					unknown
186	No significant homology					unknown
56	phage Mu complete genome	1.7e-05	42.3% in 52 aa	AAF01116	AF083977	HI1510 homologue
	phage Mu DNA for ORF1, sheath protein gpL, ORF 2,3	1.7e-05	42.3% in 52 aa	BAA19194	AB000833	ORF1 major tail subunit
498	*H. influenzae* HI1511	4.9e-56	38.4% in 495aa	YF11_HAEIN	P44233	hypothetical protein
	phage Mu DNA for ORF1, sheath protein gpL, ORF 2,3	9.4e-50	35.8% in 500 aa	BAA19195	AB000833	sheath protein gp L
	N. meningitidis MC58	2.0e-41	29% in 486 aa	AAF41495	AE002460	sheath protein
	M. xanthus gene difE	3.4e-02	24.2% in 248 aa	AAC27632	AF076485	fruiting body formation and motility
118	No significant homology					unknown
89	*P. aeruginosa* phage phi CTX	8.0e-01	27.1% in 85 aa	BAA36251	AB008550	ORF 24 similar to E gene of P2 (essential tail protein)
116	*S. typhimurium* sop E gene	1.2e-09	45.2% in 104 aa	AAD54238	AF153829	similar to protein G of phage 186 (tail protein)
	P. aeruginosa phage phi CTX	2.1e-03	40% in 121 aa	BAA36253	AB008550	ORF 25 similar to T gene of phage P2 (essential tail protein)

hosts (Table 4.2). These phages with their corresponding gene coordinates were: *Pseudomonas aeruginosa* phage D3 ORF2 (nt 419-2110); *Staphylococcus aureus* phage ϕPVL proviral protein ORF2 (nt 537-2231); *Bacillus subtilis* phage ϕphi-105 ORF22 (nt 612-1862); *Lactobacillus casei* phage A2gp3 ORF5 (nt 1718-3409); *Streptomyces* phage ϕC31 gene 33(nt 1282-2841); *Streptococcus thermophilus* bacteriophage Sfi19 ORF623 (nt 2107-3978); *Streptococcus thermophilus* bacteriophage Sfi21 ORF623 (nt 1738-3609); *Streptococcus thermophilus* phage DT1 ORF3 (nt 1103-1792); *Lactococcus lactis* phage bIL170 gene 12 (nt 779-2401); *Lactococcus lactis* phage p008 ORF2 (nt 1130-2752); *Lactococcus lactis* phage sk1 ORF2 (nt 788-2410); and *Lactobacillus* phage adh ORF624 (20183-22057). The higher percent identities were observed between ORF577 and proteins encoded in bacterial genomes: the region of homology in *E. coli* is in prophage e14 gene region with a 96.5% identity to ORF577 over 367 amino acid residues; a 31% identity over 559 amino acids with that of *Rhodobacter capsulatus* hypothetical protein; and 29% identity over 552 amino acids with that of *Haemophilus influenzae* genetic island 1 ORF10. These data indicated that bacterial terminase genes could be encoded by a cryptic phage. More expected was the second highest homology score (46.3% in 570 amino acids), which was traced to a bacteriophage D3 of *Pseudomonas aeruginosa*. Its ORF1692 was reported to have significant similarity to the same phage and bacterial proteins as those homologous to phage SfV [Gilakjan, 1999 #278]. Multiple sequence alignment of the *E. coli* YmfN hypothetical protein, phage D3 ORF1692 and phage SfV ORF577 revealed conserved regions within a considerably large area of the protein (Figure 4.4). The conserved regions may represent functional

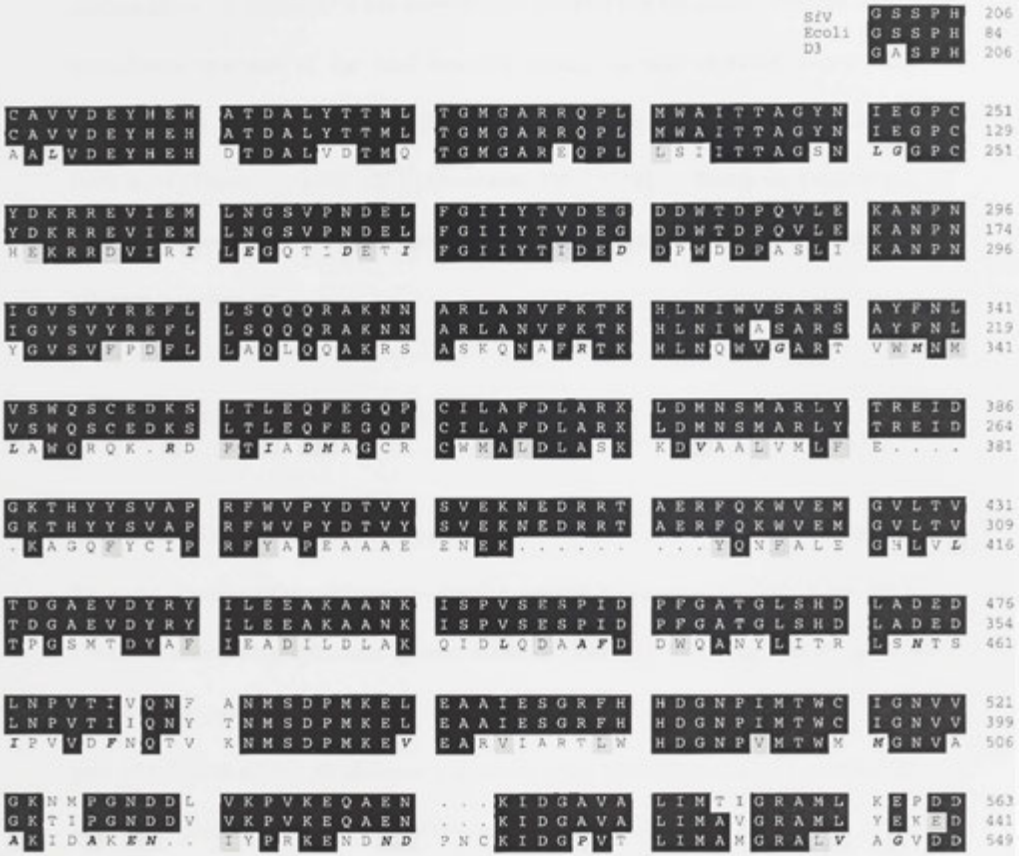


Figure 4.4. Multiple sequence alignment of phage SfV ORF577 encoding the putative large terminase subunit, *E. coli* hypothetical protein YMFN_ECOLI (Accession No. P75978), and *P. aeruginosa* phage D3 ORF 1692 (Accession No. AF165214). Alignment was performed using the Eclustalw program in WAG with a gap penalty of 10. Identical residues are in black boxes, similar residues are boxed in dark gray, somewhat similar residues are italicised, and non-matching residues are not shaded.

terminase active sites which could also exhibit similarity in their secondary structural conformation. In addition, it was interesting to discover that the distinct position *orf577* immediately upstream of the head assembly genes, is also observed among other bacteriophages [Duda, 1995B #289][Lucchini, 1998 #318][Smith, 1999 #276][Kaneko, 1998 #204][Tremblay, 1999 #207][Altermann, 1999 #206]. Based on location and homology to several bacteriophage terminases, evidences indicate that *orf577* encodes the SfV terminase subunit.

4.2.4. *ORF367*

ORF367 is 271 nucleotides downstream of ORF577 and shares significant homology to the portal proteins of the following: *Bacillus subtilis* phage phi-105 ORF25 (nt 2512-3762); *Haemophilus influenzae* genetic island 1 ORF25; Coliphage HK97 gp 3 ; *Pseudomonas aeruginosa* phage D3 ORF4 (nt 2264-3568); *Streptomyces* phage ϕC31 gene 34 (nt 2886-4259); *Rhodobacter capsulatus* strain SB1003 protein ID AAC16225; and *Staphylococcus aureus* phage ϕPVL proviral protein ORF4 (nt 2451-3701). The predicted gene product had a similar isoelectric point (9.45) and molecular mass (41.9 kDa) to the portal proteins of phage D3 and phage DT1 portal protein [Gilakjan, 1999 #278][Tremblay, 1999 #207]. However, other phages had larger portal protein products. For example, phage ϕC31 has a 54 kDa product and phage λ has a 59 kDa product which unlike other phage portal protein are processed. This is why the SfV putative portal protein ORF367 and those of phage D3 and DT1 had higher molecular

masses [Catalano, 1995 #272][Smith, 1999 #276]. The organisation of *orf577* and *orf367* is similar to other phage terminases and portal proteins. Based on the position of the gene, the location of *orf367* which is immediately downstream of the putative terminase gene is similar to the gene arrangement found in other bacteriophages [Duda, 1995B #289][Lucchini, 1998 #318][Smith, 1999 #276][Kaneko, 1998 #204][Tremblay, 1999 #207][Altermann, 1999 #206]. With this observation and considering that the protein product of most genes from the database homologous to ORF367 is the portal protein, it is likely that *orf367* also encodes for a product with a similar function. In addition, several portal proteins were compared and found to be conserved among bacteriophages [Smith, 1999 #276]. Portal protein serves an important role in directing incoming DNAs into the head during the formation of mature phage particles. Thirteen subunits form an annular structure situated at the opening of the prohead shell and serve as passageway for DNA during the packaging process [Dube, 1993 #320]. In phage HK97, the portal protein participates in the capsid maturation process [Duda, 1998 #319].

4.2.5. *ORF200*

A 22.7 kDa protein with an isoelectric point of 4.92, is encoded by *orf200*. FASTA search detected significant homology to *Lactobacillus casei* phage A2 protein, *B. subtilis* phage phi-105 ORF26 (nt 3752-4378) and *S. aureus* phage ϕPVL proviral protein ORF5a (nt 3694-4278) proteins all of unknown function. The search also showed high sequence homology to the protease proteins of *Streptomyces* phage ϕC31 gene 35 (nt 4240-4881)

and coliphage HK97 gp4. Bacterial proteins which are similar to the ORF200 protein include the ORF12 of *H. influenzae* genetic island 1 and *M. tuberculosis* segment 118/162 MTCY441.20C, both of which were reported as similar to coliphage HK97 gp4 protein (SwissProt ID VP4_BPHK7)[Cole, 1998 #328]. Protease functions were assigned to phage ϕC31 23.5 kDa gp35 and coliphage HK97 25 kDa gp4, both have similar molecular mass to the 22.7 kDa phage SfV ORF200. Based on its similarity to several bacteriophage proteases in size and isoelectric point (*pI*), it is tempting to predict a similar function encoded by *orf200*. However, further experiments should be conducted in order to test for its functional properties (Chapter 5). Protease is a compotent of the head assembly genes and has proteolytic action that processes the major head protein (capsid) prior to covalent crosslinking to form the mature head shell [Kaneko, 1998 #204][Gilakjan, 1999 #278][Duda, 1995B #289]. The processing and head maturation of HK97 has been characterised in detail [Duda, 1995C #301]. Similar to the position of phage SfV *orf200,* the protease gene has been located immediately upstream of the capsid gene of other bacteriophages [Kaneko, 1998 #204][Gilakjan, 1999 #278][Duda, 1995B #289][Smith, 1999 #276][Altermann, 1999 #206]. Alignment of SfV ORF200 product with *L. casei* phage A2 putative protease revealed significant amino acid sequence conservation extending throughout the entire length of the proteins (Figure. 4.5).

4.2.6. *ORF409*

The closest homologue to SfV ORF409 is the major capsid protein of temperate phage phi-C31 encoded by gp36. The gene position of the capsid protein is reminiscent of the

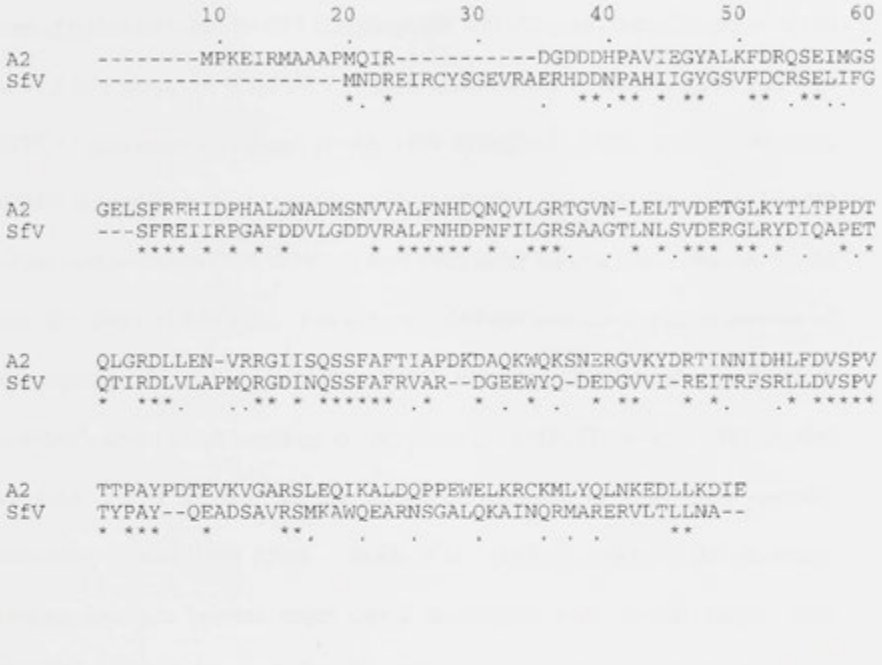


Figure 4.5. Sequence comparison of *Shigella flexneri* phage SfV ORF200 putative protease and *Lactobacillus casei* phage A2 protease gene products (Accession No. AJ251790). Homology is evident throughout the entire protein. Identical residues are indicated by asterisks and conservative changes are indicated by periods. Alignment was performed using the Eclustalw program in WAG.

gene arrangement of the late structural proteins in double-stranded DNA bacteriophages including phi C31, HK97, ϕPVL, D3, ϕadh and phage SfV (Figure 4.6) whose putative protease ORF200 is immediately upstream of ORF409 with only 11 intergenic bases. In terms of molecular mass, the 45.8 kDa phage SfV ORF409 gene product is comparable to the 41.7 kDa protein of phage phi-C31 gp36 which in turn is homologous to coliphage HK97 42 kDa major capsid gp5 [Smith, 1999 #276][Duda, 1995C #301]. Although ORF409 showed the highest homology to phage phi-C31, its probability value (1.1e-09) and percentage identity (23% in 397 aa) were much lower than the scores obtained for the other SfV ORFs (Table 4.2). Furthermore, ORF409 was not similar to proteins of bacteriophages phi PVL, phi-105, HK97 and phage D3 whose gene products consistently showed high homology to other phage SfV ORFs (Table 4.2). This implied that major capsid proteins are less conserved compared to other head assembly components [Smith, 1999 #276]. Smith *et al.*, (1999), compared the percentage sequence identities between major capsid homologues from diverse phages using BESTFIT. They reported that the major capsid proteins generally have less conservation than the proteases and portal proteins. This could be the reason why phage SfV ORF409 only shows significant homology to phage ϕC31 gp 36 and could also explain why similarity to *Pseudomonas* phage D3 capsid was not evident despite the high 46% identity of its ORF1692 terminase gene product with phage SfV ORF577 (Table. 4.2)

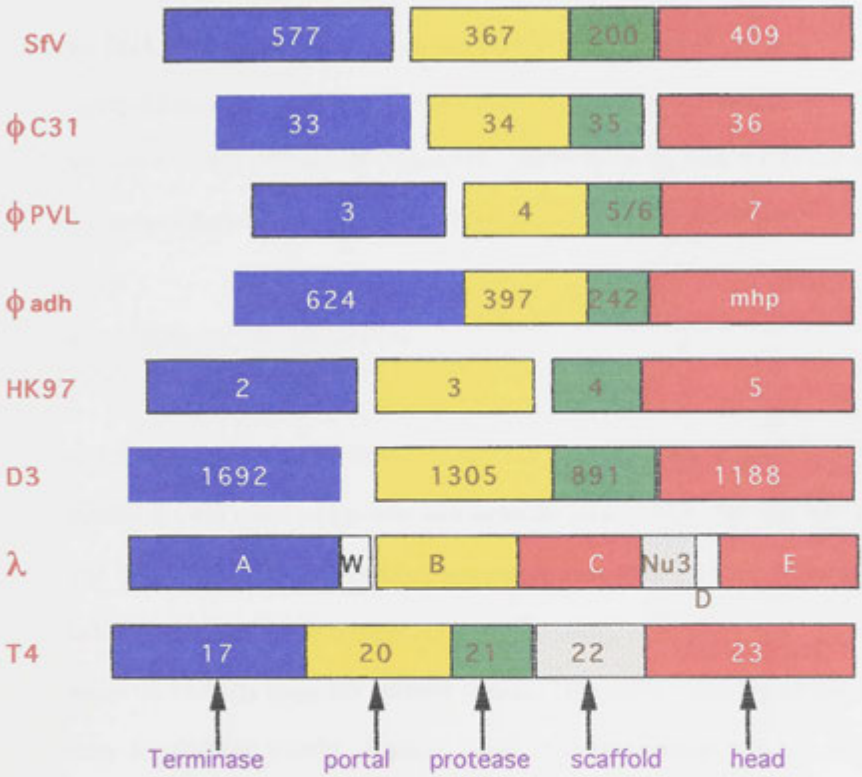


Figure 4.6. Comparison of head assembly gene organization among representative bacteriophages: *Shigella flexneri* phage SfV (Chapter 4); *Streptomyces* phage ϕC31 (Smith *et al.*, 1999); Coliphage HK97 (Duda et al., 1995A); *Pseudomonas aeruginosa* phage D3 (Gilakjan *et al.*, 1999); Bacteriophage λ (Murialdo *et al.*, 1978, Hendrix *et al.*, 1974); *Lactobacillus gasseri* phage ϕadh (Altermann *et al.*, 1999); *Staphylococcus aureus* phage ϕPVL (kaneko *et al.*, 1998); and Bacteriophage T4 (Laemmli, 1970, Dickson *et al.*, 1970, Powell *et al.*, 1990). Note the similarity in the sequential order of the genes or open reading frames. Coloured boxes represent related genes. The terminase protein genes are shown in blue, portal protein genes in yellow, protease genes in green, and the head protein genes in red. Lambda gene D represents the head-DNA stabilization gene and W represents the head-tail joining gene.

At this stage, we can only assume that *orf409* codes for phage SfV capsid protein with an isoelectric point of 5.08. However, the striking conservation in the organisation of the DNA packaging genes positioned immediately upstream of the head assembly genes among different bacteriophages, provided a strong evidence as to the identity of phage SfV *orfs 577, 367, 200* and *409* (Figure 4.6). Based on the location of ORF409 and its protein homologies, it appears that *orf409* encodes the capsid protein of SfV.

4.2.7. *ORFs 56, 498, 89, and 116*

orf498 extends from nt position 2615 to 1119 of the 10.1 kb *Bam*HI fragment A, translating a 498 amino acid protein with molecular mass of 54.8 kDa and a *pI* value of 5.54 (Table 4.1). Following GenBank sequence database search, ORF498 was found to be homologous to the 55 kDa phage Mu tail sheath ORF L (9.4e-56), which is similar in size to the 54.8 kDa phage SfV ORF498 [Takeda, 1998 #321]. The highest probability score for ORF498 FASTX database search (4.9e-56) matched with the sequence designated as HI1511 of the Mu-like prophage integrated into *Haemophilus influenzae* chromosome [Fleischmann, 1995 #210] which also matched highly with phage Mu ORF L [Takeda, 1998 #321]. Immediately upstream of *Haemophilus* HI1511 is HI1510, a protein homologue of phage Mu ORF1 which is located upstream of ORF L. HI1510 showed significant homology to phage SfV ORF56 which is also found upstream of ORF498 (Figure. 4.7 A and B). Similarity searches of predicted ORF89 protein matched with *Pseudomonas aeruginosa* phage ϕCTX ORF24 which is similar to gene E of phage

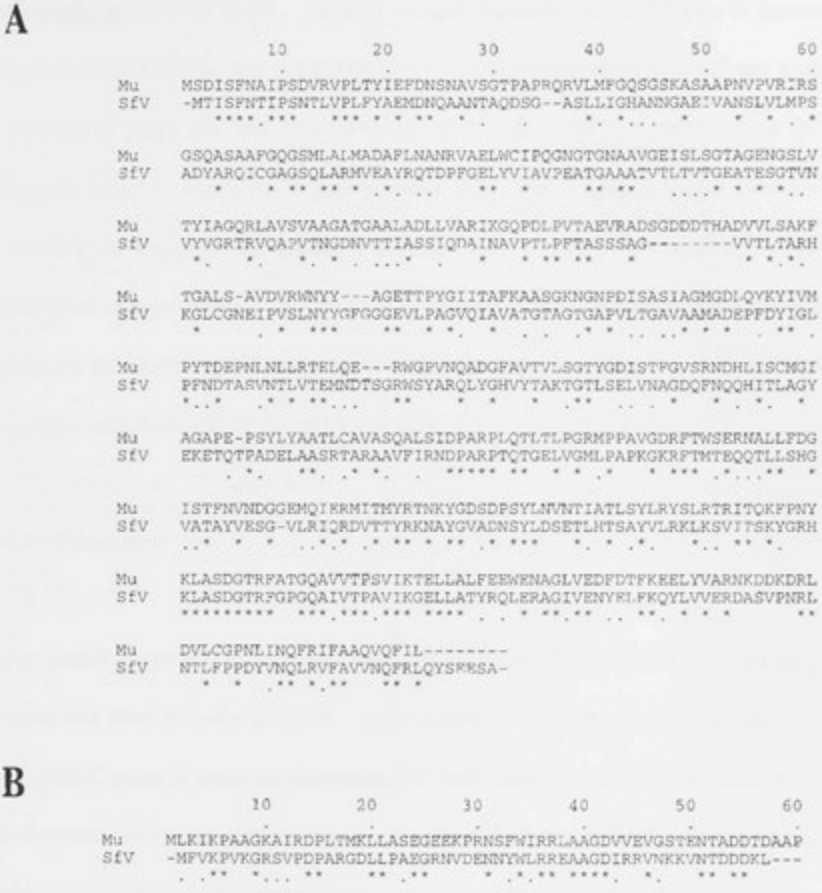


Figure 4.7. Sequence comparison of tail proteins. A, *Shigella flexneri* phage SfV ORF498 putative sheath and Coliphage Mu sheath protein gene L (Accession No. AB000833). B, *Shigella flexneri* phage SfV ORF56 and Coliphage Mu ORF1. Identical residues are indicated by asterisks and conservative changes by periods. Alignment was performed using the Eclustalw program in WAG.

P2 [Nakayama, 1999 #203]. In the GenBank database, phage P2 gene E (accession number AF063097; protein ID AAD03291) was identified as an essential tail protein. ORF116 of phage SfV had 40% homology to *Pseudomonas aeruginosa* phage ϕCTX ORF25 which is similar to T gene of phage P2. The T gene was also identified as encoding an essential tail protein of phage P2. With the pending functional characterisation of phage SfV ORFs498, 56, 89, and 116 and based on sequence homolgy and clustering of these genes in a particular genomic region, we can only deduce that these genes encode for the major and minor tail components of phage SfV.

4.3 Discussion

The *Bam*HI fragment C of bacteriophage SfV flanking the the O-antigen modification gene region has been well-characterised. The important function of proteins encoded by fragment C genes in serotype conversion has been established through sequence analysis and several cloning experiments [Huan, 1997B #117] [Huan, 1997A #118]. For example, Huan and colleagues have constructed a fully serotype-converted *S. flexneri 5a* strain, SFL1168 (Type antigen V) by introducing a 5.0 kb *Bam*HI fragment C of phage SfV into *S. flexneri* serotype Y, SFL124 (Group antigen 3,4). SFL1168 was indistinguishable from the *S. flexneri* 5a wild-type strain. The discovery of phage SfV genomic portion encoding important factors for serotype conversion was followed by other experiments that lead to the localisation of other serotype conversion modules [Adhikari, 1999 #112][Adams, 2000 #523]. Inspired by the wealth of information derived from such

limited segment of the genome, it was deemed necessary to proceed in the analysis and characterisation of the genetic blueprint in order to unravel and fully understand the biology of phage SfV.

In this project, the 10.1 kb portion of *Bam*HI fragment A was sequenced and analysed. This portion yielded putative ORFs, eight of which showed similarity to other phage proteins. The higher FASTA similarity scores of most phage SfV ORFs were traced to the genomes of prophages. Examples were the *E. coli* hypothetical 50.9 kDa protein of e14 prophage, and prophages in *Rhodobacter capsulatus* strain SB1003 and *Haemophilus influenzae* genetic island 1. This indicated a pattern of genetic exchange and transfer of DNA across species mediated by mobile gene carriers like the bacteriophages. Bacteriophages have been known to be important transposing agents of genetic elements from various bacterial species that paved the way to evolutionary progress [Cheetham, 1995 #30]. Bacteriophage infection can relay not only genes that are unnecessary to host existence but also important genes like virulence factors that can transform avirulent bacterial strain into a harmful pathogenic organism [Bishai, 1988 #352][Miao, 1999 #524].

ORF577 showed significant similarity to several phage terminase subunits (Table 4.2). Based on the statistical significance of the E value from the database search and the conserved position of ORF577 immediately upstream of the putative portal protein gene *orf367* (Figure 4.6) [Lucchini, 1998 #318], ORF577 may also have a terminase enzyme

function. The terminases play an active role in DNA packaging during the phage head assembly [Casjens, 1988 #266]. Comprising a small and large subunit, the terminase complex initiates either a sequence specific (*cos*) or non-specific (*pac*) packaging cut in the adjoining head-tail concatemeric DNA and presents the DNA through the portal vertex of the prohead [Earnshaw, 1980 #270][Bazinet, 1985 #326]. The terminase and portal protein interact to form a headfilling packasome complex which facilitates DNA translocation coupled with ATP hydrolysis. When the head is filled with a certain DNA density or a specific sequence is reached, the terminase cuts the DNA [Black, 1995 #325].

Considering the constant cycle of lytic and lysogenic existence of bacteriophages, it was not surprising to detect a high percentage identity (96%) between *orf577* and *E. coli ymfN*. However, the extent of amino acid sequence homology of the *E. coli YMFN* encompassed only a limited region of ORF577 compared to other terminase homologues of bacteriophage origin. It is possible that the putative terminase region in *ymfn* which was introduced by a cryptic phage, had became latent and had acquired several mutations which contributed to the observed homology to a limited region of ORF577 [Campbell, 1996 #324][Blattner, 1997 #4]. It is also possible that phage e14 terminase is encoded by a mosaic gene originally derived from various phage sources through gene recombinations. This could account for its limited sequence identity to ORF577.

Another possible product of bacteriophage promiscuity leading to genetic mosaicism could be SfV *orf164*. Although homologies to proteins of bacteriophages and prophages

were detected, these proteins were reported to have unknown or hypothetical function (Table 4.2). The size and location of *orf164* is similar to other bacteriophage small terminase subunit. It is immediately upstream of *orf577*, which is a location similar to the 20.4 kDa small terminase subunit (gpNu1) of bacteriophage λ gpA-gpNu1 holoenzyme complex [Catalano, 1995 #272][Tomka, 1993 #353]. In phage D3 the predicted small terminase subunit gene *orf381* also lies upstream of the large terminase subunit gene *orf1692* [Gilakjan, 1999 #278]. These terminase protein complex were also observed in other bacteriophages such as the gp3 and gp16 of phi-29, gp2 and gp3 of P22, gpM and gpP of P2, gp16 and gp17 of T4 and gp18 and gp19 of T3 [Powell, 1990 #354][Guo, 1987 #355][Hashimoto, 1992 #356][Eppler, 1991 #357]. Analogies pertaining to *orf164*'s location relative to terminase gene position in other bacteriophages are interesting to note, however, it would be very presumptuous to consider *orf164* as being a terminase component since it only showed significant homology to bacteriophage proteins that are not yet identified. Further studies must be conducted on *orf164* in order to determine its real identity. The terminases together with the portal protein at the prohead portal vertex comprise the packasome [Hendrix, 1978 #358][Earnshaw, 1980 #270][Black, 1995 #325].

The identity of phage Mu gene L and ORF2 as the respective tail sheath and tube genes was confirmed by N-terminal sequencing of the purified Mu tail proteins [Takeda, 1998 #321]. The tail components were purified by sucrose density gradient ultracentrifugation from the cell lysate of heat-induced lysogenic Su⁻ strain containing a phage Mu *Tam*

mutant defective in the phage head synthesis [Shore, 1982 #360]. Significant homology of phage SfV ORF498 protein to the gene L product of phage Mu suggests a possible tail sheath protein function for the ORF498 protein. In considering gene organisation, the immediate upstream location of phage SfV ORF56 to ORF498 is similar to the arrangement of phage Mu gene L which is directly downstream of gene ORF 1 (Figure.4.8). However, phage Mu tail tube gene ORF 2, transcribed after gene L, was not identified as homologous to phage SfV ORF118. ORF118 did not show significant homology to any database protein and yielded high levels of gap insertion when aligned with other phage tail protein. The gene may represent a dysfunctional ORF of host origin repeatedly transferred as the phage undergoes multiple lytic-lysogenic cycles, or a remnant tail gene of viral origin that has become vestigial in function since it was positioned between genes of putative tail protein identities (Figure 4.3). Predicted amino acid sequences for ORF89 and ORF116, downstream of ORF118, were homologous to essential tail proteins encoded by phage P2 gene E and phage P2 gene T, respectively [Nakayama, 1999 #203]. This clustering of bacteriophage SfV putative tail structural genes strongly suggested that ORF56, 498, 89, and 116 comprise another SfV gene module which codes for the phage tail assembly components.

There are several methods adapted in identifying the late protein functions in these bacteriophages. Protein functions were derived using approaches such as comparison with conserved domains, functional assessment experiments, and identification of essential structural features common among known proteins [Smith, 1999 #276][Kaneko,

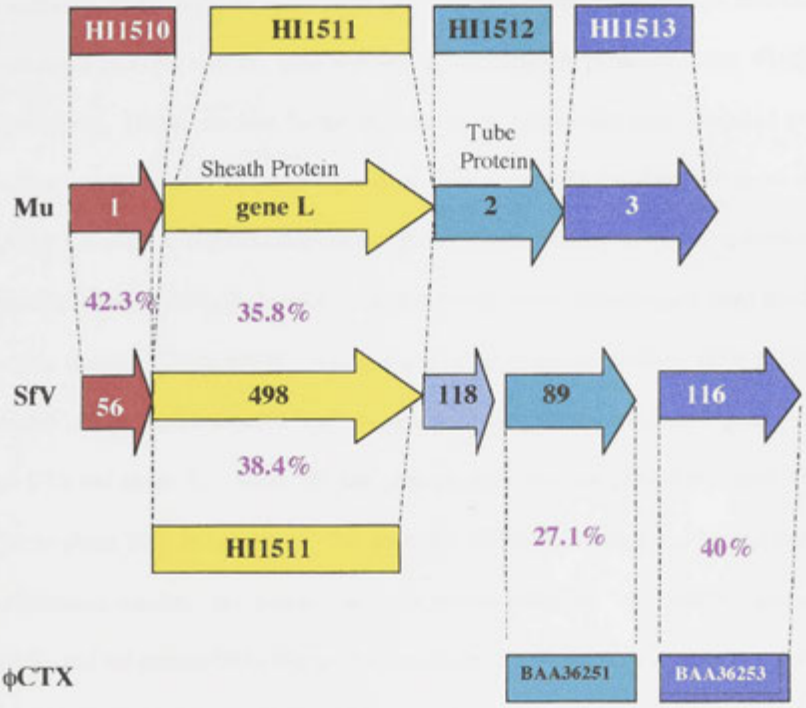


Figure 4.8. Schematic diagram illustrating the homology of bacteriophage SfV tail region proteins to the gene products of bacteriophage Mu and *Haemophilus influenzae* Mu-like prophage. Dashed lines indicate homology between two predicted proteins with corresponding similarity scores shown in pink text. Represented by coloured arrows are genes with related protein products. Phage Mu ORF1, sheath protein gpL and ORFs 2 and 3 has been submitted to GenBank sequence database with accession number AB000833. HI1510, HI1511, HI1512, HI1513 are proteins encoded in the Mu-like prophage of *Haemophilus influenzae* (Accession No: U32827), BAA36251 is the protein ID for phage ϕCTX similar to phage P2 gene E tail protein (Accession No. AB008550); and BAA36253 is for phage ϕCTX protein similar to phage P2 gene T tail protein (Accession No. AB008550).

1997 #280][Gilakjan, 1999 #278][Tremblay, 1999 #207][Desiere, 1998 #162][Kobayashi, 1998 #361][Garcia, 1997 #362]. Recent reports cited nonsequence alignment technique that used ORF length and isolelectric point (*pI*) to make a comparative analysis between the unknown proteins and the gene products of coliphage λ [Chandry, 1997 #316]. Tremblay *et al.*, (1999) provided further evidence of its utility when they compared the molecular weight and *pI* of *Streptococcus thermophilus* phage DT1 proteins with those of coliphage λ in order to support assignment of putative functions for 15 ORFs, a method inspired by researches on gene function prediction among Siphoviridae phages using the λ gene map [Lucchini, 1998 #318]. Using a similar approach [Tremblay, 1999 #207], phage SfV ORFs were compared with the molecular weights and isoelectric points of phage DT1 and phage λ. Some SfV late gene products showed properties which are similar to phage DT1 and phage λ. For example, ORF577 encoding for the putative large terminase subunit, the putative protease protein ORF200, the putative capsid ORF409, and tail protein ORFs 498 and 89 have molecular weights and isoelectric points which are comparable to those of phage DT1 and λ (Table 4.3).

The properties of the putative protease ORF200 (molecular weight, 22.7 and isoelectric point, 4.9) are similar to phage DT1 ORF7 scaffolding protein (MW 24.5, *pI* 5.0) and phage λ scaffolding protein Nu3 (MW 20.8, *pI* 4.5). Comparison between the protease and scaffolding protein was done based on their conserved gene location adjacent to the major head protein and in consideration of their analogous function in providing support

Table 4.3. Comparison between the molecular weight and isoelectric point of phage SfV predicted proteins with phage DT1 and coliphage λ protein products.

	SfV			DT1			λ		
Protein Function	ORF	MW	pI	ORF	MW	pI	Protein	MW	pI
Terminase large subunit	577	65.3	5.3	4	42.7	4.4	A	73.3	6.1
Scaffolding/Protease	200	22.7	4.9	7	24.5	5.0	Nu3	20.8	4.5
Major Head Protein	409	45.8*	5.1	8	32	5.2	E	38.2	5.3
Tail component	498	54.8	5.5	17	58.3	5.6	L	25.7	4.7
Tail component	89	10	5.2	14	13.5	4.7	G	15.6	5.1

*Unprocessed form of head protein (refer to Chapter 5)

during the formation of the head shell [Duda, 1995B #289][Duda, 1995C #301]. Likewise, the putative major head protein of SfV showed similar properties to its counterpart proteins in DT1 and λ, except for a higher molecular weight of 45.8 kDa for phage SfV major head protein (Table 4.3). However, several reports involving bacteriophage head assembly have experimentally confirmed the requirement for the major head protein to be cleaved in order to attain its mature form [Gilakjan, 1999 #278][Duda, 1995A #275][Smith, 1999 #276][Kaneko, 1998 #204]. If such a process occurs in phage SfV ORF409 protein, then we would expect a reduction in its molecular size possibly down to the size of phage DT1 major head protein (MW 32 kDa) or to that of phage λ (MW 38.2kDa) (Table 4.3). Therefore, functional studies involving experiments which would elucidate the occurrence of cleavage in ORF409 putative head protein should be investigated not only to ascertain its role as the capsid protein but also to determine other factors required for such a process to take place.

Variations in protein properties among bacteriophages SfV, DT1 and λ were also observed for the tail components suggesting a divergent gene pool source. Another example is ORF367 encoding the putative portal protein which showed a higher *pI* value of 9.4 compared to DT1 ORF 6 (*pI* 5.2) and phage λ protein B (*pI* 5.7), despite similar molecular weights for their protein products. Taken together, this method of comparative analysis between predicted protein properties could be useful in providing additional support in the assignment of putative gene functions for phage SfV ORFs.

However, it should be used with caution and should not be considered as a substitute but only as an adjunct to the more definitive protein functional analysis using experimentations.

Many SfV late proteins are homologous to the late proteins of other Siphoviridae bacteriophages which infect diverse bacterial hosts. The organisation of the late structural genes showed remarkable conservation (Figure. 4.6). It is possible that the genome of bacteriophages such as ϕC31, ϕPVL, D3, ϕ105, , DT1, Sfi19, Sfi21 and phage A2 was assembled from a common gene pool source [Hendrix, 1999 #166]. Another evidence showing linkage of phage SfV to the Siphoviridae family is its morphology (Figure 4.9). Phage SfV belongs to the group B category of Bradley's classification, possessing an isometric head and a long non-contractile tail which is also inherent among bacteriophages of the Siphoviridae family [Ackerman, 1987 #287][Bradley, 1967 #179]. Due to these genetic and structural similarities, a genetic pool lineage of phage SfV to bacteriophage members of the Siphoviridae family was observed.

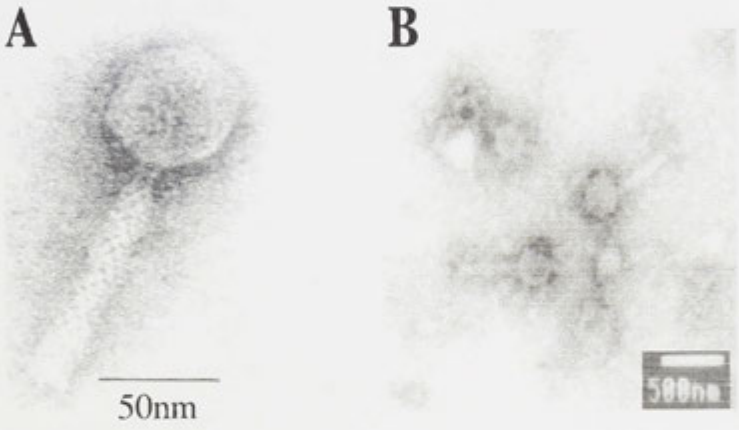


Figure 4.9. Transmission electron micrograph of bacteriophage SfV. A, phage SfV stained with 2% phosphotungstic acid (Huan et al., 1997C). B, phage SfV in cluster stained with 2% uranyl-acetate.

V RESULTS

CHAPTER 5

Post-translational Cleavage of the 45.8 kDa Major Prohead Capsid Subunit is Catalysed by *Shigella flexneri* SfV Bacteriophage-encoded Protease

5.1. Introduction

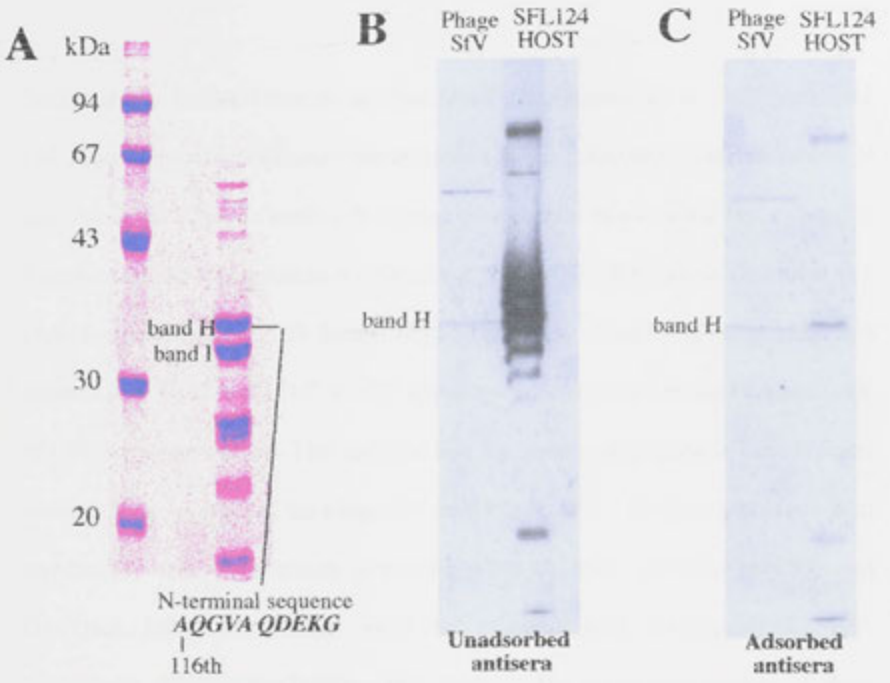
The 10.1 kb *Bam*HI fragment A has been shown to encode genes which are homologous to the late structural genes of many bacteriophages and prophages (Table 4.2). It was also noted that the organisation of bacteriophage SfV late genes have a similar gene arrangement to those found in other double-stranded bacteriophages such as D3, HK97, ϕC31, ϕPVL, ϕadh, T4 and λ (Figure 4.6). The order of transcription proceeds from the DNA packaging proteins, the terminases and portal protein to the head assembly genes comprised of the protease and the major head protein. SfV ORF200 and ORF409 products showed significant homology to the protease and capsid proteins of several bacteriophages including phage ϕC31, D3 and HK97. Recent comparative studies have revealed evidences supporting the similarity in the capsid structure and head assembly process of phage ϕC31 and D3 to phage HK97 [Gilakjan, 1999 #278][Smith, 1999 #276]. The capsid processing and maturation of HK97 had been studied extensively and was considered unique, since the process does not require the use of an independently encoded scaffold protein as support in the formation of a mature head shell but relied on

the assembly of capsid subunits that had undergone molecular modification and structural processing [Duda, 1995C #301]. In order to determine if phage SfV follows the same initial head processing step during head assembly involving cleavage of the capsid protein prior to crosslinking, we have conducted experiments that would show the capability of a functional ORF200 to induce reduction in the size of ORF409 upon overexpression. Our experiments demonstrated the processing of a 45 kDa SfV major capsid subunit after its translation. Our experiments have established that an intact protease gene *orf200* immediately upstream of the capsid gene *orf409*, is required to effectively cleave the unprocessed capsid subunit into a 32 kDa form of the prohead.

5.2. Results

5.2.1. *Protein Band H N-terminal Sequence was Located within ORF409*

In order to support and establish the identity of some SfV *orfs*, the phage structural proteins were separated on SDS-PAGE. The protein profile showed several bands which were labelled with alphabetical letters starting from the largest protein as band A (Figure 5.1A). Using Western immunoblotting procedure, the protein bands were blotted and hybridised with adsorbed and unadsorbed rabbit polyclonal antisera raised against the whole phage SfV particles which were propagated in the *S. flexneri* host, SFL124. Adsorption procedure using SFL124 acetone powder was performed in order to reduce non-phage SfV specific antibodies in the crude polyclonal antiserum preparation



D

```
  1 MKLHELKQKR  NTIATDMRAL  NEKIGDNAWT  EEQRTEWNKA  KSELEALDI
 51 IAREEELRRQ  DQAYIESNEE  EQRQNLDPEN  NSQQDEKRAQ  VFDKWMRHC
101 SELTSEERKA  LRELRAQGVA  QDEKGGYTVP  ETFLAKVVEK  MKSYGGIAS
151 AQILTTSDGR  TMEWATADGT  SEVGVLLGEN  EEAGEEDTDF  GMGSLGALI
201 TSKIIRVSNE  LLQDSAIDME  AYLARRIAER  IGRGEARYLI  QGTGAGTPI
251 PKGLAASVTG  TTQTAAANAV  KWQEILALKH  SIDPAYRRGP  KFRLAFNDI
301 LKLISEMEDG  QGRPLWLPDI  VGVAPASVLN  VPYVIDQEID  DIGAGKKFI
351 CGDFDRFIIR  RVRYMILKRL  VERYAEYDQT  GFLAFHRFDC  ILEDTSAII
401 LVGKGSVGG*
```

Figure 5.1. A, 12% SDS-PAGE gel of PEG precipitated SfV phage particles. B, Western blot of phage SfV protein hybridised with rabbit polyvalent unadsorbed antisera raised against whole phage SfV particles. Note the SFL124 host protein band of similar molecular mass as the protein band H of phage SfV. C, An identical Western blot as in B except that the rabbit polyvalent antisera used was adsorbed with SFL124 acetone powder. Note the reduced background and non-specific binding with crossreacting host protein on membrane using adsorbed antisera. D, N-terminal amino acid sequence of protein band H revealed the first residue to be internal to the predicted amino acid sequence of the ORF409 protein. The first ten N-terminal sequences are shown in bold print and italicized.

(Figure 5.1B and 5.1C). Anti-SfV antibodies detected band H at approximately 32 kDa level.

Six bands were extracted from the gel, transferred electrophoretically to PVDF membrane and pieces of blotted membrane were analysed using an automated protein sequencer at the John Curtin School of Medical Research, Biomolecular Resource Facility. Only the N-terminal amino acid sequence of protein band H, AQGVAQDEKG, corresponded to a protein encoded in the 10.1 kb *Bam*HI fragment A region. These N-terminal amino acid residues were found as the 116th to 125th amino acid residues of the predicted amino acids of ORF409 (Figure 5.1D). This indicated that the mature phage protein band H starts internal to the sequence of the phage SfV capsid ORF409. Similar results have been reported for other bacteriophages such as phage HK97, PVL, phi-105, phi-C31, and D3 [Duda, 1995A #275][Duda, 1995B #289][Kaneko, 1997 #280][Kobayashi, 1998 #361][Smith, 1999 #276][Gilakjan, 1999 #278]. The major head protein of these bacteriophages is proteolytically cleaved. This process is well characterised in HK97. The proteolytic cleavage of the capsid is conferred by a protease which is encoded immediately upstream of the capsid gene.

5.2.2. *Localisation and Amplification of the Putative Protease and Capsid Gene Insert*

Database homology searches have detected several bacteriophage-encoded capsid homologues as significantly similar to phage SfV ORF409. Further support to the

identity of this ORF was derived after the first ten amino acid residues of phage SfV SDS-PAGE protein band H detected by anti-SfV antibodies, were traced within the predicted amino acid sequence of ORF409. With ORF200 having a similar location as the protease genes in other bacteriophages, the function of ORF200 in processing ORF409 was investigated by assessing the fate of ORF409 protein in the presence and absence of a functional ORF200.

Phage SfV *orf200* (nt 6467-5865) and *orf409* (nt 5854-4625) are located at the upper 5' end of the 10.1 kb *Bam*HI fragment A, within the head protein cluster of the genome's late region (Figure 5.2A and B). Oligonucleotide primers were designed to capture the entire protease and capsid *orfs*. The forward primer 5'(5'-AAT GAA TTC ATC TGA CGG GGC TTT TAC-3'), designated as proteasestart, is located 207 nucleotides upstream of the *orf200* start codon (ATG) and the reverse primer 3'(5'-AAT GGA TCC GAC TAA TCA ACC ACC AAC-3'), called capsidend, flanks the complementary termination codon (TGA) of *orf409* (Figure 5.2C). A gene segment containing part of *orf200* and the entire *orf409* was also amplified. PCR primer proteasemiddle 5'(5'-AAT GAA TTC GCA CGC TGA ATC TCT CAG-3') was designed to anneal at the upper third portion of the protease DNA sequence amplifying the 3' half of the gene (Figure 5.2C). The 5' end of the primer was situated at exactly 232 nucleotides downstream from the start codon of *orf*200. When paired with primer capsidend, PCR yielded an insert comprising the 371 bp 3' half of *orf200* and the entire *orf409* DNA sequence. This insert with a truncated putative protease gene and a complete putative capsid gene

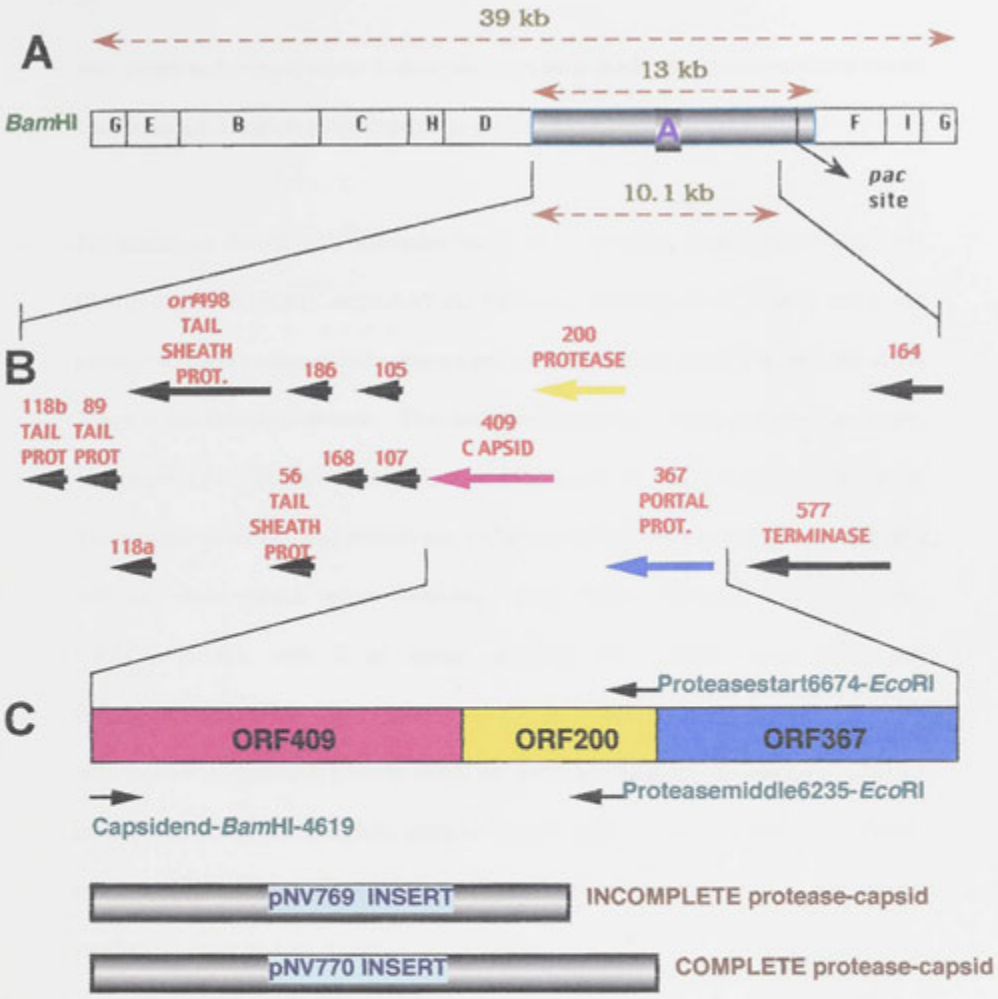


Figure 5.2. A, Physical map of bacteriophage SfV showing the whole genome's *Bam*HI fragments; B, Map of the sequenced 10.1 kb *Bam*HI fragment A with corresponding ORF designation and putative function, arrows indicate transcriptional direction; C, Map of the head assembly genes showing coordinates of primers used to amplify the complete protease-capsid insert and the incomplete protease-capsid insert, the primers were tagged with *Eco*RI or *Bam*HI linkers for directional cloning. The PCR products were cloned into the *Eco*RI/*Bam*HI sites of pT7-5 to create pNV769 and pNV770, respectively.

was cloned and overexpressed to compare its protein products with recombinant strains containing the complete putative protease gene.

For directional cloning with expression vector pT7-5 [Studier, 1990 #300][Tabor, 1985 #245][Tabor, 1987 #262], extra AAT-*Eco*RI linkers were tagged on to the 5' end of the proteasestart and proteasemiddle primers and an AAT-*Bam*HI linker was included at the 5' end of the capsidend primer. This facilitated directional cloning ensuring the correct transcriptional orientation of the insert with reference to the vector's promoter sequence. The proteasestart/capsidend primers successfully amplified 2055 bp flanking the putative protease and capsid genes including their likely ribosomal binding sites, GAGGGAAAAA with 3 nt spacer sequence from *orf200* start codon and GtAAAGGAAA with 4 nt spacer sequence from the *orf409* start codon. The proteasemiddle/capsidend primers amplified 1616 bp flanking the truncated putative protease gene and the complete putative capsid gene. Both inserts were cloned, overexpressed and the protein profiles compared with SDS-PAGE and immunoblotting results.

5.2.3. *ORF*200 *is Required to Cleave ORF*409 *Gene Product*

The PCR products were digested and ligated into pT7-5 and transformed into P4189. *E. coli* P4189 has a single copy of the T7 RNA polymerase gene under the control of the IPTG-inducible *lac*UV5 promoter. After transformation, the identity of the

recombinant plasmids were confirmed through restriction digest analysis. Plasmids were cut with *Sac*I and *Pst*I which have restriction sites located at the polycloning site as well as nucleotides 5591 and 5220 of the capsid gene, respectively. This generated three fragments indicating the transformation of strains with the desired recombinant plasmids. The resultant transformants were strain B876 with pNV769 containing the 2.4 kb pT7-5 vector and 2 kb complete protease-complete capsid insert and strains B877 and B878 with pNV770 containing the vector and the 1.6 kb truncated, incomplete protease-complete capsid insert (Figure 5.3). The protein products were overexpressed using 0.4mM IPTG to induce a log phase culture, and aliqouts were taken at 0, 60, and 180 minutes intervals prior to visualization in 12% SDS-PAGE stained with Coomassie blue.

Upon induction, the 45 kDa capsid protein accumulated over time in strains B877 and B878 without a pronounced increase in protein band intensity at the processed 32 kDa level (Figure 5. 4A, lanes 4-7 and Figure 5.4B, lanes 4-5). This indicated that the 45 kDa capsid proteins were not being cleaved because the incomplete protease was non-functional, thus the unprocessed capsid at this level accumulated. A separate 12% SDS-PAGE was prepared that clearly illustrated the increasing band intensity at the 45 kDa capsid level in B877 over time (Figure 5.5, lanes 6-9). The accumulation of the 45 kDa protein was not observed in B876, instead, a gradual increase in intensity of the 32 kDa capsid band was observed over time (Figure 5.5, lanes 1-4; Figure 5.4A, lanes 2-3; and Figure 5.4B, lane 3). Neither protein accumulated in the control strain B866 containing the vector only. This suggested that the complete protease insert *orf200* in B876

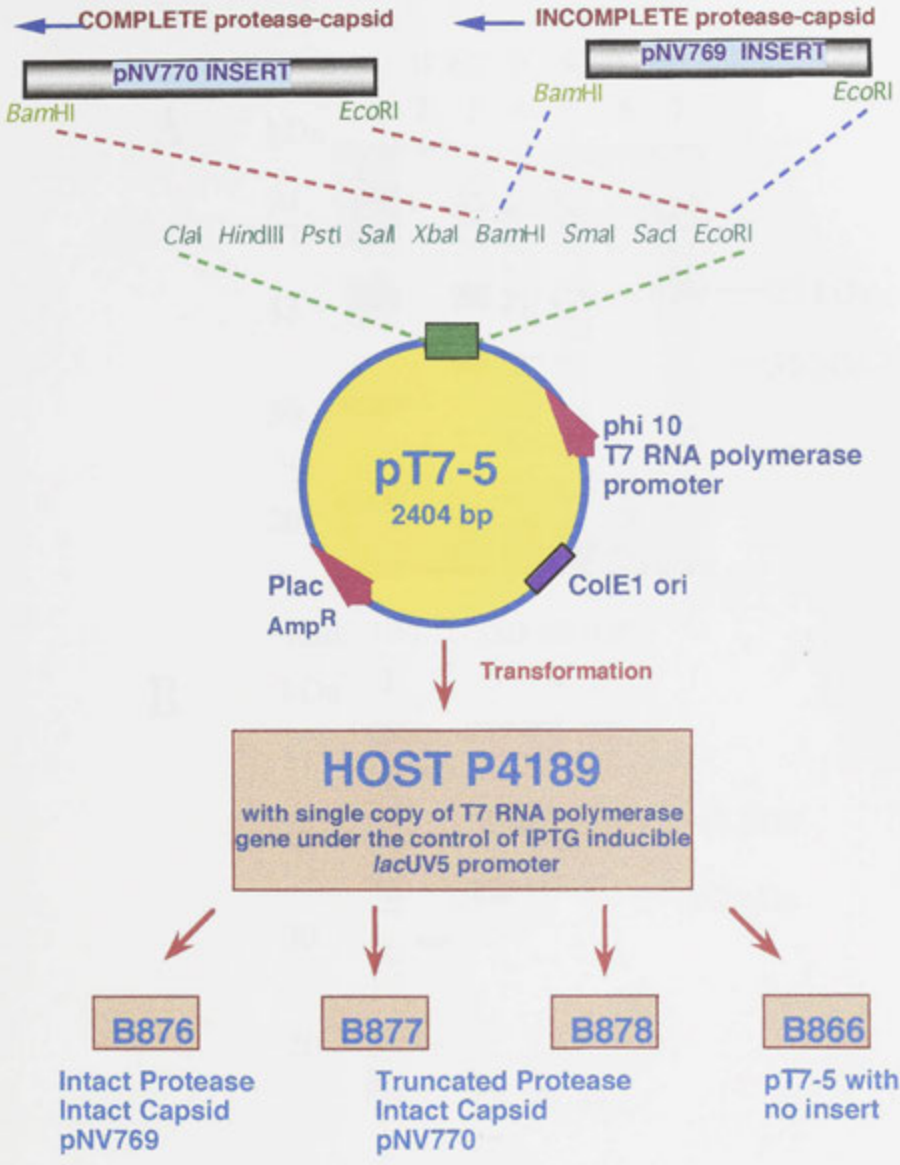


Figure 5.3. Schematic diagram of pNV769 and PNV770 cloning showing the designation of corresponding recombinant strains.

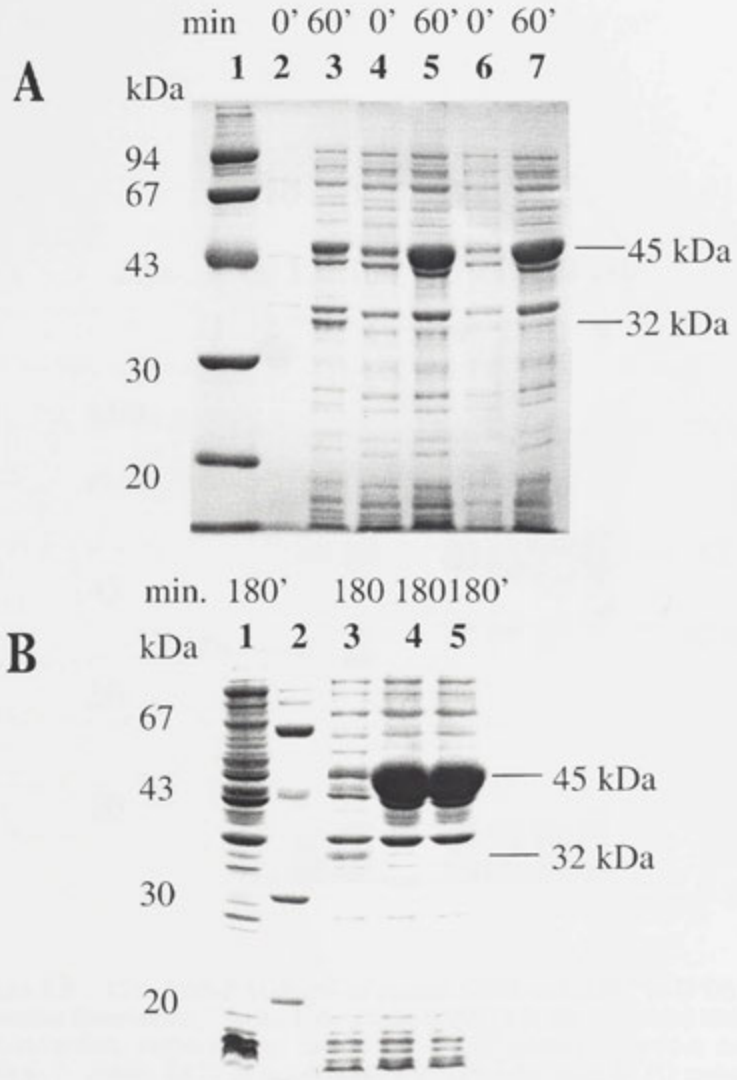


Figure 5.4. 12% SDS-PAGE gel of IPTG induced log phase growth of strains B866, B876, B877 and B878. lane 1: LMW protein calibration marker; lanes 2 and 3: strain B876 at 0 and 60 min. post-induction; lanes 4 and 5: strain B877 at 0 and 60 min. post-induction; lanes 6 and 7: strain B878 at 0 and 60 min. post-induction. B, lane 1: B866, control strain containing the vector only at 180 min. post-induction; lane 2: LMW protein calibration marker; lane 3: B876 at 180 min. post-induction; lane 4: B877 at 180 min. post-induction; and lane 5: B878 at 180 min. post-induction.

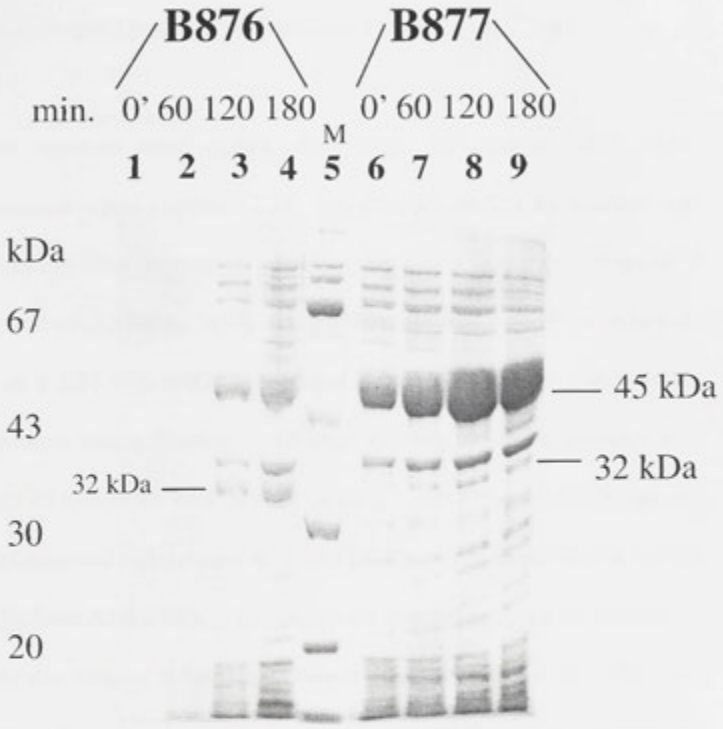


Figure 5.5. 12% SDS-PAGE gel of strains B876 and B877 in IPTG post-induction time series. lanes 1-4: strain B876 at 0, 60, 120, and 180 min. post-induction, respectively; lane 5: M, LMW protein calibration marker; lanes 6-7: strain B877 at 0, 60, 120, and 180 min. post-IPTG induction, respectively.

encoded for a functional protein that caused the cleavage of the 45 kDa capsid protein generating a 32 kDa protein digest product.

5.2.4. *Immunoblot Analysis Confirmed ORF409 Gene Product as the Capsid*

Polyclonal rabbit antiserum raised against whole phage SfV particles which earlier detected the processed capsid (section 5.2.1), was used to confirm the position and expression of capsid proteins from clones containing the ORF409 insert. Aliquots of phage SfV stock and strains B866, B876, and B877 at 120 min post-IPTG induction, were separated on a 12% SDS-PAGE then blotted onto to Hybond N+ nitrocellulose membrane for Western immunoblotting. Although the sera had been adsorbed with saturating amount of competitor host SFL124 protein, anti-*Shigella* SFL124-specific antibodies that crossreacted with proteins of *E. coli* P4189 control strain B866 were still present (Figure 5.6 lanes A4 and B8). Six bands were detected representing proteins of B866 which were also detected in B877 and B876 (Figure 5.6 lanes 5, 6, 8). This was expected as all three strains are derivative strains of *E. coli* P4189. Two major phage SfV structural proteins were detected, one representing the deduced 32 kDa processed capsid band, the only form composing the mature phage head, and a ~55 kDa protein band suspected as representing the putative tail protein sheath ORF498 with a 54.8 kDa molecular weight (Figure 5.6 lanes A3 and B7).

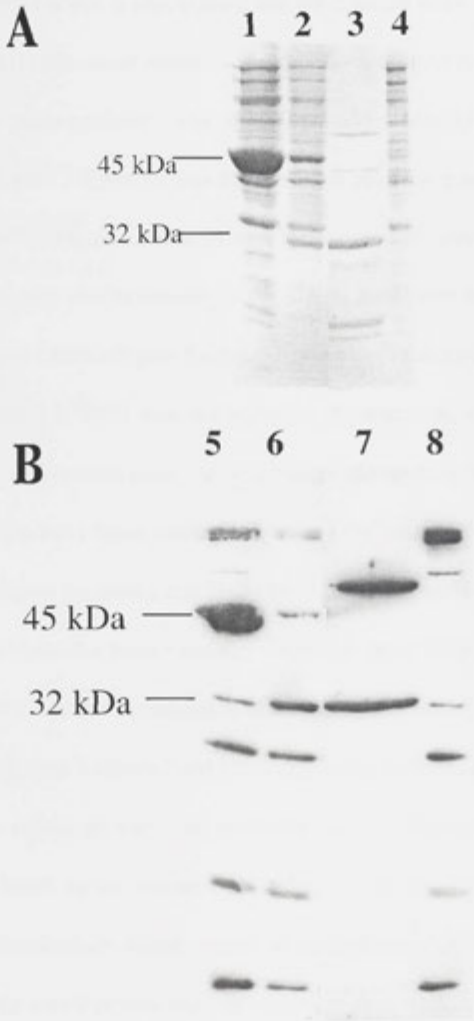


Figure 5.6. Western immunoblot of strains B866, B876, B877, and phage SfV protein. A, lanes 1-2: 12% SDS-PAGE gel of strains B877 and B876 respectively, at 120 min post-IPTG induction; lane 3: phage SfV; and lane 4: strain B866 (control strain with pT7-5 vector). B, Western immunoblot using rabbit anti-SfV polyclonal antisera, lanes 5-6: strains B877 and B876, respectively; lane 7: phage SfV; and lane 8: B866 control strain.

The deduced 45 kDa capsid which is unprocessed was not detected in the phage SfV lane, suggesting that the 32 kDa processed capsid is the only form of head protein utilised in the formation of mature phage particles. This 32 kDa processed form of the capsid was also detected in B876 (Figure 5.6 lane 6) with the complete protease gene insert but was not as pronounced in B877 (Figure 5.6 lane 5) with the incomplete protease gene insert. Autoradiographic signals with similar intensity at the 32 kDa level were detected for both B877 and the control strain B866 (Figure 5.6 lanes 5 and 8). This signifies that the 32 kDa protein band visualised in B877 does not represent the processed capsid but an *E. coli* host protein of similar molecular mass, an observation clearly illustrated upon direct comparison of the 32 kDa band signal intensities between the processed capsid in B876 and phage SfV capsid (Figure 5.6 lanes 6 and 7) and the E. coli host proteins in B877 and the control strain B866 (Figure 5.6 lanes 5 and 8). Another proof of the absence of the processed capsid in B877 was the accumulation of the 45 kDa unprocessed form which was as evident in B876 (Figure 5.6 lanes 5 and 6). Considering the homology of ORF409 to other bacteriophage capsid protein, its molecular size, N-terminal amino acid sequence and identity based on the immunoblot results, our functional experiments provided additional evidence which would support the designation of *orf409* as the phage SfV gene encoding for the capsid protein that gets cleaved in the presence of the complete *orf200*-encoded protease protein.

5.3. Discussion

In this study we have identified the genes encoding the protease and capsid protein of the serotype-converting *S. flexneri* bacteriophage SfV. Alanine was the first amino acid residue of the mature processed capsid band H whose first 10 amino acid residues were determined (Figure 5.1). It was found to be the 116th amino acid residue of the mature capsid protein sequence. This indicated that the start of the mature protein sequence is internal to ORF409 and suggested that processing of the capsid protein may be involved.

Functional studies of the ORF200 indicated its role as the protease protein required in the processing of the capsid through cleavage. SfV ORF200 has significant homology to putative proteases of bacteriophages such as phage A2 of *Lactobacillus casei* ORF242 gene product which is similar to ClpP-like (protease Ti) [Chung, 1996 #508] proteins in the chromosome of bacteria and eukaryotes such as *Caenorhabditis elegans* (Q27539), *Yersenia enterocolitica* (Q60170), *Homo sapiens* (Q16740), *Paracoccus denitrificans* (P54414), *E. coli* (P19245) and *Haemophilus influenzae* (P43867) [Altermann, 1999 #206]. ClpP is a proteolytic subunit of the Clp complex which exhibits potent proteinase activity when combined with the ClpA, C, X subunits of the Clp enzyme system [Wawrzynow, 1996 #359]. The Serine108 and Histidine85 residues of phage ϕadh ORF242 protease with 22 interpeptide residues were found to correspond to *E. coli* ClpP protein residues Serine111 and Histidine136 with 24 interpeptide residues. These important amino acid components were reported to form part of the enzyme's catalytic

site [Maurizi, 1990 #329]. Similarly, these sites were found in phage SfV ORF200 residues Serine 61 and Histidine 85 and residues Serine 45 and Histidine 69 of *L. casei* phage A2 protease protein. It is also interesting to note that for both bacteriophages, the putative serine and histidine catalytic sites are separated by 23 interpeptide residues (Figure. 5.7). Again, we can observe the similarity in protein properties suggestive of the probable protease characteristics of SfV ORF200 product.

The diversity of capsid assembly processes adapted by double-stranded bacteriophages has been well-studied [Casjens, 1988 #266][Casjens, 1992 #265][Hendrix, 1994 #263][Hendrix, 1999 #166][King, 1974 #264]. While the basic mechanism involves a common transition of structural components from the detailed arrangement of subunits through the formation of a mature icosahedral head, the steps and intermediate molecules involved in the capsid development process vary. For example, phage P22 requires multiple copies of a cigar-shaped scaffold protein that interacts with the viral coat subunit to ensure the proper shell configuration [Fuller, 1982 #267]. This protein is absent in the developed phage particles and only found inside the protein shell coat of immature proheads, signifying its transient role in the development of the head [Murialdo, 1978 #268]. Other phages encoding the scaffold protein are phage T4 genes 22 and 67 [Black, 1974 #269]; phage λ gpNU3 [Earnshaw, 1980 #270][Black, 1978 #271][Catalano, 1995 #272]; phage P2 gene P2N and phage P4 *sid* gene [Marvik, 1994 #273][Marvik, 1995 #274]. In the pseudo T-even bacteriophage RB49 [Monod, 1997 #284] and the well-studied coliphage T4, a substantial portion removed from the amino terminus of its initial

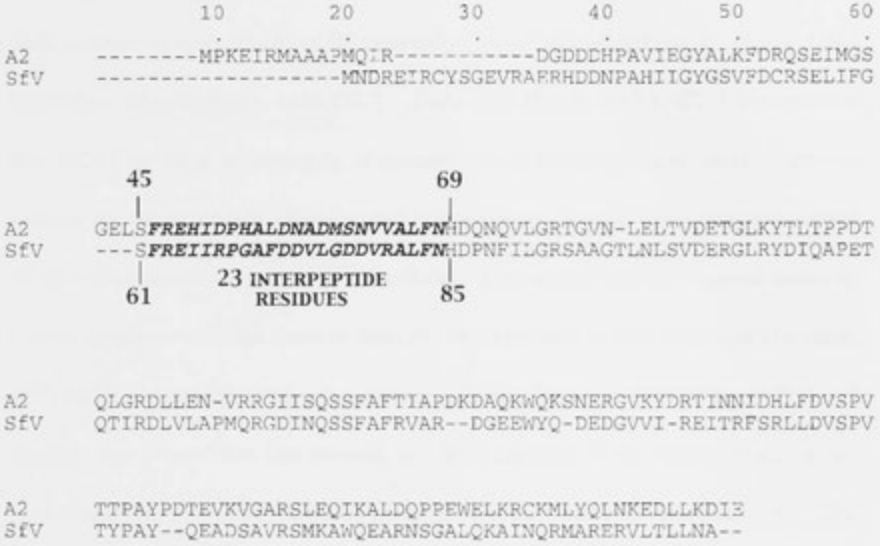


Figure 5.7. Protein sequence alignment of *Shigella flexneri* phage SfV ORF200 putative protease and *Lactobacillus casei* phage A2 protease (Accession No. AJ251790) showing the protease catalytic sites composed of serine and histidine residues which are numbered for both bacteriophages. There are 23 amino acid residues separating the serine and histidine residues indicated by italicised text. Alignment was performed using the Eclustalw program in WAG.

capsid subunit through cleavage mediated by a phage-encoded protease enzyme [Dickson, 1970 #285][Hosoda, 1970 #286]. The apparent contrast in the head assembly pathway between phage HK97 and phage T4 is that the latter use scaffold proteins to support shell structure integrity and these are eventually cleaved into several pieces prior to DNA headfilling [Mesyanzhinov, 1990 #327]. Duda *et al.* [Duda, 1995A #275] demonstrated that HK97 capsid is an assembly of pentameric and hexameric head protein subunits without the use of scaffold proteins, and referred to the cleaved amino terminus portion of the capsid protein as a substitute scaffold. Conversely in P22, capsid assembly utilises monomer units that serve as assembly intermediates in shell formation [Prevelige, 1993 #279]. Although there are variations between the many double-stranded phages studied, one central fact that prevails is, that subunits of the mature head are not produced as pre-fabricated molecules ready to be incorporated as an intact shell. The components of the head need to undergo proteolytic digestion, structural changes and molecular interaction prior to the final assembly (Figure 5.8).

About a third of ORF409 protein was released after proteolytic cleavage, a mechanism not unique to phage SfV [Casjens, 1988 #266]. For example, both Vibriophage KVP20 [Matsuzaki, 1998A #281] and KVP40 [Matsuzaki, 1992 #283][Matsuzaki, 1998B #282] capsid had 62 N-terminal amino acid residues of pro-MCP (Major Capsid Protein) released upon maturation to MCP. These similarities were also observed in the head morphogenesis of bacteriophage members of the Siphoviridae family, possessing long, non-contractile tail [Ackermann, 1987 #287][Ackermann, 1997 #288]. Phage SfV capsid

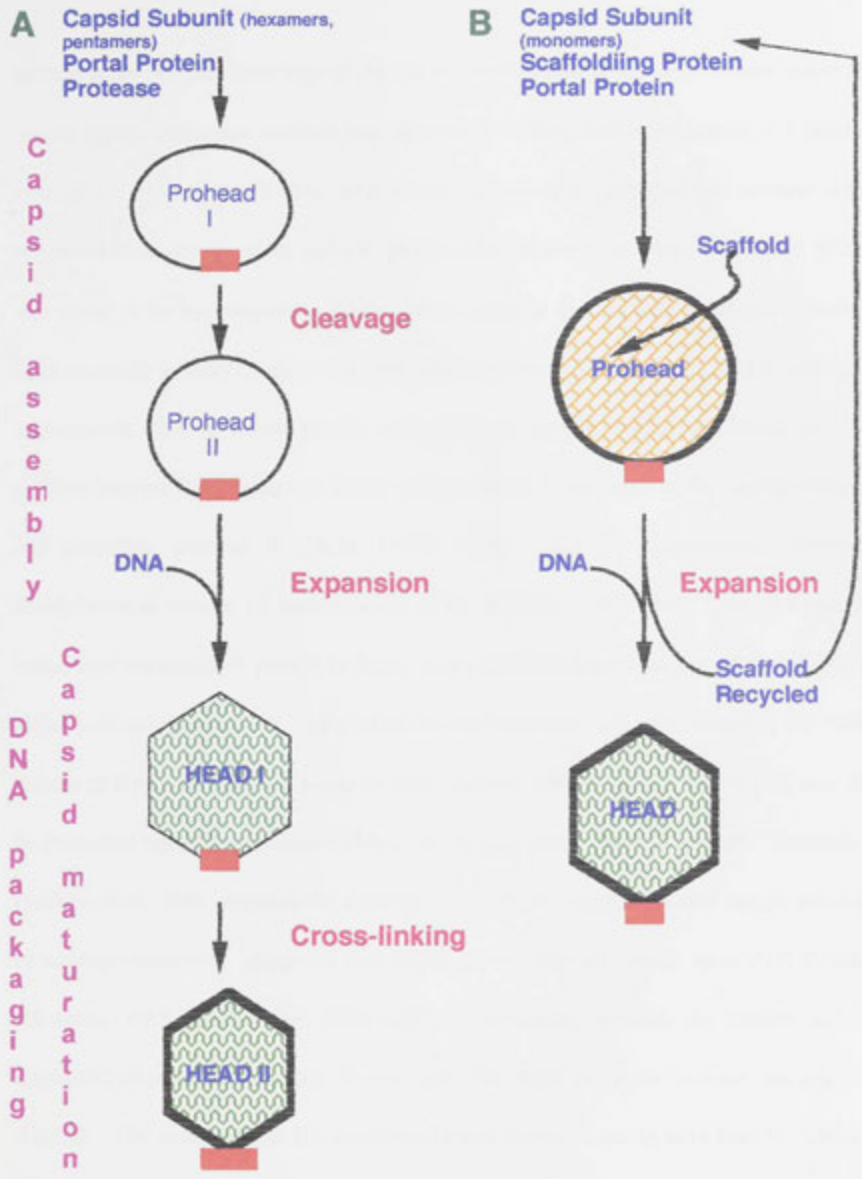


Figure 5.8. Schematic representation of two types of bacteriophage head assembly process. A, Pathway involves cleavage and structural crosslinking in the absence of a scaffold protein. B, Pathway requiring scaffold protein. (Adapted from Hendrix *et al.*, 1994)

protein shared closest homology (31% ID) to *Streptomyces* phage ϕC31 whose putative capsid (gp36) translation product was discovered to have had an additional 111 amino acids at the N-terminus [Smith, 1999 #276]. A putative gp35 protease product was suggested to have cleaved the gp36 to produce the mature form. Gp35 of phage ϕC31 was found to be homologous to HK97 gene 4 which is the prohead protease. During head assembly in HK97, the initial shell structure prohead I composed of 420 units of unprocessed 42 kDa capsid protein and 12 portal protein units, is cleaved by the protease enzyme eliminating 102 amino acids from the 5' terminus of the capsid subunit and generating prohead II [Duda, 1995B #289]. In the mitomycin-C induced *Staphylococcus aureus* V8 bacteriophage ϕPVL [Kaneko, 1997 #280], the N-terminal amino acid sequences of protein B (major head) are located between the 117^{th} and 146^{th} amino acid residue of ORF 7. Internal amino acid sequences were also found at the 11th residue of the ORF4 putative portal protein [Kaneko, 1998 #204]. The 34 kDa size of its processed capsid approximates that of its 32 kDa phage SfV counterpart. Recently, Gilakjan *et al.*, 1999, reported the discovery of a 186 kDa high molecular weight protein in serotype-converting phage D3 that failed to enter the gel matrix upon SDS-PAGE [Gertman, 1987 #290][Miller, 1974 #291]. Sequencing revealed the protein as an aggregated cross-linked hexamer derived from the head monomer subunit starting at Ala112. The mature phage D3 head protein was demonstrated to have lost 111 amino acid residues from its N-terminus upon proteolytic cleavage. These events were analogous to our phage SfV in which N-terminal sequencing of the mature capsid yielded an Ala116 as its first residue with a total 115 amino acids cleaved off from the precursor

protein (Figure 5.9). Head morphogenesis in phage HK97 and D3 is also similar, both having a 31 kDa processed head unit [Gilakjan, 1999 #278]. From the discussion presented so far, it can be deduced that the processing of major head proteins is well-established and is universal among dsDNA bacteriophages.

Other similarities between phage SfV and HK97 head morphogenesis were also observed. For example, phage SfV has a slightly larger 45.8 kDa unprocessed head subunit compared to its 42 kDa prohead I precursor counterpart in phage HK97. It is also possible that phage SfV primary cleavage site located between Arg115 and Ala116 could have similar properties as the HK97 cleavage site. Based on experimental results, Duda and his colleagues [Duda, 1995B #289] located the phage HK97 head protein cleavage site between Lys103 and Ser104 and had proven that the site sequence requirement is not absolute since there was only a modest decrease in cleavage when lysine was changed to leucine. This suggests the role of substrate conformation in the specificity of cleavage and the possibility of having intermediate cleavage sites.

Adding support to the designation of phage SfV *orfs 200*, and *409* as the protease and capsid proteins, respectively, is the identical organisation of the these head assembly genes among representative bacteriophages belonging mostly to the Siphoviridae family (Figure 4.6). Taken together, we have proved that cleavage of the phage SfV 45 kDa ORF409 capsid only occurs when the complete and intact protease gene *orf200* is present and that the phage encoded proteins required for head subunit processing are the capsid

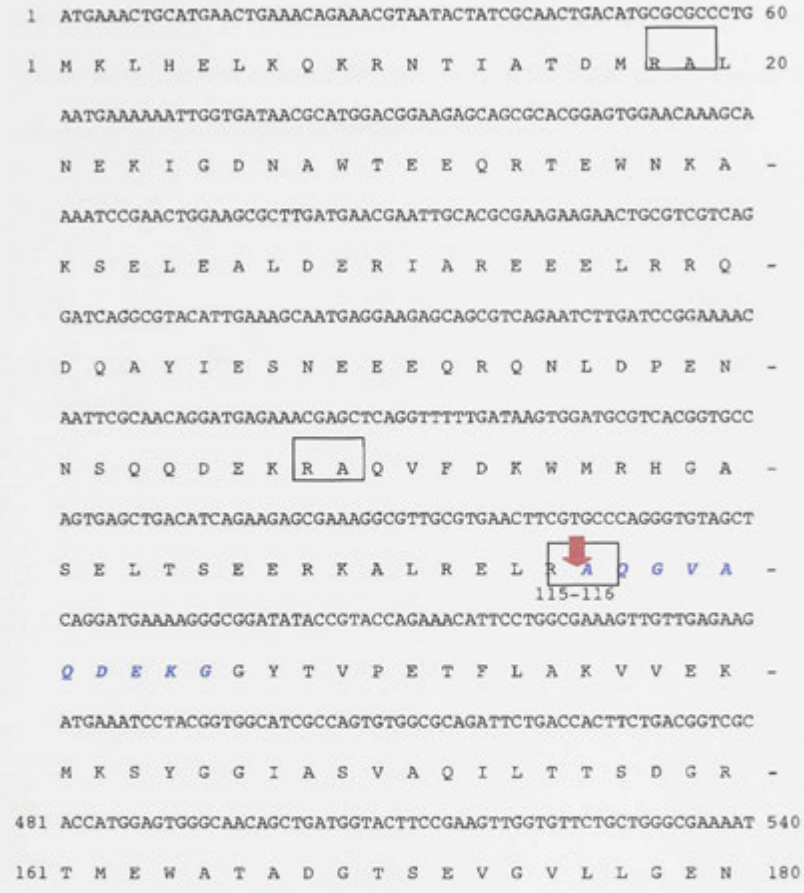


Figure 5.9. Partial sequence of bacteriophage SfV *orf*409 encoding the capsid protein. The primary cleavage site containing the arginine-alanine tandem is boxed and indicated by a red arrow. Also boxed are other putative cleavage sites upstream of the primary site. The first ten N-terminal amino acid sequence of protein band H are written in italics and blue coloured text.

substrate and the protease. Finally, in considering the remarkable similarity in capsid assembly and processing not only in physical and molecular properties of the genes and their products, we have observed that bacteriophage SfV shares an evolutionary linkage to the bacteriophages belonging to the Siphoviridae family.

VI RESULTS

CHAPTER 6

The Role of *Shigella flexneri* Bacteriophage SfV Genome in Host Virulence

6.1. Introduction

Majority of bacteriophage SfV genome has been characterised and further studies on the protein products should be performed in order to determine their function. Several studies on the role of bacteriophage in pathogenesis have been reported recently. The aim of this project was to determine if phage SfV plays a role in *S. flexneri* pathogenesis, a study based on recent reports on the role of bacteriophages in bacterial host pathogenesis [Waldor, 1998 #537][Miao, 1999 #524].

The initial step in the pathogenesis of shigellosis is the invasion of intestinal epithelium [LaBrec, 1964 #331]. Penetration of epithelial cells by *Shigella* involves a three main stage process. Initially, bacterial entry occurs by induced phagocytosis. This is characterised by the bacterium being engulfed by long pseudopods [Sansonetti, 1992 #24]. Shigella evades fusion with the phagosome enabling them to escape from the phagocytic vacuole [Sansonetti, 1986 #339]. The release of bacteria is followed by intracellular organelle-like movement (olm) along stress fibers [Vasselon, 1991 #340] or along the actin filament ring at the perijunction area [Vasselon, 1992 #341]. Passage of bacteria into adjacent cells is mediated by a bundle of actin filaments which push the

bacteria and form a protrusion into the other cells. This allows intercellular transfer of bacteria to occur without bacterial exposure to the extracellular fluid, a movement designated as intra- and intercellular spread [Ogawa, 1968 #342]

Animal models have been used to study the virulence and infection process of *Shigella*. For example, the Sereny test is performed in guinea pigs and involves testing the capacity of *Shigella* to cause keratoconjunctivitis via invasion of the corneal epithelium [Sereny, 1957 #109]. Ligated rabbit ileal loops have also been utilised. Injection of *Shigella* into the loops results in fluid secretion into the gut lumen and ulcer formation if the bacterium is invasive [Gots, 1974 #332][Wassef, 1989 #333]. Pathogenic mechanisms of *Shigella* were also observed in orally challenged Rhesus monkeys [Takuchi, 1968 #334]. However, most knowledge regarding *Shigella* invasiveness has been derived from mammalian tissue culture experiments [Gerber, 1961 #335][Oaks, 1985 #336][Sasakawa, 1988 #337].

In addition to invasion, it would be useful to identify genes in bacteriophage SfV that could convey important properties to its *Shigella* host upon lysogeny. In this study, we investigated how the subtraction or addition of phage SfV, or the addition of segments of the SfV genome affect the capacity of *S. flexneri* to invade mammalian HeLa cells.

6.2. Results

6.2.1. *Preparation of SFL124 and SFL1 lysogen strains*

Bacteriophage SfV particles were serially titrated ten-fold and propagated on wild type host SFL1339 (SFL1 with Congo red binding ability, Pcr^{+}) and SFL124 (attenuated derivative of SFL1) expressing group antigen 3,4. SFL124 lysogen designated SFL1333 and SFL1339 lysogen designated SFL1338 were identified using various genetic and serological procedures. To ensure phage SfV had integrated into the host chromosome, it was induced by UV irradiation and isolated. The genomic DNA was extracted and the characteristic DNA fingerprint band pattern typical of phage SfV was visualised following restriction digest with *Bam*HI and *Eco*RI (Figure 3.1A). Phage sensitivity test was also used to confirm that the lysogen strain was protected against phage SfV superinfection (Table 6.1A). For the lysogen strains SFL1333 and SFL1338, growth was evident on an area in the bacterial lawn where 20 µl of phage SfV suspension was dropped. This indicated that the lysogen strains have acquired immunity from phage SfV infection which was not observed in SFL124 and SFL1339.

The lysogen strains strongly reacted with monovalent antisera containing agglutinins specific to *S. flexneri* type antigen V and group antigen 3,4. This suggested serotype Y (SFL1339 and SFL124) conversion to serotype V (Denka Seiken Co., Ltd., Japan), and denoted successful integration of the phage genome into the host chromosome. The

Table 6.1A Analysis of SFL124, SFL1339 and their derivative lysogen strains

Strains	Phage SfV Sensitivity	UV induction of SfV	PCR Amplification of gtrV	Serogroup 3,4 Agglutination	Serotype V Agglutination	Congo Red Binding
SFL124	S	NI	NA	+	-	+
EW595/52	R	I	A	+	+	+
SFL1333	R	I	A	+	+	+
SFL1339	S	NI	NA	-	-	+
SFL1338	R	I	A	+	+	+

Table 6.1B Analysis of putative cured derivatives of EW595/52

Strains	Phage SfV Sensitivity	PCR Amplification of gtrV	Serogroup 3,4 Agglutination	Serotype V Agglutination	Congo Red Binding
SFL124	S	NA	+	-	+
EW595/52	R	A	+	+	-
SFL1336	R	A	+	+	+
SFL1337(I12)	S	NA	+	-	+
E1	R	A	+	+	+
F2	R	A	+	+	+
F3	R	A	+	+	+
G4	R	A	+	+	+
G5	R	A	+	+	+
H6	R	A	+	+	+
H7	R	A	+	+	+
I8	R	A	+	+	+
I9	R	A	+	+	+
I10	R	A	+	+	+
I11	R	A	+	+	+
J13	R	A	+	+	+

S= sensitive; R=resistant; I=induced; NI= not induced; A=amplified; NA=not amplified

identity of the lysogen strains was confirmed when the 1.3 kb glucosyltransferase gene (*gtrV*) was amplified from representative colonies using the primer pair created by D. Bastin: DB2-5'-AGAGAATTCCTACCATTCAACATTAAGGCT-3' and DB3-5'-AGAGGATCCACAT CGCCCAAAATACATCAT. The presence of the large invasion plasmid associated with *S. flexneri* virulence was endorsed by the lysogen's ability to bind Congo red and grow in Congo red supplemented Trypticase Soy agar (Per^-)[Sansonetti, 1982 #349]. Strain SFL1338 bound Congo red and satisfied all other criteria that confirmed it was an SfV lysogen. The virulence characteristics of SFL1338 was compared to its parental strain SFL1339 in the cell invasion assay in order to assess the contribution of the phage genome to its invasive capability (section 6.2.4).

6.2.2. *Preparation of cured wild-type EW595/52 Type V strain*

Two methods were used to cure *S. flexneri* EW595/52 (Type V, SfV lysogen). Initially, the mitomycin C method of Stanley *et al.*, [Stanley, 1999 #248] was used. The strain was plated onto LB medium containing varied concentrations of mitomycin C. Thirty surviving colonies were screened by phage sensitivity testing but all were found to be resistant to phage SfV. This indicates that the prophage was still present and providing immunity to superinfection. The second method used involved inducing the phage out of the bacterial host chromosome by overexpression of the excisionase gene (*xis*) [Leffers, 1998 #246]. Phage SfV *xis* gene was obtained from strain B377 with pNV324 containing the 7.3 kb *Bam*HI fragment B in pUC19 (Genbank Accession No. SFU82619)[Huan,

1997C #230]. The recombinant plasmid was digested with *Hin*dIII, which cleaved at nt 8590 upstream of the 5' end of *xis*, and at nt 8761 in the pUC19 multiple cloning site (MCS). The MCS is adjacent to the *Bam*HI nt 9538 site which is located at the 3' end of the *xis* gene (Figure 6.1). The 1 kb *xis*-containing *Hin*dIII fragment was ligated into the pTrc 99A expression vector which has a strong IPTG-inducible *trc* promoter upstream of the MCS. The ligation mix was then transformed into rubidium chloride competent JM109 cells. Plasmids were obtained from several transformants and digested with both *Hin*dIII and *Bam*HI. The plasmid DNA from strain B885, designated pNV775, yielded the 1 kb *xis* insert and the 4.1 kb pTrc 99A vector, indicative of having the *xis* gene in the correct orientation with respect to the vector promoter. Plasmid pNV775 was then transformed into electrocompetent wild-type EW595/52 (Type V) and designated as SFL1336 (Figure 6.1). The Xis protein was overproduced in SFL1336 through IPTG induction at two timepoints of 60 minutes and 180 minutes, and 784 of the resultant colonies were inoculated onto duplicate plates for colony blotting hybridisation (Figure 6.2). Each plate included a positive control strain SFL1334 (EW595/52 with pTrc 99A) which was uninduced , and a negative control strain, SFL1266 (SFL124 containing pUC18).

Alkali blotted DNA from test colonies was probed with radioactive ^{32}P-labelled glucosyltransferase gene, *gtrV*. This probe was amplified from the plasmid pNV323 which contains serotype-conversion cluster, $gtrA_{(V)}$, $gtrB_{(V)}$, and *gtrV*. Thirteen colonies that did not hybridise to the *gtr*V probe were screened using the phage

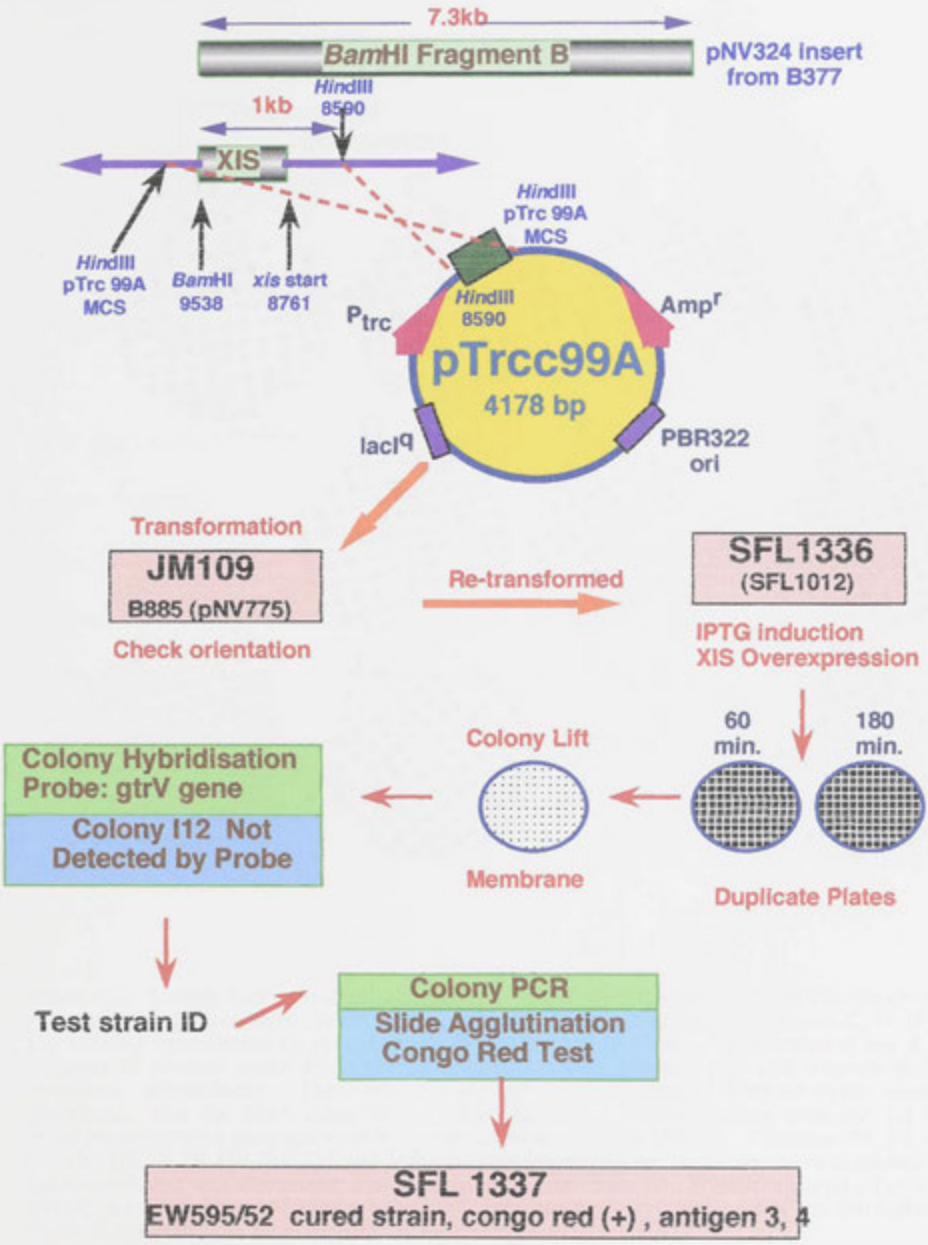


Figure 6.1. Flow diagram showing the steps and methods used in the preparation of EW595/52 cured strain

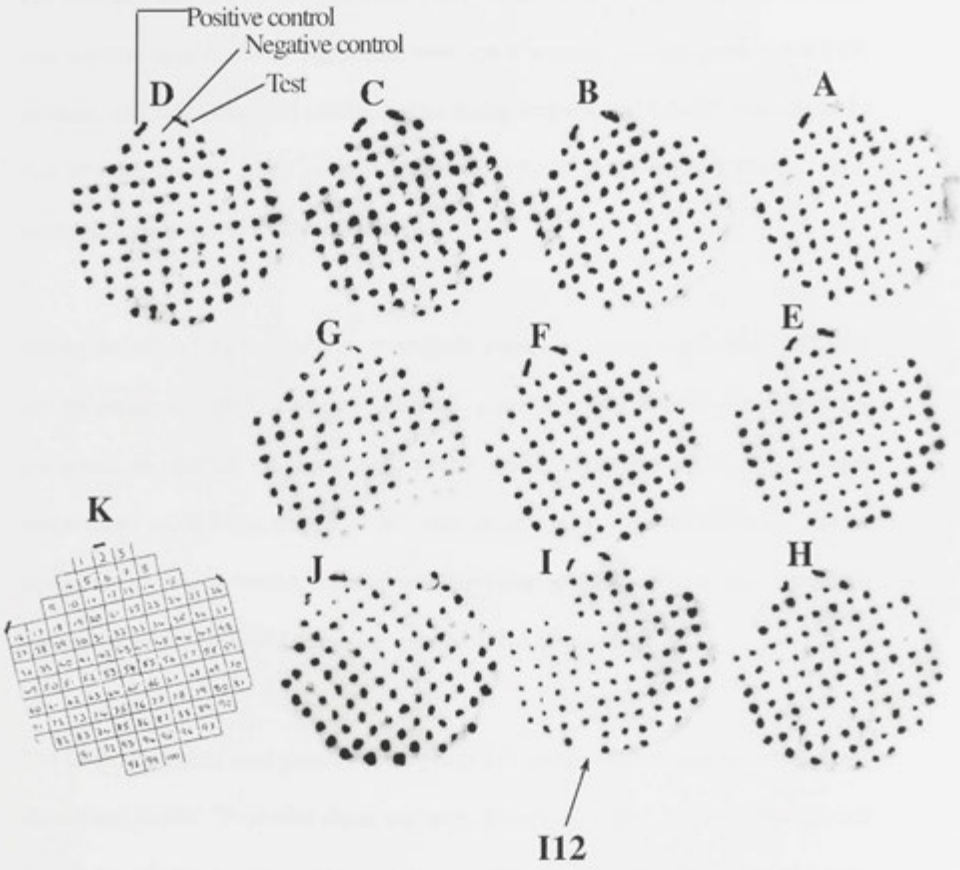


Figure 6.2. Colony blot hybridisation experiment of SFL1336 colonies after IPTG induction. The colonies were grown overnight on duplicate plates after 60 minutes (membranes C, D, E, F) and 180 minutes (membranes G, H, I, J) post-IPTG induction (0.4mM). Membranes A and B contain colonies of control strain SFL1334 inoculated after 60 minutes and 180 minutes post-IPTG induction, respectively. The colonies were transblotted (coloony lifts) on nylon membranes, denatured, and the DNA subjected to alkali fixation, before probing with the 1.3 kb SfV serotype-conversion gene (*gtrV*) PCR amplified from pNV323 (B376). Colonies E1, F2, F3, G4, G5, H6, H7, I8, I9, I10, I11, I12 and J13 were not detected by the probe and were examined further. Each membrane was inoculated with a positive control strain (\), EW595/52 with pTrc99A, and a negative control strain SFL1266 which is SFL124 with pUC18 (.), followed by the first test strain (/). Figure K represents the template pattern used during colony inoculation.

sensitivity test, colony PCR and slide agglutination test to ensure the phage SfV genome had been excised from EW595/52 (Table 6.1B). Only colony 112 designated SFL1337 was sensitive to SfV, did not agglutinate with type V antisera, did not yield a *gtr*V PCR product, and was Congo red positive. Our curing frequency of $1.3x10^{-3}$ was similar to that of phage Lambda cI857 lysogens which was $3.8x10^{-3}$ [Leffers, 1998 #246]. Test results suggested that SFL1337 was cured of SfV.

Having determined the loss of type V antigenic expression in the cured strain SFL1337 and the presence of *gtrV* in lysogen SFL1338, a Southern hybridisation experiment was performed to confirm the integration of the whole SfV genome into the lysogen chromosome and the loss thereof in the cured strain. Chromosomal DNA of relevant *Shigella* strains was extracted, digested with *Bam*HI and *Eco*RV, blotted onto membrane and probed with ^{32}P-labelled phage SfV genomic DNA (Figure 6.3).

The DNA fingerprint band patterns of the phage SfV genome was visualised in the phage *Bam*HI and *Eco*RV ^{32}P-labelled digest fragments (Figure 6.3, lanes 2 and 3) that showed similar *Bam*HI band pattern observed in an earlier gel electrophoresis (Figure 3.1A). These *Bam*HI and *Eco*RV SfV restriction band patterns were also seen in the SfV lysogen EW595/52, its transformed clone SFL1336 containing pNV775 with the *xis* insert, SfV lysogens SFL1333 and SFL1338 (Figure 6.3 lanes 4-7, 12-13, and 16-17). Conversely, the phage DNA restiction pattern was not evident in the cured strain SFL1337 and the serotype Y strains SFL124 and SFL1. (Figure 6.3, lanes 8-9, 10-11, and 14-15). Strains

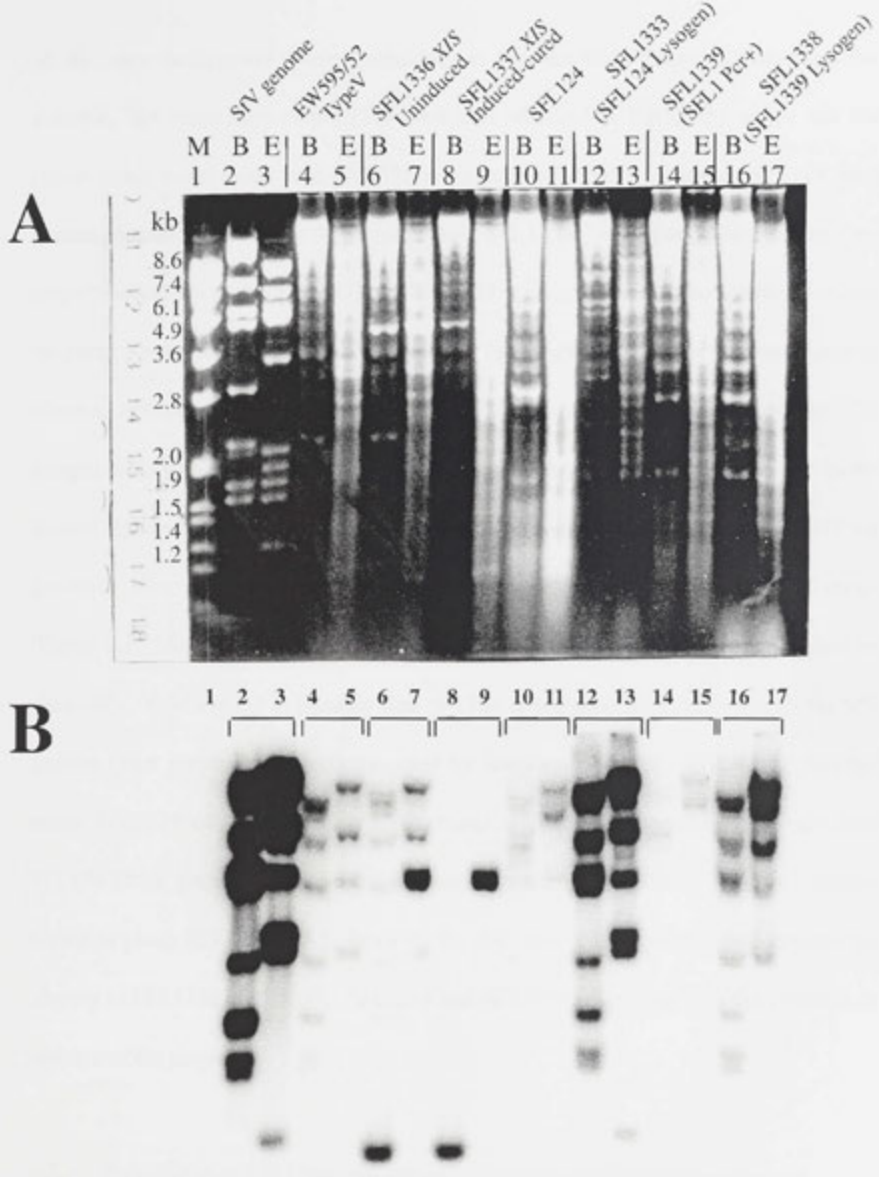


Figure 6.3. Southern hybridisation of chromosomal DNA of prepared lysogen, cured and SFL1339 strains probed with the ^{32}P-labelled SfV genome. A. 0.7% agarose gel electrophoresis run at 12V for 17 hours and blotted to Hybond N+ nucleic acid transfer membrane. B. Autoradiograph of A. M, SPPI cut with *Eco*RI; B, *Bam*HI digests; E, *Eco*RV digests.

of the same background showed similarity in the restriction bands produced. For example, the bands seen in SFL1336 were also observed in EW595/52 which was the parent strain transformed with pNV775 to produce SFL1336. Likewise SFL124 band pattern showed similarity to its parental strain SFL1339 (SFL1, Pcr$^+$) and between their respective lysogen strains SFL1333 and SFL1338. This supported the identity between the parent strains and their derivative strains. The slight difference in the band patterns observed between the lysogen strains and phage SfV is attributed to the presence of SfV-integrated bacterial host DNA in the lysogen strains. The ~1 kb extra band detected in the *Bam*HI digests and the ~5 kb band in the *Eco*RV digests of SFL1336 and SFL1337 are generated from pNV775 which contains the SfV *xis* gene ligated to pTrc99A vector (Figure 6.3, lanes 6-9). SFL124 was the bacterial host used during the propagation of phage SfV, therefore, it is possible that SFL124 DNA could still be present in the SfV genome DNA preparation which was used for labelling. Both SFL124 and its parental strain SFL1339 only showed DNA fragments which cross-reacted with phage host SFL124 DNA present in the probe preparation, and did not show the band pattern typical of phage SfV (Figure 6.3, lanes 10, 11, 14, 15). Above results thus confirm the identity of SFL1336, SFL1337, SFL1338 and SFL1333 with respect to the presence or absence of the prophage.

6.2.3. *Preparation of SFL1 isogenic strains transformed with plasmids containing genomic portions of phage SfV*

SFL1 derivatives carrying plasmids with various phage SfV genomic fragments were constructed in order to investigate the role if any, of specific phage SfV genes on host invasiveness. Five isogenic strains of SFL1339 (SFL1 Pcr^{+}) containing recombinant plasmids with various portions of the phage genome were produced (Figure 6.4). These include SFL1342, SFL1346, SFL1347, SFL1348 and SFL1349 (Figure 6.4). SFL1342 is SFL1339 with only the pUC18 vector, this was used as a baseline control. SFL1346 is SFL1339 containing pNV731 which has pUC18 with PCR amplified (~2.8 kb) three-gene serotype-conversion cluster, $gtrA_{(V)}$, $gtrB_{(V)}$ and *gtrV*. SFL1347 is SFL1339 containing pNV724 which has pUC18 with the *Bam*HI fragment E portion (~3.1 kb). This was originally cloned in the B828 *E.coli* strain and codes for the putative origin of replication, DNA adenine methylase and other regulatory proteins. SFL1348 is SFL1339 containing pNV314 which has pUC19 with the *Eco*RI fragment D portion (~3.4 kb). This was originally cloned in the B367 *E. coli* strain and codes for the excisionase (Xis), the amino terminal half of the integrase protein (Int) and some unknown protein [Huan, 1997A #118][Allison et al., submitted for publication]. SFL1349 is SFL1339 containing pNV324 which has pUC19 with the *Bam*HI fragment B portion (~7.3 kb). This was originally cloned in the B377 *E. coli* strain and codes for the putative early regulatory and immunity proteins. These strains were tested for their capacity to invade confluent HeLa cell monolayer.

6.2.4. *HeLa cell invasion assay*

The strains containing SfV genes were assessed for their invasive capabilities using the HeLa cell invasion assay. The invasion level of the lysogen strain SFL1338 was evaluated against its serotype Y parent strain SFL1339. Also tested were the cured strain SFL1337 versus its serotype V parent strain SFL1336 and the SFL1339 isogenic

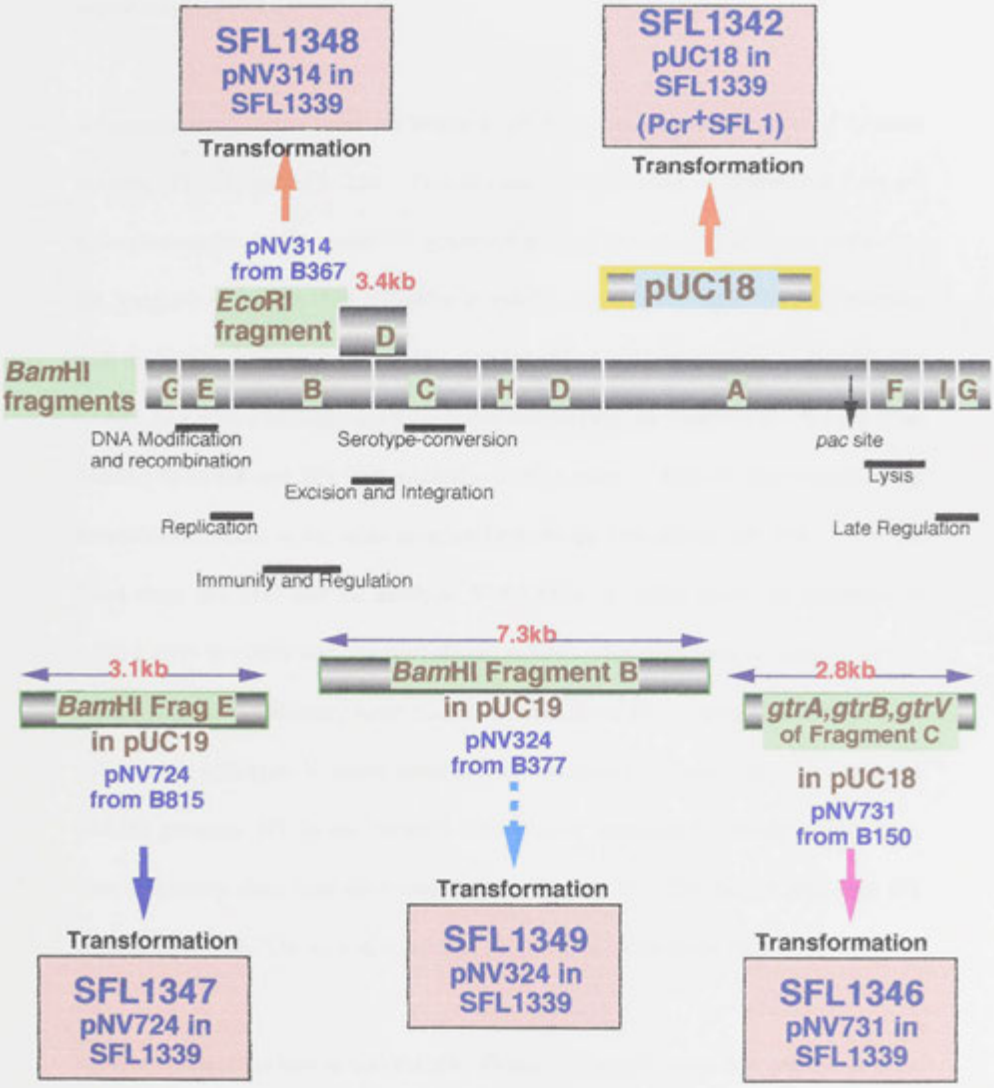
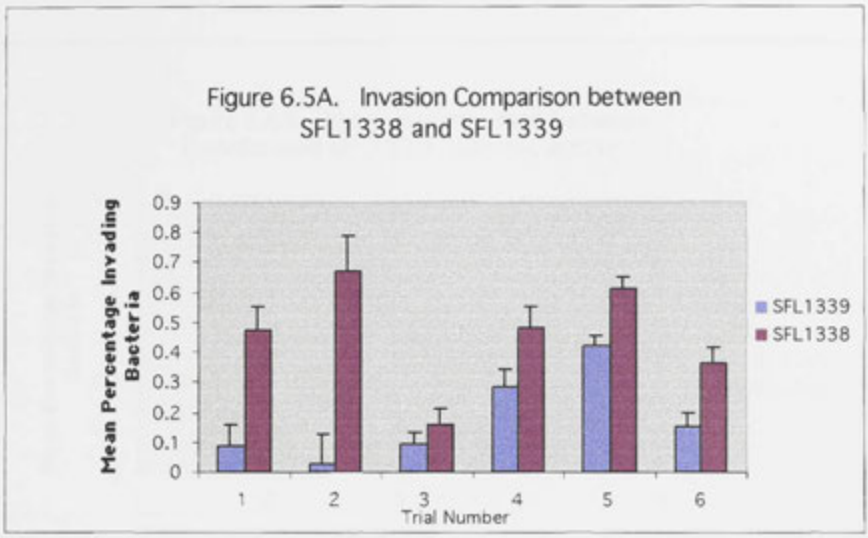


Figure 6.4. Diagram showing phage SfV segments cloned in isogenic SFL1339 strains . The functional map of the phage genome, as derived by Allison *et al.*, is included

strains were compared against the baseline invasin value for SFL1342 (SFL1339 with the vector only).

Six invasion assay trials were performed in duplicate to compare the level of invasion between SFL1338 and SFL1339. This was performed in order to determine if there are factors encoded in the integrated SfV genome that could enhance the invasive potential of the lysogenic strain SFL1338. Results revealed a consistently higher level of invasion with the lysogenic *Shigella*, SFL1338, than for SFL1339 (wild-type SFL1, Pcr^{+})(Figure 6.5A). At a 95% confidence interval, paired comparison one-sample t-test for the mean invading SFL1338 and SFL1339 yielded a 2.669 t value. This denotes a statistically detectable difference in the mean invasion between the two groups (p=.044). For the cured strain SFL1337 and the serotype V SFL1336, a higher statistical difference of 6.738 t value (p=.007) was observed (Figure 6.5B). The mean level of invasion for the cured strain was significantly lower than the invasion level for the control strain SFL1336 which is the wild-type V strain containing the uninduced *xis* gene. The data indicates that the presence SfV in the bacterial chromosome enhances *S. flexneri* invasiveness. This is apparent since both the lysogenic type Y strain SFL1338 and the wild-type SfV lysogen strain SFL1336 showed consistently higher percent invasion values.

In order to determine how invasiveness is effected by specific phage SfV genomic regions, the level of invasion by strains SFL1342, SFL1346, SFL1347, SFL1348 and SFL1349 was compared (Figure 6.6A). The Tukey multiple comparison test was used to detect



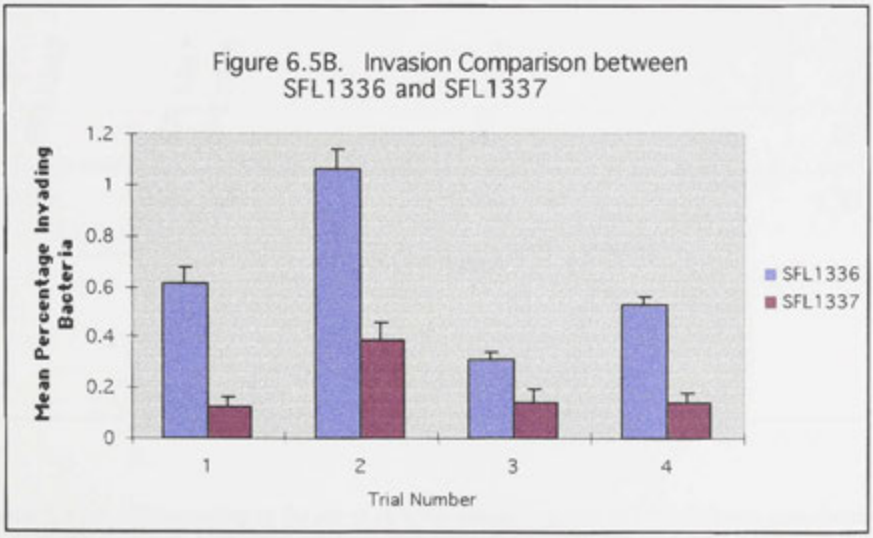
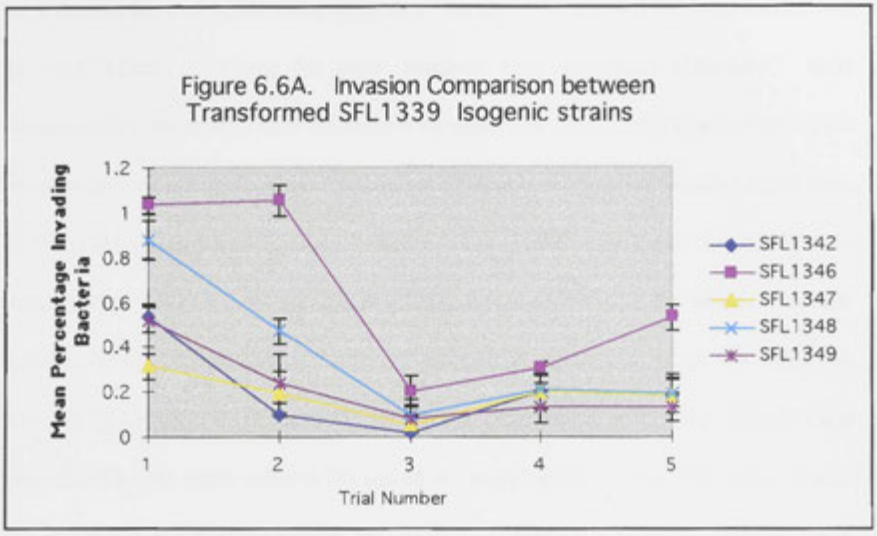


Figure 6.5. A, Comparison of the invasion level between the serotype V lysogen strain SFL1338 and its serotype Y parent SFL1339. B, Comparison of the invasion level between the serotype V strain SFL1336 and its serotype Y cured strain SFL1337.



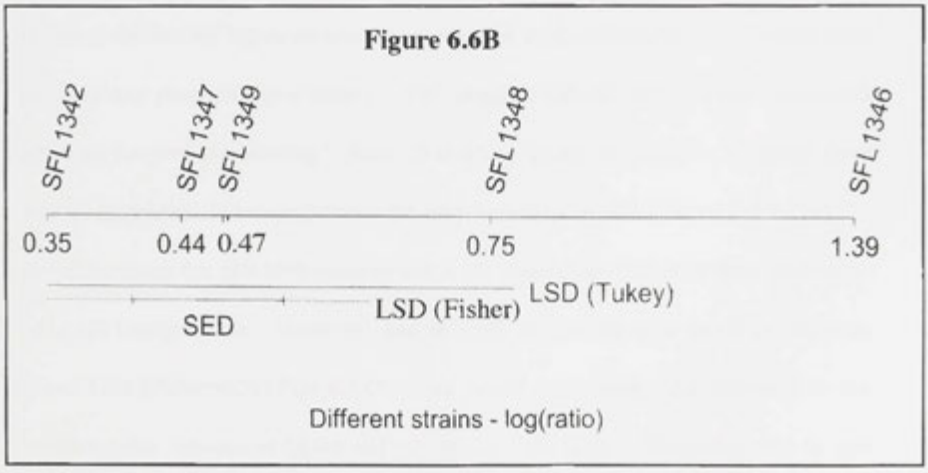


Figure 6.6. A. Comparison of the invasion level among isogenic SFL1339 strains transformed with recombinant plasmids containing different segments of SfV genome as insert. The location of the segments in the SfV physical map with their functions in the functional map are indicated in Figure 6.4. B. Tukey multiple comparison showing the ratio values of the different strains on a logarithmic scale. SED, standard error of difference is equal to 0.26; LSD (Tukey), is more conservative with the least significant difference being equal to 0.79; LSD (Fisher), is more liberal with the least significant difference being equal to 0.55.

significant differences between the mean invasion values (Figure 6.6B). Ratio values were derived for each strain using analysis of variance that allowed for treatment as well as batch effects. Using the more stringent least significant difference, LSD (Tukey=0.79), the strain's ratio values on a log scale were compared (Figure 6.6B). This analysis showed that there was no significant difference between the invasive capabilities of SFL1342 (ratio value=0.35) and all the other strains except for SFL1346 (ratio value=1.39). Likewise, no significant difference was observed in the level of invasion between SFL1347 (ratio value=0.44) and all the other strains except for SFL1346, between SFL1349 (ratio value=0.47) and all the other strains except for SFL1346 and between SFL1348 (ratio value=0.75) and all the other strains. From this data, it was concluded that SFL1346 containing the serotype conversion genes was the only strain showing significantly higher invasion values than the control strain SFL1342 (which does not have any phage SfV gene insert). This suggests that the SfV serotype-conversion genes are involved in increasing *S. flexneri's* ability to invade HeLa cells. Although there was no significant difference between the invasion values of SFL1348 and SFL1346, it should be noted that SFL1348 invasiveness is not significantly different from that of the SFL1342 baseline value. However, had we made our comparisons based on the more liberal LSD (Fisher=0.26)(Figure 6.6B), we would see a significant difference in the invasion value between SFL1348 and the baseline SFL1342. Therefore, due to this discrepancy between LSD (Tukey) and LSD (Fisher), the contribution of SFL1348 containing genes encoding the phage's site-specific integration proteins, in stimulating *Shigella* invasiveness should be further investigated.

6.2.5. *Invasiveness of gtrV mutant (gtrV')*

The role of the three gene serotype-conversion cluster in invasion was investigated further. The results described above had shown statistically significant increase in invasion by SFL1346, which carries $gtrA_{(V)}$, $gtrB_{(V)}$ and *gtrV*. In order to determine if the increase in invasion was mediated by serotype-conversion, mutations that would cause a frameshift in the *gtrV* gene were constructed. Initially, pNV731 NcoI site (nt 1074) was digested, filled-in, re-ligated and transformed into JM109 host (B1039). The resulting recombinant plasmid, pNV909 was sequenced and the presence of the predicted frameshift mutation confirmed. Four bases G, T, A and C were inserted between nt 1076 and 1077 of the *gtrV* gene causing a frameshift from frame a to c. The amino acid sequence was conserved until Pro158, afterwhich, the succeeding Trp159 was displaced with a cysteine residue followed methionine and aspartic acid. Plasmid pNV909 was re-transformed into SFL1339. However, the resulting recombinant strain SFL1367 still exhibited strong agglutination with anti-TypeV antisera, with the supposed *gtrV* mutation (4 base insertion at the 180th nucleotide upstream from the C terminus). This suggested the *gtrV* gene was still functional and capable of adding glucosyl groups into the O-antigen to mediate serotype-conversion. It also indicated that the product encoded by the 180 nucleotide bases downstream from the *gtrV* *Nco*I site was not required to mediate serotype-conversion and that the active site of the glucosyltransferase is still intact.

An alternative *gtrV* restriction site, *BclI* was also used to disrupt the *gtrV* gene. The *Bcl*I enzyme used was sensitive to Dam methylation, so pNV731 was initially transformed into the Dam$^-$ host, B834, creating B1040. The non-methylated pNV731 was digested with *Bcl*I at *gtrV* nt 425, filled-in, re-ligated (the recombinant plasmid was designated as pNV934), and transformed into the Dam$^-$ *E. coli* host to yield strain B1076. pNV934 sequence revealed four bases C,T,A,and G inserted into the *Bcl*I site of the *gtrV* gene between nt 427 and 428 (Figure 6.7). The amino acid sequence was conserved until Asp 312, afterwhich His 313 was displaced with an arginine residue followed by serine and leucine found along the d frame. When digested with *Bcl*I, plasmid pNV934 was not cut because it had lost its *Bcl*I site due to the alteration of *Bcl*I palindrome sequence brought about by the four base insertion into the restriction site (Figure 6.7). The latter strain did not agglutinate with anti-TypeV antisera but only with the anti-Group 3,4 antisera. This suggested that the serotype conversion glucosyltransferase gene (*gtrV*) in SFL1394 became non-functional due to the frameshift mutation (Figure 6.7). After confirmation of pNV934 genotype, it was then electroporated into SFL1339 (SFL1, Pcr$^+$) to produce SFL1394.

SFL1394 was tested against SFL1342 and SFL1346 in the invasion assay. Using the Tukey multiple comparison test, the percent mean invading SFL1342, 1346 and 1394 cells were compared using the LSD value on a log scale equivalent to 0.99. There was no statistically significant difference between the ratio values of invading SFL1346 (ratio value=0.66) and SFL1394 (ratio value=0.31). And the mean percentage invasion of

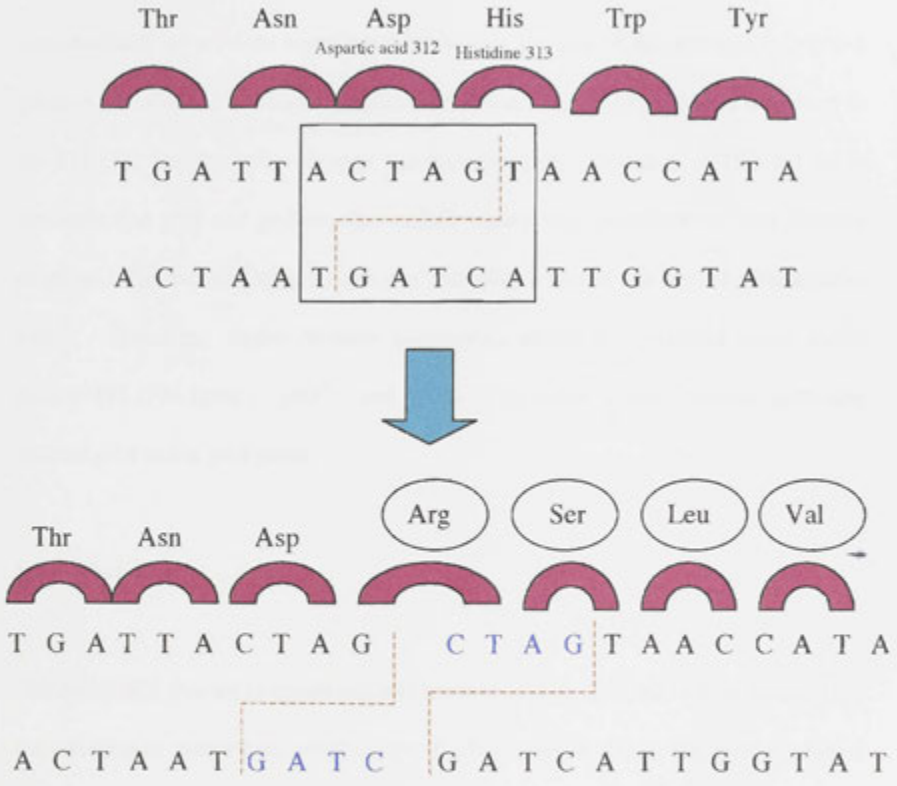
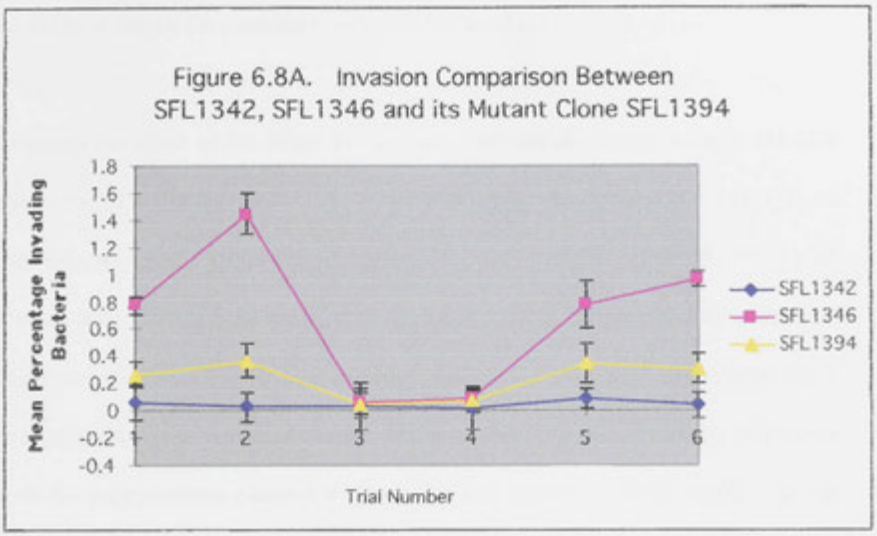


Figure 6.7. Diagram showing insertion of four nucleotide bases (blue letters) after pNV731 restriction with *Bcl*I (site indicated by red dashed lines). Frameshift mutation resulted in changes in the translated amino acid residues (encircled). Shown are six amino acids: Threonine310, Asparagine311, Aspartic Acid312, Histidine313, Tryptophan314, and Tyrosine315, out of a total 417 amino acid residues of the GtrV protein.

SFL1394 was consistently lower than SFL1346 but consistently higher and significantly different from the baseline SFL1342 invasion ratio value (ratio value=0.07) (Figure 6.8A and B). The observed difference between SFL1346 and SFL1342 confirmed the reproducibility of previous experiments implicating the role of the serotype-conversion genes in the increased invasion capability of *S. flexneri* host. SFL1394 did not revert to the SFL1342 baseline value despite the introduced *gtrV* mutation. This led us to speculate that *gtrA* and *gtrB* together or individually may contribute to host invasive potential. The data also suggests that *gtrV* may play a partial role in conferring invasive traits. Therefore, further invasion experiments should be conducted which would include SFL1394 (*gtrA*$^+$, *gtrB*$^+$, and *gtrV*$^+$) being tested against strains containing mutated *gtrA* and/or *gtrB* genes.

6.3. Discussion

The ability of *S. flexneri* to invade and proliferate in colonic epithelial cells is essential for the pathogen to induce bacillary dysentery [LaBrec, 1964 #331]. The primary factors involved in the invasive process are expressed by complex genes on the large 230-kb virulence plasmid [Watanabe, 1985 #497][Sasakawa, 1992 #52] as demonstrated by tissue culture studies [Hong, 1998 #141][Perdomo, 1994a #41]. It has been shown that virulence genes such as bacterial toxins and adhesion molecules can be encoded by bacteriophages [Sjogren, 1994 #500][Cheetham, 1995 #30]. Thus the aim of this study



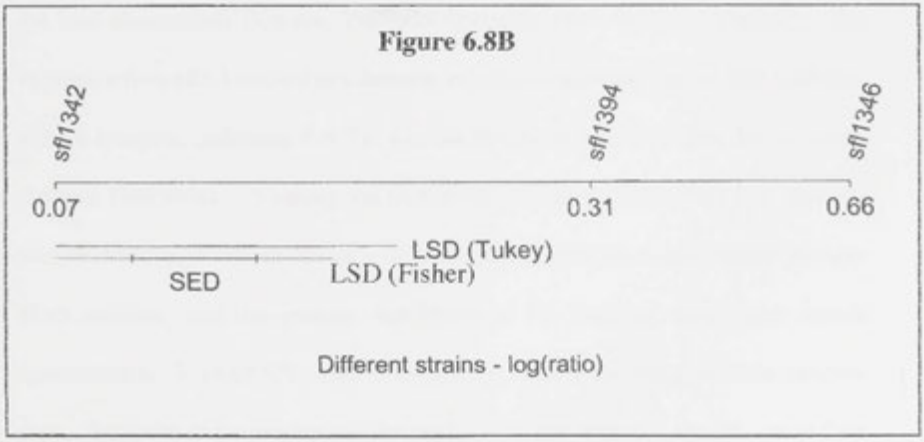


Figure 6.8. A, Comparison of the invasion level among SFL1342 (SFL1339 containing the pUC18 vector), SFL1346 (SFL1339 containing the serotype-conversion genes in pUC18), and SFL1394 (SFL1339 containing pNV731 with a mutated gtrV gene). B, Tukey multiple comparison showing the ratio values of the different strains on a logarithmic scale. SED, standard error of difference is equal to 0.36; LSD (Tukey), is more conservative with the least significant difference being equal to 0.99; LSD (Fisher), is more liberal with the least significant difference being equal to 0.80.

was to investigate whether serotype converting bacteriophage SfV was able to influence the ability of *Shigella* to penetrate mammalian HeLa cells.

To assess the effect of the phage SfV genome, we initially tested lysogen SFL1338 against its wild-type progenitor strain SFL1339 (SFL1 Pcr^+), and the wild-type V SFL1336 carrying the uninduced *xis* gene against the cured derivative SFL1337 via an invasion assay. During the selection of the strains, the Congo red binding phenotype (Pcr^+) was maintained. This was necessary since lack of ability to bind Congo red is correlated with loss of virulence [Sereny, 1957 #109][Murayama, 1986 #498] and occurs when the large virulence plasmid is mutated or lost [Maurelli, 1984B #499]. In our curing experiments, excess Xis production successfully induced the SfV prophage from the host chromosome [Kaneko, 1997 #280][Stanley, 1999 #248]. Similarly, Xis expression from pRK5 resulted in a dramatic increase in the curing rate for both wild-type and *lon* lysogens, indicating that Xis was the limiting factor rather than Int for curing [Leffers, 1998 #246]. Weisberg and Gottesman [Weisberg, 1971 #501] first observed that excessive quantities of Xis *in vivo* inhibit phage integration and instead promote phage excision, and that protease degradation of Xis leads to rapid phage lambda lysogenisation. In phage SfV, ORF 8 encodes for Xis, and overlaps with the integrase gene. Similarity in the overlap and the organisation and sequence of *att*P, *int* and *xis* between phage SfV and P22 were also observed [Huan, 1997A #118][Vander Byl, 2000 #351][Allison, 2000 #199].

Part of the screening process used to ensure phage integration and deletion was phage sensitivity. This was observed in earlier experiments in which *S. flexneri* serotype Y (3,4), became resistant to superinfection with homologous phages after phage Sf6 infection of *S. flexneri* [Lindberg, 1978 #120][Clark, 1991 #119]. In our experiments, the lysogen strains SFL1333 and SFL1338 had successfully acquired phage SfV which conferred resistance to further SfV infection. The lysogens grew and did not lyse when reacted with phage SfV. Conversely, the cured strain SFL1337 had lost phage SfV which made it susceptible to phage SfV infection as demonstrated by a clear zone where the phage suspension was placed. This supports the role of O-antigen modification in eliciting immunity against infecting phage.

The consistently higher invasive capacity of wild type V strain and lysogenic *S. flexneri* strain suggested one or several factors encoded by the phage genome may be involved in invasiveness. This led us to create various SFL1339 isogenic strains which were basically identical except that they carried various portions of the phage SfV genome introduced through recombinant plasmids. These strains with an identical genotypic composition except for the introduced plasmid insert were ideal for determining the specific genomic part which contributes to the invasive potential of the bacteria. Of the constructs prepared, each carried specific SfV gene fragment inserts encoding for different proteins (serotype conversion, attachment, excision, integration, replication, methylation, immunity and regulation) [Huan, 1997B #117](Allison *et al.*, submitted for publication). The strain SFL1346, had the highest level of invasion, it was significantly

different and higher than the other mutant constructs, most specifically to the baseline value of SFL1342. This indicated that the addition of a glucosyl residue to the rhamnose II of the O-antigen sugar units via an α1,3 linkage as occurred in SFL1346 confers a conformational change in the LPS structure which made it more reactive or accessible to the epithelial cell membrane.

The alteration of the LPS through O-antigen modification has been consistent with the knowledge that LPS has important roles in bacterial entry, intercellular spread in the intestinal epithelium as well as resistance to host defences [Hong, 1998 #141][Okada, 1991A #22][Rajakumar, 1994 #128]. Studies involving rough mutants of *S. flexneri* lacking O-antigen and those having mutated O-antigen chain length and composition were observed to be capable of cell invasion but incapable of intra- and intercellular spread [Okamura, 1983 #447][Hong, 1997 #545][Okada, 1991A #22][Sandlin, 1996 #544]. However, the specific role of O-antigen conversion on virulence in *S. flexneri* has not been investigated. This made our experiments important and unique in that SfV is the first *S. flexneri* bacteriophage characterised to have played a role in the invasion potential of its host.

The *gtrV* gene is the main gene involved in the addition of the glucosyl moiety into the O-antigen. Therefore, a frameshift mutation was introduced in this gene in an attempt to hamper the attachment the glucosyl residue to the repeating sugar units of the LPS and

thus prevent the modification of the O-antigen. However, only partial reduction in the invasion level of SFL1394 was observed.

It is also possible that *gtrA* and *gtrB* may have a role in enhancing the invasive properties of *S. flexneri*. Therefore, further experiments should be conducted to prove this. These will involve recombinant clones containing plasmids with varied combinations of mutated serotype-conversion genes insert. These strains should then be tested for invasiveness in order to refute or add support to our findings and provide information as to the precise contribution of each of the three serotype-conversion genes in conferring invasive traits. Also, further work can be performed to determine if other O-antigen glucosylating systems of other *S. flexneri* bacteriophages such as SfX, SfII, and SfI, would also exhibit similar effect on host invasiveness.

VII GENERAL DISCUSSION

CHAPTER 7

General Discussion

The primary objective of this study is to characterise additional portions of the bacteriophage SfV genome. Specifically, we have sequenced a 10.1 kb *Bam*HI fragment A segment which contains the late gene region. Eight open reading frames were found to have homology to proteins which are essential for the structural integrity and morphogenesis of the phage. Some of these include the terminase, portal protein, tail assembly proteins, the capsid and the protease whose functions were subsequently confirmed through experiments.

7.1. The serotype conversion and early region of phage SfV genome

The serotype-conversion genes of several *S. flexneri* bacteriophages including SfII, SfV, SfX and Sf6 have been sequenced and characterised in the previous decade [Huan, 1997B #117][Mavris, 1997 #16][Adhikari, 1999 #112][Guan, 1999 #151][Verma, 1993 #116]. An additional 15 kb portion of the phage SfV early region has recently been characterised in our laboratory (Allison *et al.*, submitted for publication). This region contains genes which are concerned with regulation, DNA modification, recombination, replication and immunity. Important genes identified were *orf129*, *orf224*, *orf66*, *orf217* and *orf77*. The *orf129* gene encodes for a putative recombination factor and is also involved in DNA repair. The gene product has homology to the RusA endonuclease encoded by *E. coli*

DLP12 prophage and phage 82 (Allison *et al.*, submitted for publication) [Mahdi, 1996 #502][Sharples, 1999 #503]. Other genes *Orf224* and *orf66* were shown to have homology to lambda phage *cI* and *cro* repressor genes, respectively (Allison *et al.*, submitted for publication). The CI repressor protein promotes lysogeny by attaching to operator sequences adjacent to the *cI* gene. This binding prevents transcription of lytic genes and stimulating transcription of *cI*, while the Cro protein prevents the transcription of *cI*. Conversely, the Cro protein prevents transcription of *cI* by attaching to the operator sequences upstream of the gene [Campbell, 1994 #164]. Functional studies confirmed that the protein product of *orf217* induces DNA adenine methylase (Dam) activity (Allison *et al.*, submitted for publication). Based on functional and sequence analysis, Allison et al., also reported on three superinfection immunity mechanisms adapted by phage SfV. These include the O-antigen modification, repressor-mediated lambdoid immunity system and immunity involving transcription termination (Allison *et al.*, submitted for publication). Since serotype-conversion genes and most early genes have been identified and characterised, we have characterised further the genes downstream and to the right of the serotype-conversion locus, specifically the 10.1 kb portion of *Bam*HI fragment A adjacent to the *pac* site.

7.2. Fragment A encodes for phage SfV structural proteins necessary for viral packaging of DNA and morphogenesis.

The 10.1 kb *Bam*HI fragment A portion of the genome contains late region genes involved

in DNA packaging and viral morphogenesis. Our experiments demonstrated that ORF200 is a phage SfV protease while ORF409 is a phage SfV capsid protein (Chapters 4 and 5). It also indicated that ORF409 is a processed primary translation product suggesting the occurrence of capsid cleavage similar to the head assembly process of other bacteriophages such as coliphage HK97 [Duda, 1995A #275]. Database homology searches of fragment A found bacteriophages HK97, ϕC31, D3, ϕPVL and ϕ105 showed similar head morphogenesis gene function and organisation. These observations suggest that the portal, protease and capsid proteins encoded in these bacteriophages may exhibit similar head assembly mechanisms.

Observations and experimental results indicate there are similarities between the head assembly process of phage SfV and HK97. Thus, it is possible that phage SfV may also utilise other steps employed by HK97 in its head assembly process. The HK97 head assembly pathway illustrates the serial transitions which occur in the capsid subunits of phages that do not utilise a scaffold protein [Hendrix, 1994 #263]. Initially the 42 kDa pentameric and hexameric capsid subunits assemble into the first form of the shell in the absence of a scaffold protein. These subunits are then cleaved by the viral protease from its amino terminus to attain their mature conformation. Following cleavage, the head shell expands, resulting in a major conformational change [Duda, 1995A #275]. This triggers an ultimate shell strengthening transition where all shell subunits become covalently crosslinked in an autocatalytic reaction. Although it is tempting to consider a parallel mechanism being employed for phage SfV head assembly, further experiments

should be conducted before conclusions can be drawn as to the similarity in head processes between the two phages.

The tail assembly components contained in fragment A are similar to other viral counterparts. SfV *orf56* and *orf498* are homologous to phage Mu tail tube and sheath protein, respectively [Takeda, 1998 #321]. Downstream of *orf498* are *orf89* and *orf116* which both have homology to the phage P2 essential tail protein *E* gene and *T* gene, respectively. This grouping of SfV tail protein genes is another example of gene module clustering which was also observed in the head assembly gene region. In the assembly process of the T4 tail, the presence of a complete T4 tail baseplate triggers a conformational change in gp19. The gp19 polymerises into a tail tube on the baseplate and is surrounded by co-polymerising sheath monomers [Arisaka, 1988 #525][Eiserling, 1983 #509]. When infection occurs the sheath contracts and the tube penetrates the outer membrane of the host *E. coli*. The T4 genome is then transferred through the tube into the periplasmic space of the bacterial cell. This suggests that other phage SfV tail region genes may encode for tailspike proteins that possess endorhamnosidase activity. These proteins would be capable of digesting the O-antigen to allow passage of the phage through the LPS barrier and onto the surface of the outer membrane where tight binding occurs [Steinbacher, 1997 #510]. Recently, functional studies involving assay of cytoplasmic extracts containing the *S. flexneri* bacteriophage Sf6 *orf1*-encoded tailspike protein demonstrated the ability of the tailspike to hydrolyse the O-antigen [Chua, 1999 #547]. Tailspike proteins possess one of the most vital functional roles in viral

propagation in the host, since they are involved in the initial steps of attachment and penetration [Steinbacher, 1997 #510]. Therefore, the tail component genes of SfV should be further characterised.

7.3. The role of phage SfV ORF577 putative terminase in DNA packaging

There are components in DNA packaging which are common among different phage. These are a mature head shell which acts as the DNA container and two subunits of packaging proteins, the terminases [Fujisawa, 1994 #511]. The smaller subunit is the DNA binding protein while the larger subunit contains metal and ATP binding motifs and is able to cut the DNA concatemer [Guo, 1987 #512]. DNA concatemers are mature replicated phage DNA that are joined together in a head to tail fashion through terminal repetitions, and these are the substrates for terminase protein action. When the prohead shell is full of DNA, terminase cutting of concatemeric DNA molecule occurs. For phages using the headful packaging mechanism, this cleavage takes place via recognition of specific sequences at the *pac* site [Black, 1995 #325].

The experiments conducted to elucidate the role of phage SfV ORF577 putative large terminase subunit involved cloning *orf577* into the IPTG-inducible overexpression vector pT7-5. Upon induction, DNA samples from the clones were observed on agarose gel for DNA smearing as an indicator of terminase cleavage activity. This procedure was previously used by Bhattacharyya and Rao (1993) who showed that the overexpression

of the 18 kDa gp16 (small subunit) with the 69 kDa gp17 (large subunit) results in cleavage of the resident plasmid DNA plus also the host *E. coli* genomic DNA [Bhattacharyya, 1993 #514]. Compared to the uninduced clones, our preliminary trials showed reduction in viable cell growth over time after induction, suggesting a lethal effect of overproduced terminase by excessive cutting of the DNA. However, no smearing of cleaved DNA was observed. This implied that there may be additional factors that need to be cloned and expressed along with the putative large terminase ORF577. One factor could be encoded by *orf164*, which is located immediately upstream of ORF577 and has a similar size to other phage small terminase subunits. Studies have observed that the small subunit enhances DNA packaging and has ATPase activities associated with phage T4 gp17 [Bhattacharyya, 1993 #514][Bhattacharyya, 1994 #515]. However, the action of the small terminase subunit in phage T4 is thought to be redundant since the large subunit (gp17) alone can catalyse the cleavage of both plasmid and host DNA when expressed in *E. coli*. Also, gp17 is highly active during *in vitro* DNA packaging where it exhibits ATPase activity [Bhattacharyya, 1994 #515][Leffers, 1996 #516]. Therefore, the possible role of the of phage SfV *orf164* putative small terminase subunit should be investigated further.

Another protein which could be incorporated with the terminase insert is the putative portal protein ORF367. The portal protein is a dodecameric ring structure at the base of the prohead and contains a central channel of diameter 3-4 nm where the DNA is held prior to prohead entry and exit [Valpuesta, 1994 #517]. *Orf367* could have been

expressed with the putative terminase gene in order to mediate DNA cleavage since it could have bound the DNA securely before the terminase could execute more precise cutting. Recent findings on phage T3 reported that the portal-bound terminase is essential for DNA translocation, as well as for processing the DNA concatemer [Black, 1995 #325][Morita, 1995 #518]. In another study, functional domains of bacteriophage T4 large terminase (gp17) and portal protein (gp20) were determined by mutant analysis and sequence localisation [Lin, 1999 #519]. It was discovered that DNA packaging occurs due to an ATP-driven translocation of concatemeric DNA into the prohead, a process driven by the phage terminase which is complexed with the portal vertex dodecamer of the prohead [Black, 1995 #325]. These findings support the theory that the portal protein and large terminase are vital components of the packasome complex and that they are involved in the processing and packaging of DNA into the prohead. Recombinant plasmids could be created which would include SfV *orf367* along with *orf577* as gene inserts for the overexpression experiment. Constructs of this genotype may elicit DNA cleavage which would be visualised as smearing, not observed on earlier trials. Cloning of the terminase and portal protein was attempted but not completed due to time constraints. Therefore, the role of the putative portal gene should be investigated in future functional analysis of the phage SfV terminase protein.

7.4. Serotype-conversion genes are phage SfV-encoded virulence factors.

There are several reasons for characterising the genomic composition of phage SfV. This

includes the discovery of genes which encode for structural proteins, such as those encoded in the fragment A, and genes encoding for essential functions such as regulation, immunity, replication, methylation, and morphogenesis. In addition, phage SfV-encoded virulence factors may be identified. Several studies on bacteriophage-encoded virulence genes have been carried out recently [Miao, 1999 #524][Waldor, 1998 #537][Cheetham, 1995 #30]. For example, *Vibrio cholerae* bacteriophage VPIϕ encodes for a pilus that functions as a colonisation factor in the human intestine in addition to being the receptor for the cholera toxin-encoding ϕCTX bacteriophage [Karaolis, 1999 #165]. Another example is the *Salmonella* bacteriophage encoded SopE Φ [Hueck, 1998 #139][Hardt, 1998 #521]. The *sop*E gene product, is an effector of the *Salmonella* type III secretion systems (TTSS) that has been described as a sort of bacterial syringe for simultaneous injection of multiple effector proteins which induce host cell physiological modifications [Hueck, 1998 #139]. SopE is involved in the induction of *in vitro* epithelial cell invasion by induction of membrane ruffling as well as stimulation of gastrointestinal inflammation [Hardt, 1998 #521]. Virulence factors such as toxins may be encoded by phage SfVand could provide the bacterium with mechanisms to breach host structural barriers. Examples of where this occurs are diphtheria toxins, cholera toxins and *E. coli* serotype 0157:H7 toxin that causes hemolytic-uremic syndrome in children [Groman, 1984 #522][Waldor, 1996 #241][O'Brien, 1984 #242]. Several studies have demonstrated the role of LPS as a virulence factor regulating the ability of *S. flexneri* to spread intercellularly and to evade host immune defences [Hong, 1997 #545][Okada, 1991A #22][Rajakumar, 1994 #128], however, the effect of LPS O-antigen modification

in *S. flexneri* virulence has not been investigated, and this has made our experiments very important and unique.

As exemplified above, virulent bacterial strains can evolve by acquiring virulence factors from bacteriophages. Hence, we have tested the contribution of phage SfV genome in enhancing *S. flexneri*'s capacity to invade mammalian cells as a measure of virulence. Our results showed that *S. flexneri* strains with integrated phage SfV had significantly increased invasion capability compared to their isogenic non-lysogens. The serotype-conversion genes were identified as the portion of the SfV genome which enhanced *Shigella* invasiveness. Invasive *Shigella* strains transformed with the serotype-conversion genes showed a statistically significant increase in invasive potential. This indicate that the serotype-conversion may permit *S. flexneri* to evade the host defence system by varying the bacteria's surface LPS molecular and structural conformation. This change in LPS structure may have permitted increased elaboration of *S. flexneri* type III secretion system facilitating bacterial host capacity to invade. It is possible that the modified LPS may have become more reactive to the target epithelial cell receptor, resulting in increased cell surface binding and subsequent facilitated intrusion. The new LPS conformation may have also effected the mobilising components of the host phagocytic mechanism such as the actin filaments that form under the phagocytic vacuolar membrane during invasion. We then went on to construct a *gtrV*⁻ mutant in order to determine its influence on invasion level. Disruption of the *gtrV* gene resulted in a partial reduction in the level of invasion, suggesting that the other serotype-conversion

genes may also be involved in enhancing invasiveness. Therefore, further studies on the effects of $gtrA_{(V)}$ and $gtrB_{(V)}$ should be conducted.

7.4. Conclusion and Future Direction

This study has reported on the molecular characterisation of the 10.1 kb portion of phage SfV *Bam*HI fragment A. The protein sequence homology data established the genetic linkage between phage SfV and other double stranded bacteriophage which infect bacteria of a broad phylogenetic range. We have located *orfs* which encode for DNA packaging and structural morphogenetic factors within fragment A. These include genes which we have confirmed as being involved in capsid cleavage during phage head assembly. Given the complexity of the process and the broad similarities within bacteriophage groups, it is important to conduct further experiments in order to understand subsequent steps in the phage assembly process after the capsid cleavage. Other SfV *orf* functions should also be investigated. For example, the N-terminal sequence of the ~55 kDa protein detected by the anti-SfV antisera could be determined and aligned with the predicted amino acid sequences of the SfV *orfs*. Morphological studies of tail components could also be performed by doing purification and visualisation through transmission electron microscopy. Lysogenic strain mutants containing aberrant SfV tail genes can also be created, induced out and their morphology observed. The putative terminase subunits and portal protein genes comprising the putative SfV DNA packaging module could also be investigated using mutation studies to establish protein identity and function.

Plasmids carrying inserts of various combinations of the DNA packaging genes could be constructed and compared in their ability to cut DNAs which can be visualised as a smear along the agarose gel lane.

Further studies on the role of SfV genome in *S. flexneri* virulence could be investigated. Aside from HeLa cell invasion assay, experiments involving mutagenesis, other virulence tests involving animal models such as guinea pig eyes or mouse ileal loops and other functional assays such as the plaque assay and serum sensitivity assay can be performed in order to establish and support the true role of SfV genes not only in their contribution to bacterial invasiveness but also in other virulence properties like their ability to manifest intra- and intercellular spreading and survival within phagocytic cells. Our results implicating the influence of the serotype-conversion genes in bacterial invasion is preliminary and other experimental set-ups can be considered. For instance, the invasion level of SFL1339 SfV lysogen and SFL1346 containing the O-antigen glucosylation genes can be compared to assess if there are factors other than the serotype-conversion region that are active in enhancing invasiveness. The role of $gtrA_{(V)}$ and $gtrB_{(V)}$ should be examined by mutational studies since only partial reduction in *S. flexneri* invasion level was observed when the *gtrV* gene was mutated. This implicated $gtrA_{(V)}$ and $gtrB_{(V)}$ as factors working in adjunct with the *gtrV* gene in eliciting increased level of *S. flexneri* invasion.

Most of SfV genes have now been determined and characterised. However, completion

of phage SfV genome analysis should be accomplished in order to discover not only essential regulatory and structural genes but other sequences which could be utilised in the development of large-capacity vectors and site-specific integration vectors necessary in the application of genetic engineering techniques. Finally, decoding the complete genetic blueprint will ultimately lead to a comprehensive understanding of the physiology and molecular biology of bacteriophage SfV.

APPENDIX

APPENDIX

Bacterial Growth Media

LB Broth

Tryptone	10 g
Yeast Extract	5 g
NaCl	5 g

Make up to 1 L in water and sterilize by autoclaving.

LB Agar

As for LB broth, adding 15 g/L agar before autoclaving.

Ampicillin

Add filter-sterilised ampicillin (10 mg/mL) to cooled broth or agar (<55°C) to a final concentration of 100 μg/mL.

IPTG / X-Gal Plates

Add 20μL each of X-Gal and IPTG solutions (20 mg/mL) and spread on plate.

Bacteriophage Techniques

NZCYM Medium

NZ Amine	10 g
NaCl	1.7 g
Cassamino acids	1 g
$MgSO_4.7H_2O$	0.7 g
Yeast extract	1.7 g

Dissolve in 1 L of water and sterilize by autoclaving.

SM Buffer

1 M Tris.HCl (pH 7.5)	5 mL
NaCl	0.4 g
$MgSO_4.7H_2O$	0.2 g
2% gelatine solution	0.5 mL

Make up to 100 mL with water and sterilise by autoclaving.

Plasmid Miniprep

Solution I

50 mM glucose
25 mM Tris.HCl pH 8.0
10 mM EDTA pH 8.0
Sterilised by autoclaving. Stored at 4°C.

Solution II

0.2M NaOH
1% SDS
Freshly prepared from sterile stock solutions (10% SDS, 3 M NaOH) before use.

Solution III

5 M CH_3COOK	60 mL
Glacial ethanoic acid	11.5 mL
Water	28.5 mL

Prepared from sterile stock solutions and stored at 4°C.

Chromosomal DNA Preparation

TES Buffer

1 M Tris.HCl pH 8.0	5 mL
5 M NaCl	1 mL
0.5 M EDTA pH 8.0	1 mL

Make up to 100 mL with water and sterilise by autoclaving.

Lysis Solution

1 M Tris.HCl pH 8.0	5 mL
0.5 M EDTA pH 8.0	12.5 mL
10% SDS	50 mL
MQ H_20	32.5 mL

TE Buffer

1 M Tris.HCl pH 8.0	200 µL
0.5 M EDTA pH 8.0	1 mL

Make up to 100 mL with MilliQ water and autoclave.

Agarose Gel Electrophoresis

0.6% TBE / Agarose Gel

0.6 g DNA grade agarose (Progen)
100 mL 0.5X TBE buffer

5X TBE Buffer

Tris Base	54 g
Boric acid	27.5 g
0.5 M EDTA pH 8.0	20 mL

Make up to 1 L with water.

Agarose Gel Loading Buffer

Glycerol	1.2 mL
10µg/mL Bromophenol blue	2.8 mL
MilliQ water	1 mL

Southern Hybridisation

20 X SSC

3 M NaCl
0.3 M tri-sodium citrate

20X SSPE

0.6 M NaCl
0.2 M Sodium Phosphate
0.02 M EDTA
Dissolve in distilled water, adjust to pH 7.4 and autoclave.

100X Denhardt's Solution

2% w/v BSA
2% w/v Ficoll™
2% polyvinyl pyrrolidone

Colony Hybridisation

Denaturing solution

1.5 M NaCl
0.5 M NaOH

Neutralising Solution

1.5 M NaCl
0.5 M Tris.HCl pH 7.2
1 mM EDTA

Immunogold Labelling

Phosphate Buffered Saline (PBS)

NaCl	8 g
KH_2PO_4	0.24 g
Na_2HPO_4	1.44 g
KCl	0.2 g

Make up to 1 L with water and sterilise by autoclaving.

REFERENCES

REFERENCES

Achi, R., Cam, P. D., Forsum, U., Karlsson, K., Saenz, P., Mata, L. & Lindberg, A. A. (1992). Titres of class-specific antibodies against *Shigella* and *Salmonella* lipopolysaccharide (LPS) antigens in colostrum and breast milk of Costa Rican, Swedish and Vietnamese mothers. *Journal of Infectious Diseases* **25**, 89-105.

Achi, R., Mata, L. & Lindberg, A. A. (1994). Serum antibody titers to *Shigella* lipopolysaccharides and invasion plasmid antigens in healthy Costa Rican and Swedish women. *Scandinavian Journal of Infectious Diseases* **26**, 329-337.

Ackerman, H. W. & DuBow, M. S. (1987). General properties of bacteriophages. In *Viruses of prokaryotes*, Vol. I. CRC Press, Boca Raton, Florida.

Ackerman, H. W. & Krisch, H. M. (1997). A catalogue of T4-type bacteriophages. *Archives of Virology* **142**, 2329-2345.

Adachi, T., Mizuuchi, M., Robinson, E. A., Appella, E., O'Dea, M. H., Gellert, M. & Mizuuchi, K. (1987). DNA sequence of the *E. coli gyrB* gene: Application of a new sequencing strategy. *Nucleic Acids Research* **15**, 771-784.

Adams, M. M., Allison, G. E. & Verma, N. K. (2001). Characterisation of the type-IV O-antigen modification genes in the genome of *Shigella flexneri* NCTC 8296. *Microbiology* **Accepted**.

Adhikari, P., Allison, G., Whittle, B. & Verma, N. K. (1999). Serotype 1a O-antigen modification: Molecular characterisation of the genes involved and their novel organisation in the *Shigella flexneri* chromosome. *Journal of Bacteriology* **181**(15), 4711-4718.

Adkins, H. J. e. a. (1987). Two-year survey of etiologic agents of diarrheal disease at San Lazaro Hospital. Manila, Republic of the Philippines. *Journal of Clinical Microbiology* **25**, 1143-1147.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J. D. (1994). Basic genetic mechanisms. In *Molecular Biology of the Cell* 3 edit. Garland Publishing, Inc., New York.

Allison, G. A. & Verma, N. K. (2000). Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends in Microbiology* **8**(1), 17-23.

Alonso, J. C., Luder, G., Stiege, A. C., Chai, S., Weise, F. & Trautner, T. A. (1997). The complete nucleotide sequence and functional organization of *Bacillus subtilis* bacteriophage SPP1. *Gene* **204**, 201-212.

Altermann, E., Klein, T. R. & Henrich, B. (1999). Primary structure and features of the genome of the *Lactobacillus gasseri* temperate bacteriophage ϕ adh. *gene* **236**, 333-346.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**(17), 3389-3402.

Amann, E. (1988). *Gene* **69**, 301.

Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayama, S. V. L., Pierson, L. S., Sternberg, N. & Leong, J. M. (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO Journal* **5**, 433-440.

Arisaka, F., Nakako, T., Takahashi, H. & Ishii, S. (1988). Nucleotide sequence of the tail sheath gene of bacteriophage T4 and amino acid sequence of its product. *Journal of Virology* **62**(4), 1186-1193.

Bachmann, B. J. (1990). Linkage map of *Escherichia coli* K-12. *Microbiology Reviews* **54**, 130-197.

Bahrani, F. K., Sansonetti, P. J. & Parsot, C. (1997). Secretion of Ipa proteins by *Shigella flexneri*: inducer molecules and kinetics of activation. *Infect Immun* **65**(10), 4005-10.

Baskin, D. H., Lax, J. D. & Barenberg, D. (1987). *Shigella* bacteremia in patients with the acquired immune deficiency syndrome. *American Journal of Gastroenterology* **82**, 338-341.

Bastin, D. A., Stevenson, G., Brown, P. K., Haase, A. & Reeves, P. R. (1993). Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Molecular Microbiology* **7**(5), 725-734.

Batchelor, R. A., Haraguchi, G. E., Hull, R. A. & Hull, S. I. (1991). Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of *Escherichia coli*. *Journal of Bacteriology* **173**(18), 5699-5704.

Bazinet, C. & King, J. (1985). The DNA translocating vertex of dsDNA bacteriophage. *Annual Reviews in Microbiology* **39**, 109-129.

Beatty, W. L. & Sansonetti, P. J. (1997). Role of lipopolysaccharide in signaling to subepithelial polymorphonuclear leukocytes. *Infect Immun* **65**(11), 4395-4404.

Berg, C. M., Wang, G., Starughbaugh, L. D. & Berg, D. E. (1993). Transposon-facilitated sequencing of DNAs cloned in plasmids. *Methods in Enzymology* **218**, 279-306.

Bernardini, M. I., Mounier, J., d'Hauteville, H. D., Coquis-Rondon, M. & Sansonetti, P. J. (1989). Identification of *ics*A, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin. *Proceedings of the National Academy of Science, USA* **86**, 3867-3871.

Bhattacharyya, S. P. & Rao, V. B. (1993). A novel terminase activity associated with the DNA packaging protein gp17 of bacteriophage T4. *Virology* **196**, 34-44.

Bhattacharyya, S. P. & Rao, V. B. (1994). Structural analysis of DNA cleaved in vitro by bacteriophage T4 terminase. *Gene* **146**, 67-72.

Bishai, W. R. & Murphy, J. R. (1988). Bacteriophage gene products that cause human disease. In *The Bacteriophages* (Calendar, R., ed.), Vol. 2, pp. 683-724. Plenum Press, New York.

Black, L. (1974). Bacteriophage T4 internal protein mutants: isolation and properties. *Virology* **60**, 166-179.

Black, L. & Silverman, D. (1978). Model for DNA packaging into bacteriophage T4 heads. *Journal of Virology* **28**, 643-655.

Black, L. W. (1995). DNA packaging and cutting by phage terminases: control in phage T4 by a synaptic mechanism. *Bioessays* **17**(12), 1025-1030.

Black, R. E. (1993). Epidemiology of diarrhoeal disease: implications for control by vaccines. *Vaccine* **11**(2), 100-106.

Black, R. E., Graun, G. F. & Blake, P. A. (1978). Epidemiology of common source outbreaks of shigellosis in the United States 1961-1975. *American Journal of epidemiology* **108**, 47-52.

Black, R. E. e. a. (1982). Longitudinal studies of infectious diseases and physical growth in rural Bangladesh II. Incidence of diarrhea and association with known pathogens. *American Journal of Epidemiology* **115**, 315-324.

Blaser, M. J., Hale, T. L. & Formal, S. B. (1989). Recurrent shigellosis complicating human immunodeficiency virus infection: failure of pre-existing antibodies to confer protection. *American Journal of Medicine* **86**, 105-107.

Blattner, F. R., Plunkett III, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collando-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y.

(1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453-1462.

Bradley, D. E. (1967). Ultrastructure of bacteriophage and bacteriocins. *Bacteriology Reviews* **31**, 230-314.

Brahmbhatt, H. N., Lindberg, A. A. & Timmis, K. N. (1992). *Shigella* lipoploysaccharide: structure, genetics, and vaccine development. In *Pathogenesis of Shigellosis* (Sansonetti, P. J., ed.), pp. 45-64. Springer-Verlag, Berlin.

Brandtzaeg, P., Halstensen, T. S., Keh, K., Krajci, P., Kvale, D., Rognum, T. O., Scott, H. & Sollid, L. M. (1989). Immunobiology and Immunopathology of human gut mucosa: humoral immunity and intraepithelial lymphocytes. *Gastroenterology* **97**, 1562-1584.

Brito-Alayon, N. E., Blando, A. M. & Monzon-Moreno, C. (1994). Antibiotic resistance patterns and plasmid profiles for *Shigella spp.* isolated in Cordoba, Argentina. *Journal of Antimicrobial Chemotheraphy* **34**, 253-259.

Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., d'Hauteville, H., Kunst, F., Sansonetti, P. & Parsot, C. (2000). The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Molecular Microbiology* **38**(4), 760-771.

Bushman, F. D. (1993). The bacteriophage 434 right operator. Roles of O(R)1, O(R)2 and O(R)3. *Journal of Molecular Biology* **230**, 28-40.

Calendar, R., Yu, S., Myung, H., Barreiro, V., Odegrip, R., Carlson, K. & Davenport, L. (1998). The lysogenic conversion genes of coliphage P2 have unusually high AT content. In *Horizontal gene transfer* (Syvanen, M. & Kado, C., eds.), pp. 241-252. Chapman and Hall, London.

Cam, P. D., Pal, T. & Lindberg, A. A. (1993). Immune response against lipopolysaccharide and invasion plasmid-coded antigens of *shigellae* in

Vietnamese and Swedish dysenteric patients. *Journal of Clinical Microbiology* **31**(454-457).

Campbell, A. (1994). Comparative molecular biology of lambdoid phages. *Annual Reviews in Microbiology* **48**, 193-222.

Campbell, A. M. (1962). Episomes. *Advance Genetics* **11**, 101-145.

Campbell, A. M. (1992). Chromosomal insertion sites for phages and plasmids. *Journal of Bacteriology* **174**(23), 7495-7499.

Campbell, A. M. (1993). Thirty years ago in genetics: prophage insertion into bacterial chromosomes. *Genetics* **133**(3), 433-438.

Campbell, A. M. (1996). Cryptic prophages. In *Escherichia coli and Salmonella: cellular and molecular biology 2nd edition* (Neidhardt, F. C., Curtiss III, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaecter, M. & Umbarger, H. E., eds.). American Society for Microbiology, Washington, D.C.

Carpenter, C. C. J. (1982). Shigellosis. In *Cecil Textbook of Medicine* 16th edit. (Wyngaarden, J. B. & Smith, L. H., eds.), pp. 1517-1519. WB Saunders, Philadelphia.

Casjens, S., Hatfull, G. F. & Hendrix, R. W. (1992). Evolution of dsDNA tailed bacteriophage genomes. *Seminars in Virology* **3**, 383-397.

Casjens, S. & Hendrix, R. W. (1988). Control mechanisms in dsDNA bacteriophage assembly. In *The bacteriophages vol. I* (Calendar, R., ed.), Vol. 1, pp. 15-92. Plenum Press, New York.

Catalano, C. E., Cue, D. & Feiss, M. (1995). Virus DNA packaging: the strategy used by phage l. *Molecular Microbiology* **16**(6), 1075-1086.

Chandry, P. S., Moore, S. C., Boyce, J. D., Davidson, B. E. & Hillier, A. J. (1997). Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. *Molecular Microbiology* **26**, 49-64.

Cheetham, B. F. & Katz, M. E. (1995). A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Molecular Microbiology* **18**(2), 201-208.

Chen, Y., Smith, M. R., Thirumalai, K. & Zychlinsky, A. (1996). A bacterial invasin induces macrophage apoptosis by binding directly to ICE. *EMBO Journal* **15**, 3853-3860.

Chua, J. E., Manning, P. A. & Morona, R. (1999). The *Shigella flexneri* bacteriophage Sf6 tailspike protein (TSP)/endorhamnosidase i srelated to the bacteriophage P22 TSP and has a motiff common to exo- and endoglycanases, and C-5 epimerases. *Microbiology* **145**, 1649-1659.

Chung, C. H., Seol, J. H. & Kang, M. J. (1996). Protease Ti (Clp), a multicomponent ATP-dependent protease in *Escherichia coli*. *Journal of Biological Chemistry* **377**, 549-554.

Clark, C. A., Beltrame, J. & Manning, P. A. (1991). The *oac* gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. *Gene* **107**, 43-52.

Cleary, T. G., West, M. S., Ruiz-Palacios, G., Winsor, D. K., Calva, J. J., Guerrero, M. L. & Van, R. (1991). Human milk secetory immunoglobulin A to *Shigella* virulence plasmid-encoded antigens. *Journal of Pediatrics* **118**, 34-37.

Cleary, T. G., Winsor, D. K., Reich, D., Ruiz-Palacios, G. & Calva, J. J. (1989). Human milk immunoglobulin A antibodies to Shigella virulence determinants. *Infection and Immunity* **57**, 1675-1679.

Clerc, P. & Sansonetti, P. J. (1987). Entry of *Shigella flexneri* into *HeLa* cells, evidence for directed phagocytosis involving actin polymerization and myosin accumulation. *Infection and Immunity* **55**, 2681-2688.

Cohen, D., Green, M., Block, C., Slepon, R., Ambar, R., Wasserman, S. & Levine, M. (1991a). Reduction of transmission of Shigellosis by control of houseflies (*Musca domestica*). *Lancet* **337**, 993-997.

Cohen, D., Green, M. S., Block, C., Rouach, T. & Ofek, I. (1988). Serum antibodies to liposaccharides and natural immunity to shigellosis in an Israeli military population. *Journal of Infectious Diseases* **157**(5), 1068-1071.

Cohen, D., Green, M. S., Block, C., Slephon, R. & Lerman, Y. (1992). Natural immunity to Shigellosis in two groups with different previous risk of *Shigella* is only partly expressed by serum antibodies to lipopolysaccharides. *Journal of Infectious Diseases* **165**, 785-787.

Cohen, D., Green, M. S., Block, C., Slepon, R. & Ofek, I. (1991b). Prospective study of the association between serum antibodies to lipopolysaccharide O-antigen and the attack rate of shigellosis. *Journal of Clinical Microbiology* **29**(2), 386-389.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry III, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., Mc Lean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S., Sqares, R., Sulston, J. E., Taylor, K., Whitehead, S. & Barell, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**(6685), 537-544.

Collins, L. V. & Hackett, J. (1991). Molecular cloning, characterization and nucleotide sequence of the rfc gene, which encodes an O-antigen polymerase of *Salmonella typhimurium*. *Journal of Bacteriology* **173**, 2521-2529.

Davis, B. D., Dulbecco, R., Etsen, H. N. & Ginsberg, H. S. (1990). *Microbiology*. Fourth edition edit, J.B. Lippincott Company, Philadelphia.

Desiere, F., Lucchini, S. & Brussow, H. (1998). Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions. *Virology* **241**, 345-356.

Devine, S. E., Chissoe, S. L., Eby, Y., Wilson, R. K. & Boeke, J. D. (1997). A transposon-based strategy for sequencing repetitive DNA in eukaryotic genomes. *Genome Research* **7**, 551-563.

Dinari, G., Hale, T. L., Austin, S. W. & Formal, S. B. (1987). Local and systemic antibody responses to *Shigella* infection in rhesus monkeys. *Journal of Infectious Diseases* **155**(5), 1065-1069.

Dos Santos, A. L. L. & Chopin, A. (1987). Shotgun cloning in *Streptococcus lactis*. *FEMS Microbiology Letters* **42**, 209-212.

Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Research* **16**, 6127-6145.

Dube, P., Tavares, P., Lurz, R. & Heel, M. V. (1993). The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO Journal* **12**, 1303-1309.

Duda, R. L., Hempel, J., Michel, H., Shabanowitz, J., Hunt, D. & Hendrix, R. W. (1995a). Structural transitions during bacteriophage HK97 head assembly. *Journal of Molecular Biology* **247**, 618-635.

Duda, R. L. & Hendrix, R. W. (1998). bacteriophage HK97 head assembly: a protein ballet. *Advances in Virus Research* **50**, 235-288.

Duda, R. L., Martincic, K. & Hendrix, R. W. (1995b). Genetic basis of bacteriophage HK97 prohead assembly. *Journal of Molecular Biology* **247**, 636-647.

Duda, R. L., Martincic, K., Xie, Z. & Hendrix, R. W. (1995c). Bacteriophage HK97 head assembly. *FEMS Microbiology Review* **17**, 41-46.

Earnshaw, W. C. & Casjens, S. R. (1980). DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21**, 319-331.

Echeverria, P., Taylor, D. N., Leksomboon, U., Blacklow, N. R., Pinnoi, S., Nataro, J. P., Kaper, J. & Rowe, B. (1986). Identification of enteric pathogens in the small and large intestine of children with diarrhea. *Diagnostic Microbiology and Infectious Diseases* **4**, 277-284.

Ehara, M., Shimodori, S., Kojima, F., Ichinose, Y., Hirayama, T., Albert, J. M., Supawat, K., Honma, Y., Iwanaga, M. & Amako, K. (1997). Characterization of filamentous phages of *Vibrio cholerae* O139 and 01. *FEMS Microbiology Letters* **154**, 293-301.

Eiklid, K. & Olsnes, S. (1983). Animal toxicity for Shigella dysenteriae cytotoxin: evidence that the neurotoxic, enterotoxic and cytotoxic activities are due to one toxin. *Journal of Immunology* **130**, 380-384.

Eiserling, F. A. (1983). Structure of the T4 virion. In *Bacteriophage T4* (Matthews, C. K., Kutter, E. M., Mosig, M. & Berget, P. B., eds.), pp. 11-24. American Society for Microbiology, Washington D.C.

El-Rafie, M. e. a. (1990). Effect of diarrheal disease control on infant and childhood mortality in Egypt. *Lancet* **335**, 334-338.

Enquist, L. W. & Weisberg, R. A. (1984). An integration proficient int mutant of bacteriophage l. *Molecular General Genetics* **195**, 62-69.

Eppler, K., Wyckoff, E., Goates, J., Parr, R. & Casjens, S. (1991). Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging. *Virology* **183**, 519-538.

Feinberg, A. P. & Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**, 6-13.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., A.R., K., Bult, C. J., Tomb, J. F., Dougherty, B. A. & Merrick, J. M. (1995). The genome of *Haemophilus influenzae RD*. **269**, 496-512.

Formal, S. B., Gemski Jr., P., baron, L. S. & La Brec, E. H. (1970). Genetic transfer of *Shigella flexneri* antigens to *Escherichia coli* K-12. *Infection and Immunity* **1**, 279-287.

Franzon, V. L., Arondel, J. & Sansonetti, P. J. (1990). Contribution of superoxide dismutase and catalase activities to *Shigella flexneri* pathogenesis. *Infection and Immunity* **58**, 529-535.

Fraser, C. M. & Fleischmann, R. D. (1997). Strategies for whole microbial genome sequencing and analysis. *Electrophoresis* **18**, 1207-1216.

Fujisawa, H. & Hearing, P. (1994). Structure, function and specificity of the DNA packaging signals in double-stranded DNA viruses. *Seminars in Virology* **5**, 5-13.

Fuller, M. & King, J. (1982). Assembly in vitro of bacteriophage P22 procapsids from purified coat and scaffolding subunits. *Journal of Molecular Biology* **156**, 633-665.

Garcia, P., Alonso, J. C. & Suarez, J. E. (1997). Molecular analysis of the cos region of the *Lactobacillus casei* bacteriophage A2. Gene product 3, gp3, specifically binds to its downstream cos region. *Molecular Microbiology* **23**(3), 505-514.

Gemski Jr., P., Koeltzow, D. E. & Formal, S. B. (1975). Phage conversion of *Shigella flexneri* group antigens. *Infection and Immunity* **11**, 685-691.

Gemski Jr., P. & Stocker, B. A. D. (1967). Transduction by bacteriophage P22 in nonsmooth mutants of *Salmonella typhimurium*. *Journal of Bacteriology* **93**, 1588-1597.

Gerber, D. F. & Watkins, H. M. S. (1961). Growth of *Shigellae* in monolayer tissue cultures. *Journal of Bacteriology* **82**, 815-822.

Gertman, E., Berry, D. & Kropinski, A. M. (1987). Serotype-converting bacteriophage D3 of *Pseudomonas aeruginosa*: vegetative and prophage restriction maps. *Gene* **52**, 51-57.

Gilakjan, Z. A. & Kropinski, A. M. (1999). Cloning and analysis of the capsid morphogenesis genes of *Pseudomonas aeruginosa* bacteriophage D3: another example of protein chain mail? *Journal of Bacteriology* **181**(23), 7221-7227.

Gober, L. L., Friedman-Kien, A. E., Havell, E. A. & Vilcek, J. (1972). Suppression of the intracellular growth of *Shigella flexneri* in cell cultures by interferon preparations and polyinosinic-polycytidylic acid. *Infection and Immunity* **5**, 370-376.

Goldman, R. C., Joiner, K. & Leive, L. (1984). Serum resistant mutants of *Escherichia coli* 0111 contain increased lipopolysaccharide lack an O-antigen-containing capsule, and cover more of their lipid A core with O-antigen. *Journal of Bacteriology* **159**, 877-882.

Goliger, J. A. & Roberts, J. W. (1989). Sequences required for antitermination by phage 82 Q protein. *Journal of Molecular Biology* **210**, 461-471.

Gots, R. E., Formal, S. B. & Giannella, R. A. (1974). Indomethacin inhibition of *Salmonella typhimurium*, *Shigella flexneri*, and cholera-mediated rabbit ileal secretion. *Journal of Infectious Diseases* **130**, 280-284.

Griffiths, E., Stevenson, P., Hale, T. L. & Formal, S. B. (1985). Synthesis of aerobactin and a 76000-dalton iron-regulated outer membrane protein by *Escherichia coli* K-12 *Shigella flexneri* hybrids and by enteroinvasive strains of *Escherichia coli*. *Infection and Immunity* **49**, 67-71.

Groman, N. B. (1984). *Journal of Hygiene* **93**, 405-417.

Gross, R. J., Thomas, L. V. & Rowe, B. (1979). *Shigella dysenteriae*, *S. flexneri*, and *S. boydii* infections in England and Wales: the importance of foreign travel. *British Medical Journal* **280**, 744-750.

Guan, S., Bastin, D. A. & Verma, N. K. (1999). Functional analysis of the O-antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. *Microbiology* **145**, 1263-1273.

Guan, S. & Verma, N. K. (1998). Serotype conversion of a *Shigella flexneri* candidate vaccine strain via a novel site-specific chromosome-integration system. *FEMS Microbiology Letters* **166**, 79-87.

Guo, P., Ericson, S. & Anderson, D. (1987a). A small viral RNA is required for in vitro packaging of bacteriophage phi-29 DNA. *Science* **236**, 690-694.

Guo, P., Peterson, C. & Anderson, D. (1987b). Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage f29. *Journal of Molecular Biology* **197**, 229-236.

Haggard-Ljungquist, E., Halling, C. & Calendar, R. (1992). DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *Journal of Bacteriology* **174**, 1462-1477.

Hale, T. L. & Keren, D. F. (1992a). Pathogenesis and immunology in shigellosis: applications for vaccine development. In *Pathogenesis of Shigellosis* (Sansonetti, P. J., ed.), pp. 117-137. Springer-Verlag, Berlin.

Hale, T. L. & Keren, D. F. (1992b). *Pathogenesis and immunology in Shigellosis: applications for vaccine development*. In Pathogenesis of Shigellosis (Sansonetti, P. J., Ed.), Springer-Verlag.

Hale, T. L., Oaks, E. V. & Formal, S. B. (1985). Identification and antigenic characterization of virulence-associated, plasmid-coded proteins of *Shigella* spp and enteroinvasive *Escherichia coli*. *Infection and Immunity* **50**, 620-629.

Hardt, W. D., Urlaub, H. & Galan, J. E. (1998). A substrate of the centisome 63 type III protein secretion system of *Salmonella typhimurium* is encoded by a cryptic bacteriophage. *Proceedings of the National Academy of Sciences, USA* **95**, 2574-2579.

Harlow, E. & Lane, D. (1988). *Antibodies, a laboratory manual*, Cold Spring Harbor Laboratory Press, New York.

Hashimoto, C. & Fujisawa, H. (1992). DNA sequences necessary for packaging bacteriophage T3 DNA. *Virology* **187**, 788-795.

Hendrix, R. W. (1978). Symmetry mismatch and DNA packaging in large bacteriophages. *Proceedings of teh national Academy of Sciences, U.S.A.* **75**, 4779-4783.

Hendrix, R. W. & Duda, R. L. (1992). Lambda PaPa is not the mother of all phages. *Science* **258**, 1154-1158.

Hendrix, R. W. & Garcea, R. L. (1994). Capsid assembly of dsDNA viruses. *Seminars in Virology* **5**, 15-26.

Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E. & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proceedings of the National Academy of Science, USA* **96**, 2192-2197.

Henikoff, S. (1987). Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods in Enzymology* **155**, 156-165.

Herkowitz, I. & Hagen, D. (1980). The lysis-lysogeny decision of phage 1 : explicit programming and responsiveness. *Annual Review of Genetics* **14**, 394-445.

Hess, C. B., Niesel, D. W., Cho, Y. Z. & Klimpel, G. R. (1987). Bacterial invasion of fibroblast induces interferon production. *Journal of Immunology* **138**, 3949-3953.

Hess, C. B., Niesel, D. W. & Klimpel, G. R. (1989). The induction of interferon production in fibroblasts by invasive bacteria: a comparison of *Salmonella* and *Shigella* species. *Microbial Pathogenesis* **7**, 111-120.

Hong, M., Gleason, Y., Wyckoff, E. E. & Payne, S. M. (1998). Identification of two *Shigella flexneri* chromosomal loci involved in intercellular spreading. *Infection and Immunity* **66**(10), 4700-4710.

Hong, M. & Payne, S. M. (1997). Effect of mutations in *Shigella flexneri* chromosomal and plasmid-encoded lipopolysaccharaide genes on invasion and serum resistance. *Molecular Microbiology* **24**(4), 779-791.

Horwitz, J. P., Chua, J., Curby, R. J., Tomson, A. J., Da Rooge, M. A., Fisher, B. E., Mauricio, J. & Kundt, I. (1964). Substrates for cytochemical demonstration of enzyme activity I. some substituted 3-indolyl-B-D-glycopyranosides. *Journal of Medical Chemistry* **7**, 574.

Hosoda. (1970). Analysis of T4 proteins I. Conversion of precursor proteins into lower molecular weight peptides during normal capsid formation. *Proceedings of the National Academy of Science, U.S.A.* **66**, 1275-1281.

Huan, P. T., Bastin, D. A., Whittle, B. L., Lindberg, A. A. & Verma, N. K. (1997a). Molecular characterisation of the genes involved in O-antigen modification, attachment, integration and excision in *Shigella flexneri* bacteriophage SfV. *Gene* **195**, 217-227.

Huan, P. T., Taylor, R., Lindberg, A. A. & Verma, N. K. (1995). Immunogenicity of the *Shigella flexneri* serotype Y (SFL124) vaccine strain expressing cloned glucosyl transferase gene of converting bacteriophage SfX. *Microbiology and Immunology* **39**(7), 467-472.

Huan, P. T., Whittle, B. L., Bastin, D. A., Lindberg, A. A. & Verma, N. K. (1997b). *Shigella flexneri* type-specific antigen V: cloning, sequencing and characterization of the glucosyl transferase gene of temperate bacteriophage SfV. *Gene* **195**, 207-216.

Huan, P. T., Whittle, B. L., Guan, S., Lindberg, A. A. & Verma, N. K. (1997c). *Shigella flexneri* serotype-converting bacteriophage SfV and its relationship with bacteriophage SfX. In *Glucosyl transferase gene of temperate bacteriophages SfX and SfV and their role in the modification of the O-antigen of Shigella flexneri*, pp. section V. Karolinska Institute Huddinge University Hospital, Stockholm.

Hueck, C. J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and Molecular Biology Reviews* **62**(2), 379-433.

Hulton, C. S. J., Seirafi, A., Hinton, J. C. D., Sidebotham, J. M., Waddell, L., Pavitt, G. D., Owen-Hughes, T., Spassky, A., Buc, H. & Higgins, C. F. (1990). Histone-like protein H1 (H-NS), DNA supercoiling and gene expression in bacteria. *cell* **63**, 631-642.

Hunkapiller, T., Kaiser, R. J., Koop, B. F. & Hood, L. (1991). Large-scale and automated DNA sequence determination. *Science* **254**, 59-67.

Hynes, W. L., Hancock, L. & Ferretti, J. J. (1995). Analysis of a second bacteriophage hyaluronidase gene from *Streptococcus pyogenes*: evidence for a third hyaluronidase involved in extracellular enzymatic activity. *Infection and Immunity* **63**, 3015-3020.

Izhar, M., Nuchamowitz, Y. & Mirelman, D. (1982). Adherence of *Shigella flexneri* to guinea pig intestinal cells is mediated by a mucosal adhesin. *Infection and immunity* **35**, 1110-1118.

Jepson, M. A. & Clark, M. A. (1998). Studying M cells and their role in infection. *Trends in Microbiology* **6**, 359-365.

Jiang, X. M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K. & Reeves, P. R. Structure and sequence of the rfb antigen gene cluster of *Salmonella* serovar *typhimurium* (strain LT2). *Molecular Microbiology* **5**(3), 695-713.

Joiner, K. A., Grossman, N., Schmetz, M. & Leive, L. (1986). C3 binds preferentially to long chain lipopolysaccharide during alternative pathway activation by *Salmonella montevideo*. *Journal of Immunology* **136**, 710-715.

Kadurugamuwa, J. L., Rhode, M., Wehland, J. & Timmis, K. N. (1991). Intercellular spread of *Shigella flexneri* through a monolayer mediated by membranous protrusions and associated reorganization of the cytoskeletal protein vinculin. *Infection and Immunity* **59**, 3463-3471.

Kaneko, J., Kimura, T., Kawakami, Y., Tomita, T. & Kamio, Y. (1997). Panton-Valentine leukocidin genes in phage-like particle isolated from mitomycin-C treated *Staphylococcus aureus* V8 (ATCC 49775). *Bioscience Biotechnology Biochemistry* **61**, 1960-1962.

Kaneko, J., Kimura, T., Narita, S., Tomita, T. & Kanno, Y. (1998). Complete nucleotide sequence and molecular characterization of the temperate Staphylococcal bacteriophage wPVL carrying Panton-Valentine leukocidin genes. *Gene* **215**, 57-67.

Karaolis, D. K. R., Somara, S., Maneval Jr., D. R., Johnson, J. A. & Kaper, J. B. (1999). A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**, 375-379.

Karnell, A., Stocker, B. A. D., Katakura, S., Reinholt, F. P. & Lindberg, A. A. (1992a). Live oral auxotrophic *Shigella flexneri* SFL124 vaccine with a deleted *aro*D gene: Characterisation and monkey protection studies. *Vaccine* **10**(6), 389-394.

Karnell, A., Sweiha, H. & Lindberg, A. A. (1992b). Auxotrophic live oral *Shigella flexneri* vaccine protects monkeys against challenge with *S. flexneri* of different serotypes. *Vaccine* **10**, 167-174.

Kasai, H., Isono, S., Kitakawa, M., Mineno, J., Akiyama, H., Kurnit, D. M., Berg, D. E. & Isono, K. (1992). Efficient large-scale sequencingof the *Escherichia coli* genome: Implementation of a transposon and PCR-based strategy for the analysis of ordered lambda phage clones. *Nucleic Acids Research* **20**, 6509-6515.

Katouli, M. e. a. (1990). Aetiological studies of diarrhoeal diseases in infants and young children in Iran. *Journal of Tropical Medicine and Hygiene* **93**, 22-27.

Katsushi, Y., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C. H., Kurokawa, K. & Ishii, K. e. a. (2000). Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* 0157:H7 strain derived from the Sakai outbreak. *Gene* **258**, 127-139.

Kenne, L., Lindberg, B., Petersson, K. & Romanowska, E. (1977). Basic structure of the oligosaccharide repeating unit of the *Shigella flexneri* O-antigen. *Car Research* **56**, 363-370.

Kenne, L., Lindberg, B., Petesson, K., Katzenellenbogen, E. & Romanowska, E. (1978). Structural studies of *Shigella flexneri* O-antigens. *European Journal of Biochemistry* **91**, 279-284.

Keren, D. F., Brown, J. E., Mc Donald, R. A. & Wassef, J. S. (1989b). Secretory immunoglobulin A response to Shiga toxin in rabbits: kinetics of the initial mucosal immune response and inhibition of toxicity in vitro and in vivo. *Infection and Immunity* **57**, 1885-1889.

Keren, D. F., McDonald, R. A., Wassef, J. S., Armstrong, L. R. & Brown, J. E. (1989a). The enteric immune response to *Shigella* antigens. *Current Topics in Microbiology and Immunology* **146**, 213-223.

Keusch, G. T. & Jacewicz, M. (1977). The pathogenesis of *Shigella* diarrhea VI. Toxin and antitoxin in *Shigella flexneri* and *Shigella sonnei* infections in humans. *Journal of Infectious Diseases* **135**, 552-556.

Khan, M. U. e. a. (1985). Fourteen years of shigellosis in Dhaka: an epidemiological analysis. *International Journal of Epidemiology* **14**, 607-613.

Kim, S. & Landy, A. (1992). Lambda Int protein bridges between higher order complexes at two distant chromosomal loci *attL* and *attR*. *Science* **256**, 198-203.

King, J. & Casjens, S. (1974). Catalytic head assembling protein in virus morphogenesis. *Nature* **251**, 112-119.

Klena, J. D. & Schnaitman, C. A. (1993). Function of the rfb gene cluster and the rfe gene in the synthesis of O-antigen by *Shigella dysenteriae* 1. *Molecular Microbiology* **9**(2), 393-402.

Kobayashi, K., Okamura, K., Inoue, T., Sato, T. & Kobayashi, Y. (1998). Complete nucleotide sequence of *Bacillus subtilis* phage phi-105. .

Koneman, E. W., Allen, S. D., Dowell Jr., V. R., Janda, W. M., Sommers, H. M. & Winn Jr., W. C. (1988). *Color atlas and textbook of diagnostic microbiology*. Taxonomy of enterobacteriaceae, J.B. Lippincott Company, Philadelphia.

Kopecko, D. J., Washington, O. & Formal, S. B. (1980). Genetic and physical evidence for plasmid control of *Shigella sonnei* form I surface antigen. *Infection and Immunity* **29**, 207-214.

Kotloff, K. L., Herrington, D. A., Hale, T. L., Newland, J. W., van de Verg, L., Cogan, J. P., Snoy, P. J., Sadoff, J. C., Formal, S. B. & Levine, M. M. (1992). Safety, immunogenicity and efficacy in monkeys and humans of invasive *Escherichia coli* K-12 vaccine candidates expressing *Shigella flexneri* 2a somatic antigen. *Infection and Immunity* **60**, 2218-2224.

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti, P. J., Adak, G. K. & Levine, M. M. (1999). Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bulletin of the World Health Organisation* **77**(8), 651-666.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertoro, M. G., Bessieres, P., Bolotin, A. & Borchert, S. e. a. (1997). *Nature* **390**, 249-256.

LaBrec, E. H., Schneider, H., T.J., M. & Formal, S. B. (1964). Epithelial cell penetration as an essential step in pathogenesis of bacillary dysentery. *Journal of Bacteriology* **88**, 1503-1518.

Lawlor, K. M., Daskaleros, P. A., Robinson, R. E. & Payne, S. M. (1987). Virulence of iron transport mutants of *Shigella flexneri* and utilization of host iron compounds. *Infection and Immunity* **55**, 594-599.

Lawlor, K. M. & Payne, S. M. (1984). Aerobactin genes in *Shigella spp. Journal of Bacteriology* **160**, 266-272.

Lazar, S. & Waldor, M. K. (1998). ToxR-independent expression of cholera toxin from the replicative form of phi-CTX. *Infection and Immunity* **66**, 394-397.

Leffers, G. & Rao, V. B. (1996). A discontinuous headful packaging model for packaging less than headful length DNA molecules by bacteriophage T4. *Journal of Molecular Biology* **258**, 839-850.

Leffers, J., G.G. & Gottesman, S. (1998). Lambda *xis* degradation in vivo by *Lon* and *FtsH*. *Journal of Bacteriology* **180**(6), 1573-1577.

Leong, J. M., Nunes-Duby, S. E., Oser, A. B., Lesser, C. F., Youderian, P., Susskind, M. M. & Landy, A. (1986). Structural and regulatory divergence among site-specific recombination genes of lambdoid phage. *Journal of Molecular Biology* **189**, 603-616.

Leong, J. M., Nunes-Dueby, S., Lesser, C. F., Youderian, P., Susskind, M. M. & Landy, A. (1985). The phi-80 and P22 attachment sites: primary structure and interaction with *Escherichia coli* integration host factor. *Journal of Biological Chemistry* **260**, 4468-4477.

Li, A., Cam, P. D., Islam, D., Minh, N. B., Huan, P. T., Rong, Z. C., Karlsson, K., Lindberg, G. & Lindberg, A. A. (1994). Immune responses in Vietnamese children after a single dose of the auxotrophic, live *Shigella flexneri* Y vaccine strain SFL124. *J Infect* **28**(1), 11-23.

Li, A., Pal, T., Forsum, U. & Lindberg, A. A. (1992). Safety and immunogenicity of the live oral auxotrophic *Shigella flexneri* SFL124 in volunteers. *Vaccine* **10**(6), 395-404.

Lin, H., Rao, V. B. & Black, L. W. (1999). Analysis of capsid, portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4. *Journal of Molecular Biology* **289**, 249-260.

Lin, S. R. & Chang, S. F. (1992). Drug resistance and plasmid profile of *Shigellae* in taiwan. *Epidemiology and Infection* **108**, 87-97.

Lindberg, A. A. (1973). Bacteriophage receptor. *Annual Reviews in Microbiology* **27**, 205-237.

Lindberg, A. A. (1977). *Bacterial surface polysaccharides and phage adsorption*, Academic press, New York.

Lindberg, A. A., Brown, J. E., Stromberg, N., Westling-Ryd, M., Schultz, J. E. & Karlsson, K. A. (1987). Identification of the carbohydrate receptor for Shiga toxin produced by *Shigella dysenteriae* type 1. *Journal of Biological chemistry* **262**, 1779-1786.

Lindberg, A. A., Cam, P. D., Chan, N., Phu, L. K., Trach, D. D., Lindberg, G., Karlsson, K., Karnell, A. & Eckwall, E. (1991). Shigellosis in Vietnam: seroepidemiologic studies with use of lipopolysaccharide antigens in enzyme immunoassay. *Reviews in Infectious Diseases* **13**(4), S231-S237.

Lindberg, A. A., haeggman, S., Karlsson, K., Cam, P. D. & Trach, D. D. (1984). The humoral antibody response to *Shigella dysenteriae* type 1 infection as determined by ELISA. *WHO Bulletin* **62**, 597-606.

Lindberg, A. A. & Pal, T. (1993). Stategies for development of potential candidate *Shigella* vaccines. *Vaccine* **11**, 168-179.

Lindberg, A. A., Wollin, R., Gemski, P. & Wohlhieter, J. A. (1978). Interaction between bacteriophage Sf6 and *Shigella flexneri*. *Journal of Virology* **27**, 38-44.

Lindberg, B., Lonngren, J., Romanowska, E. & Ruden, U. (1972). Location of O-acetyl groups in *Shigella flexneri* type 3c and 4b lipopolysaccharides. *Acta Chemica Scandinavia* **26**, 3808-3810.

Lowell, G. H., MacDermott, R. P., Summers, P. L., Reeder, A. A., Bertovich, M. J. & Formal, S. B. (1980). Antibody-dependent cell-mediated antibacterial activity: K lymphocytes, monocytes, and granulocytes are effective against *Shigella*. *Journal of Immunology* **125**(6), 2778-2784.

Lucchini, S., Desiere, F. & Brussow, H. (1998). The structural gene module in *Streptococcus thermophilus* bacteriophage Sfi11 shows a hierarchy of relatedness to Siphoviridae from a wide range of bacterial hosts. *Virology* **246**, 63-73.

Lugtenberg, B. & van Alphen, L. (1983). Molecular architecture and functioning of the outer membrane of *Escherichia coli* and other gram-negative bacteria. *Biochemica, Biophysica Acta* **737**, 51-115.

Macaden, R., Gokul, B. N., Pereira, P. & Bhat, P. (1980). Bacillary dysentery due to multidrug resistant *Shigella dysenteriae 1*. *Indian Journal of Medical Research* **71**, 178-185.

Mahdi, A. A., Sharples, G. J., Mandal, T. N. & Lloyd, R. G. (1996). Holliday junction resolvases encoded by homologous RusA genes in *Escherichia coli* K-12 and phage 82. *Journal of Molecular Biology*.

Makela, P. H. (1966). Genetic determination of the O-antigens of *Salmonella* groups B and D. *Journal of Bacteriology* **91**, 1115-1125.

Makela, P. H. (1973). Glucosalation of lipopolysaccharide in *Salmonella*: mutants negative for O-antigen factor 12_2. *Journal of Bacteriology* **116**, 847-856.

Makela, P. H. & Stocker, B. A. D. (1984). Genetics of lipopolysaccharide. In *Chemistry of Endotoxin* (Rietschel, E. T., ed.), Vol. 1, pp. 419. 4 vols. Elsevier Science Publishers B.V,, Amsterdam.

Makino, S., Sasakawa, C., Kamata, K., Kurata, T. & Yoshikawa, M. (1986). A genetic determinant required for continuous reinfection of adjacent cells on large plasmid in *Shigella flexneri* 2a. *Cell* **46**, 551-555.

Marvik, O. J., Dokland, T., Nokling, R. H., Jacobsen, E., Larsen, T. & Lindqvist, B. H. (1995). The capsid size-determining protein *sid* forms an external scaffold on phage P4 procapsids. *Journal of Molecular Biology* **251**(1), 59-75.

Marvik, O. J., Sharma, P., Dokland, T. & Lindqvist, B. H. (1994). Bacteriophage P2 and P4 assembly: alternative scaffolding proteins regulate capsid size. *Virology* **200**, 702-714.

Mata, L. J., Gangarosa, E. J., Caceres, A., Perera, D. R. & Mejicanos, M. L. (1969a). Epidemic Shiga bacillus dysentery in Central America. I. Etiologic investigations in Guatemala. *Journal of Infectious Diseases* **122**, 170-180.

Mata, L. J., Gangarosa, E. J., Caceres, A., Perera, D. R. & Mejicanos, M. L. (1969b). Epidemic shiga bacillus in Central America. Etiologic investigations in Guatemala. *Journal of Infectious Diseases* **122**, 170-180.

Mathan, V. I., Bhat, P., Kapadia, C. R., Ponniah, J. & Baker, S. J. (1984). Epidemic dysentery caused by the Shiga bacillus in a southern Indian village. *Journal of Diarrhoeal Disease Research* **1**, 27-32.

Matsuzaki, S., Inoue, T., Kuroda, M., Kimura, S. & Tanaka, S. (1998a). Cloning ang sequencing of major capsid protein (mcp) gene of a vibriophage, KVP20, possibly related to T-even coliphages. *Gene* **222**, 25-30.

Matsuzaki, S., Inoue, T. & Tanaka, S. (1998b). A vibriophage, KVP40, with major capsid protein homologous to gp23 of coliphage T4. *Virology* **242**, 314-318.

Matsuzaki, S., Tanaka, S., Koga, T. & Kawata, T. (1992). A broad host-range vibriophage KVP40 isoalted from sea water. *Microbiology and Immunology* **36**, 93-97.

Maurelli, A. T., Baudry, B., d'Hauteville, H. & Sansonetti, P. J. (1985). Cloning of plasmid DNA sequences involved in invasion of HeLa cells by *Shigella flexneri*. *Infection and Immunity* **49**, 164-171.

Maurelli, A. T., Blackmon, B. & Curtiss III, R. (1984b). Loss of pigmentation in *Shigella flexneri* 2a is correlated with loss of virulence and virulence-associated plasmid. *Infection and Immunity* **43**, 397-401.

Maurelli, A. T., Blackmon, B. & Curtiss, R. (1984a). Tempertaure-dependent expression of virulence genes in *Shigella* species. *Infection and Immunity* **43**, 195-201.

Maurelli, A. T. & Sansonetti, P. J. (1988). Identification of a chromosomal gene controlling temperature-regulated expression of *Shigella* virulence. *Proceedings of the national Academy of Sciences, U.S.A.* **85**, 2820-2824.

Maurizi, M. R., Clark, W. P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B. & Gottesman, S. (1990). Sequence and structure of ClpP, the proteolytic component of the ATP-dependent Clp protease of *Escherichia coli*. *Journal of Biological Chemistry* **265**, 12536-12545.

Mavris, M., Manning, P. A. & Morona, R. (1997). Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri*. *Molecular Microbiology* **26**(5), 939-950.

McGrath, B. C. & Osborn, M. J. (1991). Localization of the terminal steps of O-antigen synthesis in *Salmonella typhimurium*. *Journal of Bacteriology* **173**, 649-654.

Meier-Dieter, U., Barr, K., Starman, R., Hatch, L. & Rick, P. D. (1992). Nucleotide sequence of *Escherichia coli rfe* gene involved in the synthesis of enterobacteriaceae common antigen. Molecular cloning of the *rfe-rff* gene cluster. *Journal of Biological Chemistry* **267**, 746-753.

Menard, R., Sansonetti, P. J. & Parsot, C. (1993). Nonpolar mutagenesis of the *ipa* genes defines IpaB, IpaC, and IpaD as effectors of *Shigella flexneri* entry into epithelial cells. *Journal of Bacteriology* **175**(18), 5899-5906.

Menard, R., Sansonetti, P. J., Parsot, C. & Vasselon, T. (1994). Extracellular association and cytoplasmic partitioning of the IpaB and IpaC invasins of *Shigella flexneri*. *Cell* **79**, 515-525.

Merril, C. R. & Goldman, D. (1984). Detection of polypeptides in two-dimensional gels using silver stain. In *Two-dimensional Gel Electrophoresis of Proteins* (Celis, J. E. & Bravo, R., eds.), pp. 93. Academic Press, New York.

Merril, C. R., Goldman, D., Sedman, S. A. & Ebert, M. H. (1981). Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* **211**, 1437.

Merril, C. R., Goldman, D. & Van Keuren, M. L. (1982). Simplified silver protein detection and image enhancement methods in polyacrylamide gels. *Electrophoresis* **3**, 17.

Messing, J., Crea, R. & Seeburg, P. H. (1981). A system for shotgun DNA sequencing. *Nucleic Acids Research* **9**, 309.

Mesyanzhinov, V. V., Sobolev, B. N., Marusich, E. I., Prilipov, A. G. & Efimov, V. P. (1990). A proposed structure of bacteriophage T4 gene product 22-a major prohead scaffolding core protein. *Journal of Structural Biology* **104**, 24-31.

Miao, E. A. & Miller, S. I. (1999). Bacteriophages in the evolution of pathogen-host interactions. *Proceedings of the National Academy of Sciences, USA* **96**, 9452-9454.

Miller, H. I., Mozola, M. A. & Friedman, D. I. (1980). An *E. coli* gene product required for I site specific recombination. *Cell* **20**, 721-729.

Miller, R. V., Pemberton, J. M. & Richards, K. E. (1974). F116, D3, and G101: temperate bacteriophages of *Pseudomonas aeruginosa*. *Virology* **59**, 566-569.

Mims, C., Playfair, J., Roitt, I., Wakelin, D. & Williams, R. (1998). *Medical Microbiology*. 2nd edit. Section 4, Mosby International Ltd., London.

Mims, C. A., Dimmock, N. J., Nash, A. & Stephen, J. (1995). The immune response to infection. In *Mims' Pathogenesis of Infectious Diseases*, pp. 197-277. Academic Press, San Diego, California.

Monod, C., Repoila, F., Kutateladze, M., Tetart, F. & Krisch, H. M. (1997). The genome of the pseudo T-even bacteriophages, a diverse group that resemble T4. *Journal of Molecular Biology* **267**, 237-249.

Morita, M., Tasaka, M. & Fujisawa, H. (1995). Analysis of the fine structure of the prohead binding domain of the packaging protein of bacteriophage T3 using a hexapeptide, an analog of a prohead binding site. *Virology* **211**, 516-524.

Morona, R., Mavris, M., Fallarino, A. & Manning, P. A. (1994). Characterization of the rfc region of *Shigella flexneri*. *Jounal of Bacteriology* **176**(3), 733-747.

Mosley, W. H., Adams, B. & Lyman, E. D. (1962). Epidemiologic and sociologic features of a large urban outbreak of shigellosis. *Journal of American Medical Association* **182**, 1307-1312.

Murayama, S. Y., Sakai, T., Makino, S., Kurota, T., Sasakawa, C. & Yoshikawa, M. (1986). The use of mice in the Sereny test as a virulence assay of *Shigella* and enteroinvasive *Escherichia coli*. *Infection and Immunity* **51**, 696-698.

Murray, B. E. (1986). Resistance of *Shigella, Salmonella* and other selected enteric pathogens to antimicrobial agents. *Reviews of Infectious Diseases* **8**, S172-S181.

Nakayama, K., Kanaya, S., Ohnishi, M., Terawaki, Y. & Hayashi, T. (1999). The complete nucleotide sequence of fCTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages. *Molecular Microbiology* **31**(2), 399-419.

Neutra, M. R. (1999a). M cells in antigen sampling in mucosal tissues. *Current Topics in Microbiology and Immunology* **236**, 17-32.

Neutra, M. R., Mantis, N. J., Frey, A. & Giannasca, P. J. (1999b). The composition and function of M cell apical membranes: implications for microbial pathogenesis. *Seminars in Immunology* **11**, 171-181.

Nhieu, G. T. V., Ben-Zeev, A. & Sansonetti, P. J. (1997). Modulation of bacterial entry into epithelial cells by association between vinculin and the *Shigella* IpaA invasin. *EMBO Journal* **16**(10), 2717-2729.

Niesel, D. W., Hess, C. B., Cho, Y. J., Klimpel, K. D. & Klimpel, G. R. (1986). Natural and recombinant interferons inhibit epithelial cell invasion by *Shigella* spp. *Infection and Immunity* **52**(3), 828-833.

Nikaido, H., Nikaido, N., Nakae, T. & Makela, P. H. (1971). Glucosylation of lipopolysaccharide in *Salmonella*: biosynthesis of O-antigen factor 12_2. I. Overall reaction. *Journal of Biological Chemistry* **246**, 3902-3911.

O' Brien, A. D., La Veck, G. D., Thompson, M. R. & Formal, S. B. (1982). Production of *Shigella dysenteriae* type 1-like cytotoxin by *Escherichia coli*. *Journal of infectious diseases* **146**, 763-769.

O' Gorman, S., Fox, D. T. & Wahl, G. M. (1991). Recombinase- mediated gene activation and site-specific integration in mammalian cells. *Science* **251**(4999), 1351-1355.

O'Brien, A. D., Newland, J. W., Miller, S. F., Holmes, R. K., Smith, H. W. & Formal, S. B. (1984). Shiga-like toxin-converting phage from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* **226**, 694-696.

Oaks, E. V., Hale, T. L. & Formal, S. B. (1986). Serum immune response to *Shigella* protein antigens in rhesus monkeys and human infected with *Shigella* spp. *Infection and Immunity* **53**(1), 57-63.

Oaks, E. V., Wingfield, M. E. & Formal, S. B. (1985). Plaque formation by virulent *Shigella flexneri*. *Infection and Immunity* **48**, 124-129.

Oberhelman, R. A., Kopecko, D. J., Salazar-Lindo, E., Gotuzzo, E., Buysse, J. M., Venkatesan, M. M., Yi, A., Fernandez-Prada, C., Guzman, M., Leon-Barua, R. & Sack, R. B. (1991). Prospective study of systemic and mucosal immune responses in dysenteric patients to specific *Shigella* invasion plasmid antigens and lipopolysaccharides. *Infect Immun* **59**(7), 2341-2350.

Ogawa, H., Nakamura, A. & Nakaya, R. (1968). Cinematographic studies of tissue cell cultures infected with *Shigella flexneri*. *Japanese Journal of Medical Science and Biology* **21**, 259-273.

Okada, N., Sasakawa, C., Tobe, T., Talukder, K. A., Komatsu, K. & Yoshikawa, M. (1991b). Construction of a physical map of the chromosome of *Shigella flexneri* 2a and the direct assignment of nine virulence-associated loci identified by Tn5 insertions. *Molecular Microbiology* **5**, 2171-2180.

Okada, N., Sasakawa, C., Tobe, T., Yamada, M., Nagai, S., Talukder, K. A., Komatsu, K., Kanegasaki & Yoshikawa, M. (1991a). Virulence-associated chromosomal loci of *Shigella flexneri* identified by random Tn*5* insertion mutagenesis. *Molecular Microbiology* **5**(1), 187-195.

Okamura, N., Nagei, T., Nakaya, R., Kondo, S., Murakami, M. & Hisatsune, K. (1983). HeLa cell invasiveness and O-antigen of *Shigella flexneri* as separate and

prerequisite attributes of virulence to evoke keratoconjunctivitis in guinea pigs. *Infection and Immunity* **39**, 505-513.

Okamura, N. & Nakaya, R. (1977). Rough mutants of *Shigella flexneri* 2a that penetrate tissue culture cells but does not evoke keratoconjunctivitis in guinea pigs. *Infection and immunity* **17**, 4-8.

Olsnes, S. & Eiklid, K. (1980). Isolation and characterization of *Shigella* shiga cytotoxin. *Journal of Biological Chemistry* **255**, 284-289.

Osborn, M. J., Cynkin, M. A., Gilbert, J. M., Muller, L. & Singh, M. (1972). Synthesis of bacterial O-antigen. *Methods in Enzymology* **28**, 583-601.

Parker, C. T., Kloser, A. W., Schnaitman, C. A., Stein, M. A., Gottesman, S. & Gibson, B. W. (1992). Role of the *rfaG* and *rfaP* genes in determining the lipopolysaccharide core structure and cell surface properties of *Escherichia coli* K-12. *Journal of Bacteriology* **174**, 2525-2538.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence analysis. *Proceedings of the National Academy of Sciences* **85**, 2444-2448.

Penn, C. W. (1983). Bacterial envelope and humoral defences. In *Role of the Envelope in the Survival of Bacteria in Infection* (Easmon, C. S. F., Jeljaszewicz, J., Brown, M. R. W. & Lambert, P. A., eds.), Vol. 3, pp. 109-135. Academic, London.

Perdomo, J. J., Gounon, P. & Sansonetti, P. J. (1994a). Polymorphonuclear leukocyte transmigration promotes invasion of colonic epithelial monolayer by *Shigella flexneri*. *Journal of Clinical Investigation* **93**(2), 633-643.

Perdomo, O. J. J., Cavaillon, J. M., Huerre, M., Ohayon, H., Gounon, P. & Sansonetti, P. J. (1994b). Acute inflammation causes epithelial invasion and mucosal destruction in experimental shigellosis. *Journal of Experimental Medicine* **180**, 1307-1319.

Peterson, G. L. (1977). A simplification of the protein assay method of Lowry et al., which is more generally applicable. *Analytical Biochemistry* **83**(83), 346-356.

Petrovskaya, V. G. & Licheva, T. A. (1982). A provisional chromosome map of *Shigella* and the regions related to pathogenicity. *Acta Microbiologica Academy of Sciences, Hungary* **29**, 41-53.

Pluschke, G., Mayden, J., Achtman, M. & Levine, R. P. (1983). Role of the capsule and the O-antigen in resistance of 018:K1 *Escherichia coli* to complement-mediated killing. *Infection and Immunity* **42**, 907-913.

Poteete, A. R. (1988). Bacteriophage P22. In *The bacteriophages*, Vol. vol. 2, pp. 647-680. Plenum Press, New York.

Powell, D., Franklin, J., Arisaka, F. & Mosig, G. (1990). Bacteriophage T4 DNA packaging genes 16 and 17. *Nucleic Acids Research* **18**, 4005.

Prevelige, P., Thomas, D. & King, J. (1993). Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Journal of Biophysics* **64**, 824-835.

Prevost, M. C., Lesourd, M., Arpin, M., Vernel, F., Mounier, J., Hellio, R. & Sansonetti, P. J. (1992). Unipolar reorganization of F-actin layer at bacterial division and bundling of actin filaments by plastin correlate with movement of *Shigella flexneri* within HeLa cells. *Infection and Immunity* **60**, 4088-4099.

Ptashne, M. (1986). *A genetic switch*, Cell Blackwell, Cambridge, MA.

Raetz, C. R. H. (1990). Biochemistry of endotoxins. *Annual Reviews of Biochemistry* **59**, 129-170.

Rahaman, M. M., Khan, M. M., Aziz, K. M. S., Islam, M. S. & Kibriya, A. K. M. G. (1975). An outbreak of dysentery caused by *Shigella dysenteriae* type 1 on a coral island in the Bay of Bengal. *Journal of Infectious Diseases* **132**, 15-19.

Rajakumar, K., Jost, B. H., Sasakawa, C., Okada, N., Yoshikawa, M. & Adler, B. (1994). Nucleotide sequence of the rhamnose biosynthetic operon of *Shigella flexneri* 2a and role of lipopolysaccharide in virulence. *Journal of Bacteriology* **176**(8), 2362-2373.

Reed, W. P. & Williams, R. C. (1971). Intestinal immunoglobulins in shigellosis. *Gastroenterology* **61**(1), 35-45.

Reiter, W. D., Palm, P. & Yeats, S. (1989). Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Research* **17**, 1907-1914.

Robbins, P. W. & Uchida, T. (1962). Studies on the chemical basis of the phage conversion of O-antigen in the E-group *Salmonella*. *Journal of bacteriology* **99**, 513-519.

Robbins, P. W. & Wright, A. (1971). *Biosynthesis of O-antigen*. Microbial toxins (Weinbaum, G., kadis, S. & Ajl, S. J., Eds.), 4, Academic, New York.

Roitt, I. M. (1997). *Roitt's Essential Immunology*. 9th edit, Blackwell Science, Malden, M.A.

Rosner, J. L. (1972). Formation, induction and curing of bacteriophage P1 lysogens. *Virology* **48**, 679-689.

Rout, W. R., Formal, S. B. & Gianella, R. A. (1975). Pathophysiology of *Shigella* diarrhea in the rhesus monkey; intestinal transport, morphological and bacteriological studies. *Gastroenterology* **68**, 270-278.

Sack, R. B. e. a. (1997). Antimicrobial resistance in organisms causing diarrheal disease. *Clinical Infectious Diseases* **24**, S102-S105.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. 2nd edit, Cold Spring Harbour Laboratory, New York.

Sandlin, R. C., Goldberg, M. B. & Maurelli, A. T. (1996). Effect of O side-chain length and composition on the virulence of *Shigella flexneri* 2a. *Molecular Microbiology* **22**, 63-73.

Sandlin, R. C., Lampel, K. A., Keasler, S. P., Goldberg, M. B., Stolzer, A. L. & Maurelli, A. T. (1995). Avirulence of rough mutants of *Shigella flexneri*: requirement of O-antigen for unique polar localization of IcsA in the bacterial outer membrane. *Infection and Immunity* **63**, 229-237.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science, U.S.A.* **74**, 5463-5467.

Sansonetti, P. J. (1992). Molecular and cellular biology of *Shigella flexneri* invasiveness: From cell assay systems to shigellosis. In *Pathogenesis of Shigellosis* (Sansonetti, P. J., ed.), pp. 1-19. Springer-Verlag, Berlin.

Sansonetti, P. J. & Arondel, J. (1989). Construction and evaluation of a double mutant of *Shigella flexneri* as a candidate for oral vaccination. *Vaccine* **7**, 443-450.

Sansonetti, P. J., Arondel, J., Cavaillon, J. M. & Huerre, M. (1995). Role of IL-1 in the pathogenesis of experimental shigellosis. *Journal of Clinical Investigation* **96**, 884-892.

Sansonetti, P. J., Arondel, J., Fontaine, A., d'Hauteville, H. & Bernardini, M. L. (1991). *Omp*B (osmo-regulation) and *ics*A (cell-to-cell spread) mutants of *Shigella flexneri*: vaccine candidates and probes to study the pathogenesis of shigellosis. *Vaccine* **9**, 416-422.

Sansonetti, P. J., Hale, T. L., Dammin, G. J., Kapfer, C., Colins, H. & Formal, S. (1983). Alterations in the pathogenicity of *Escherichia coli* K-12 after transfer of plasmid and chromosomal genes from *Shigella flexneri*. *Infection and Immunity* **39**, 1392-1402.

Sansonetti, P. J., Kopecko, D. J. & Formal, S. B. (1982). Involvement of a large plasmid in the invasive ability of *Shigella flexneri*. *Infection and Immunity* **35**, 852-860.

Sansonetti, P. J. & Phalipon, A. (1999). M cells as ports of entry for enteroinvasive pathogens: Mechanisms of interaction. consequences for the disease process. *Seminars in Immunology* **11**, 193-203.

Sansonetti, P. J., Ryter, A., Clerc, P., Maurelli, A. T. & Mounier, J. (1986). Multiplication of *Shigella flexneri* within HeLa cells: lysis of the phagocytic vacuole and plasmid-mediated contact hemolysis. *Infection and Immunity* **51**(2), 461-469.

Sasakawa, C., Buysse, J. M. & Watanabe, H. (1992). The large virulence plasmid of *Shigella*. *Current Topics in Microbiology and Immunology* **180**, 21-44.

Sasakawa, C., Kamata, K., Sakai, T., Makino, S., Yamada, M., Okada, N. & Yoshikawa, M. (1988). Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase palsmid in *Shigella flexneri* 2a. *Journal of Bacteriology* **170**, 2480-2484.

Sasaki, T., Uchida, T. & Kurahashi, K. (1974). Glucosylation of O-antigen in *Salmonella* carrying epsilon 15 and epsilon 34 phages. *Journal of Biological Chemistry* **249**, 761-772.

Schnaitman, C. A. & Klena, J. D. (1993). Genetics of lipoplysaccharide biosynthesis in enteric bacteria. *Microbiological Reviews* **57**(3), 655-682.

Schuch, R. & Maurelli, A. T. (1997). Virulence plasmid instability in *Shigella flexneri* 2a is induced by virulence gene expression. *Infection and Immunity* **65**(9), 3686-3692.

Schultz, C. L., Qadri, F., Hossain, S. A., Ahmed, F. & Ciznar, I. (1992). *Shigella*-specific IgA in saliva of children of bacillary dysentery. *FEMS Microbiology and Immunology* **89**, 65-72.

Sekizaki, T., Harayama, S., Brazil, G. M. & Timmis, K. N. (1987). Localization of the *stx*, a determinant essential for high level production of shiga toxin by *Shigella dysenteriae* type 1 near *pyr*F and generation of *stx* transposon mutants. *Infection and Immunity* **55**, 2208-2214.

Sereny, B. (1957). Experimental keratoconjunctivitis shigellosis. *Acta Microbiol. Acad. Sci. Hung.* **4**, 367-376.

Seyedirashti, S., Wood, C. & Akagi, J. M. (1991). Induction and partial purification of bacteriophages from *Desulfovibrio vulgaris* (Hildenborough) and *Desulfovirbrio desulfuricans* ATCC 13541. *Journal of General Microbiology* **137**, 1545-1549.

Sharples, G. J., Ingleston, S. M. & Lloyd, R. G. (1999). Holliday junction processing in bacteria: insights from the evolutionary conservation of *RuvABC*, *RecG*, and *RusA*. *Journal of Bacteriology* **181**, 5543-5550.

Shore, S. H. & Howe, M. M. (1982). Bacteriophage Mu T mutants are defective in synthesis of the major head polypeptide. *Virology* **120**, 264-268.

Simmons, D. A. R. (1971). Immunochemistry of *Shigella flexneri* O-antigens: a study of structural and genetic aspects of the biosynthesis of cell-surface antigens. *Bacteriological Reviews* **35**, 117-148.

Simmons, D. A. R. & Romanowska, E. (1987). Structure and biology of *Shigella flexneri* O antigens. *Journal of Medical Microbiology* **23**, 289-302.

Sinha, A. K., Chakraborti, M. K. & Chakraborti, S. (1992). Gut mucosal lymphocyte subpopulations in the host defence of *Shigella* infected guinea pigs. *Immunology Letters* **32**, 65-68.

Sjogren, R., Neill, R., Rachmilewitz, D., Fritz, D., Newland, J., Sharpnack, D., Colleton, C., Fondacaro, J., Gemski, P. & Boldeker, E. (1994). Role of Shiga-like toxin I in bacterial enteritis: comparison between isogenic *Escherichia coli* strains induced in rabbits. *Gastroenterology* **106**, 306-317.

Smith, M. C., Burns, R. N., Wilson, S. E. & Gregory, M. A. (1999). The complete genome sequence of the *Streptomyces* temperate phage fC31 : evolutionary relationships to other viruses. *Nucleic Acids Research* **27**(10), 2145-2155.

Stanley, E., Fitzgerald, G. F. & Sinderen, v. D. (1999). Characterisation of *Streptococcus thermophilus* CNRZ1205 and its cured and re-lysogenised derivatives. *FEMS Microbiology Letters* **176**, 503-510.

Stark, W. M., Boocock, M. R. & Sherratt, D. J. (1992). Catalysis by site-specific recombinases. *Trends in Genetics* **8**(12), 432-439.

Steinbacher, S., Miller, S., Baxa, U., Weintraub, A. & Seckler, R. (1997). Interaction of Salmonella phage P22 with its O-antigen receptor studied by x-crystallography. *Journal of Biological Chemistry* **378**, 337-343.

Studier, W. F., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology* **185**, 61-89.

Tabor, S. & Richardson, C. C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Science, U.S.A.* **82**, 1074-1078.

Tabor, S. & Richardson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proceedings of the national Academy of Science, U.S.A.* **84**, 4767-4771.

Takeda, S., Sasaki, T., Ritani, A., Howe, M. M. & Arisaka, F. (1998). Discovery of the tail tube gene of bacteriophage Mu and sequence analysis of the sheath and tube genes. *Biochimica et Biophysica Acta* **1399**, 88-92.

Takeshita, M. & Makela, P. H. (1971). Glucosylation of lipopolysaccharide in *Salmonella*: biosynthesis of O-antigen factor 12_2. III. The presence of 12_2

determinants in haptenic polysaccharide. *Journal of Biological Chemistry* **246**, 3920-3927.

Takuchi, A., Formal, S. B. & Sprinz, H. (1968). Experimental acute colitis in the Rhesus monkey following peroral infection with *Shigella flexneri*. *American Journal of Pathology* **52**, 503-520.

Taylor, A. (1971). Endopeptidase activity of phage lambda-endolysin. *Nature New Biology* **234**, 144-145.

Thompson, J. F., Moitoso de Vargas, L., Koch, C., Kahmann, R. & Landy, A. (1987). Cellular factors couple recombination with growth phase: characterization of a new component in the lambda site-specific recombination pathway. *Cell* **50**, 901-908.

Tilney, I. G. & Portnoy, D. A. (1990). Actin filament nucleation by the bacterial pathogen, *Listeria monocytogenes*. *Journal of Cell Biology III*, 2979-2988.

Timmis, K. N., Boulonois, G. J., Bitter-Suermann, D. & Cabello, F. C. (1985). Surface components of *Escherichia coli* that mediate resistance to bactericidal activities of serum and phagocytosis. *Current Topics in Microbiology and Immunology* **118**, 197-218.

Timokov, V. D., Petrovskaya, V. G. & Bondarenko, V. M. (1970). Studies of the genetic control of *Shigella* subgroup B-type specific antigens. *Annals of Institute Pasteur, Paris* **118**, 3-9.

Tobe, T., Sasakawa, C., Okada, N., Honma, Y. & Yoshikawa, M. (1992). *vacB*, a novel chromosomal gene required for expression of virulence genes on the large plasmid of Shigella flexneri. *Journal of Bacteriology* **174**, 6359-6367.

Tobias, P. S., Soldau, K. & Ulevitch, R. J. (1989). Identification of a lipid A binding site in the acute phase reactant lipopolysaccharide binding protein. *Journal of Biological Chemistry* **264**, 10867-10871.

Tomasi, T. B., Tan, E. M. & Solomon, A. (1965). Characteristics of an immune system common to certain external secretions. *Journal of Experimental medicine* **121**, 101-124.

Tomka, M. A. & Catalano, C. E. (1993). Physical and kinetic characterization of the DNA packaging enzyme from bacteriophage lambda. *Journal of Biological Chemistry* **268**, 3056-3065.

Towbin, H., Staehelin, T. & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Science, U.S.A.* **76**, 4350-4354.

Tremblay, D. M. & Moineau, S. (1999). Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. *Virology* **255**, 63-76.

Ulmann, A., Jacob, F. & J., M. (1967). Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the B-galactosidase structural gene of *Escherichia coli*. *Journal of Molecular Biology* **24**, 339.

Valpuesta, J. M. & Carrascosa, J. L. (1994). The structure and function of viral connectors. *Quarterly Reviews in Biophysics* **27**, 107-155.

Vander Byl, C. & Kropinski, A. M. (2000). Sequence of the genome of *Salmonella* bacteriophage P22. *Journal of Bacteriology* **182**(22), 6472-6481.

Vasselon, T., Mounier, J., Hellio, R. & Sansonetti, P. J. (1992). Movement along actin filaments of the perijunctional area and de novo polymerisation of cellular actin are required for *Shigella flexneri* colonisation of epithelial Caco-2 cell monolayers. *Infection and Immunity* **60**, 1031-1040.

Vasselon, T., Mounier, J., Prevost, M. C., Hellio, R. & Sansonetti, P. J. (1991). Stress fiber- based movement of *Shigella flexneri* within cells. *Infection and Immunity* **59**, 1723-1732.

Verma, N. K., Brandt, J. M., Verma, D. J. & Lindberg, A. A. (1991). Molecular characterisation of the O-acetyl transferase gene of converting bacteriophage SF6 that adds group antigen 6 to *Shigella flexneri*. *Molecular Microbiology* **5**(1), 71-75.

Verma, N. K., Verma, D. K., Huan, P. T. & Lindberg, A. A. (1993). Cloning and sequencing of the glucosyl transferase-encoding gene from converting bacteriophage X (SfX) of *Shigella flexneri*. *Gene* **129**, 99-101.

Vogelstein, B. & Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. *Proceedings of the National Academy of Science, U.S.A.* **76**, 615-619.

Waldor, M. K. (1998). Bacteriophage biology and bacterial virulence. *Trends in Microbiology* **6**(8), 295-297.

Waldor, M. K. & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910-1914.

Wassef, J. S., Keren, D. F. & Mailloux, J. L. (1989). Role of M cells in initial antigen uptake and in ulcer formation in the rabbit intestinal loop model of Shigellosis. *Infection and Immunity* **57**, 858-863.

Watanabe, H. & Nakamura, A. (1985). Large plasmid associated with virulence in *Shigella* species have a common function necessary for epithelial penetration. *Infection and Immunity* **48**, 260-262.

Watanabe, H., Nakamura, A. & Timmis, K. N. (1984). Small virulence plasmid of *Shigella dysenteriae* 1 strain W30864 encodes a 41000-dalton protein involved in formation of specific lipopolysaccharide side chains of serotype 1 isolates. *Infection and Immunity* **46**, 55-63.

Wawrzynow, A., Banecki, B. & Zylicz, M. (1996). The Clp ATPases define a novel class of molecular chaperones. *Molecular Microbiology* **21**, 895-899.

Weisberg, R. & Landy, A. (1983). Site-specific recombination in phage lambda. In *Lambda II* (Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A., eds.). Cold Spring Harbor Laboratory, New York.

Weisberg, R. A. & Gottesman, M. E. (1971). The stability of Int and Xis functions. In *The bacteriophage lambda* (Hershey, A. D., ed.), pp. 489-500. Cold Spring Harbor Laboratory, New York.

Whitfield, C. (1995). Biosynthesis of lipopolysaccharide O antigens. *Trends in Microbiology* **3**(5), 178-185.

Whitman, W. B., Coleman, D. C. & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences, USA* **95**, 6578-6583.

WHO. (1997). The world health report 1996 - fighting disease, fostering development. *World Health Forum* **18**, 1-8.

WHO. (1987). Development of vaccines against shigellosis: memorandum from a WHO meeting. *Bulletin if the World Health Organization* **65**, 667-676.

WHO. (1991). Research priorities for diarrhoeal disease vaccines: memorandum from a WHO meeting. *Bulletin of the WHO* **69**, 667-676.

Wright, A. (1971). Mechanism of conversion of the *Salmonella* O antigen by bacteriophage epsilon 34. *Journal of Bacteriology* **105**, 927-936.

Yamada, M., Sasakawa, C., Okada, N., Makino, S. & Yoshikawa, M. (1989). Molecular cloning and characterization of chromosomal virulence region *kcpA* of *Shigella flexneri*. *Molecular Microbiology* **3**, 207-213.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Yao, Z. & Valvano, M. A. (1994). Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (*rfb*) of *Escherichia coli* K-12 W3110: Identification of genes that confer group 6 specificity to *Shigella flexneri* serotypes Y and 4a. *Journal of Bacteriology* **176**(13), 4133-4143.

Zwillich, S. H., Dubt, A. D. & Lipsky, P. E. (1989). T-lymphocyte clones responsive to Shigella flexneri. *Journal of Clinical Microbiology* **27**, 417-421.

Zychlinsky, A., Prevost, M. C. & Sansonetti, P. J. (1992). *Shigella flexneri* induces apoptosis in infected macrophages. *Nature* **358**, 167-169.

Zychlinsky, A., Thirumalai, K., Arondel, J., Cantey, J. R., Aliprantis, A. O. & Sansonetti, P. J. (1996). *In vivo* apoptosis in *Shigella flexneri* infection. *Infection and Immunity* **64**, 5357-5365.