

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

# **Unlocking the M13 (f1 and fd) virion**

**Investigation into the role of the pIII C-domain  
of F specific filamentous bacteriophage in infection**

**A thesis presented in partial fulfilment of the requirements for the degree of**

**Doctor of Philosophy**

**in**

**Biochemistry**

**at Massey University, Palmerston North,**

**New Zealand.**

**Nicholas James Bennett**

**2009**

## Abstract

Ff filamentous bacteriophage infect male ( $F^+$ ) strains of *Escherichia coli* and are assembled at the cell membranes, by a secretion-like, non-lethal process. The pIII protein, located at one end of the virion-filament, is required at both the beginning and the end of the phage life cycle. During infection, the N-terminal domains of pIII, N2 and N1, bind to the primary and secondary host receptors, F pilus and TolA protein, respectively. At the end of the life cycle, the pIII C-domain mediates the termination and release of virions. Thus, both entry and release involve structural transitions of the virus coupled to membrane transactions of the virion proteins. "Unlocking" of the highly stable virion presumably results in membrane integration during entry, whereas a reverse event, "locking" of the virion, occurs upon detachment from the membrane at termination step of assembly/secretion. Recently, it was shown that the pIII C-domain plays an active role at the step of entry. This finding implicates the C-domain of pIII in "unlocking" of the virion, presumably resulting in the exposure of the membrane anchor at the very C-terminus of pIII (Bennett & Rakonjac, 2006).

To further this work, this thesis has mapped the portion of the pIII C-domain required for infection, by constructing a set of nested deletions of the C-domain fused to the receptor binding domains N1 and N2, and then determined the infectivity of phage carrying the mutant proteins. This mapped the portion of the C-domain required for phage infection is different to that required for termination of assembly. The different requirement for entry and release suggests that the two processes are carried out by distinct mechanisms and/or depend on different sets of accessory proteins.

In addition, a system was designed for the efficient production and purification of very short virions, the length of which is 1/20 that of the wild-type fl. These short virions, called microphage, are the first step towards the structural analyses of the phage terminus cap structures, of which one contains pIII in the "locked" conformation.

## **Acknowledgements**

Firstly, I would like to thank my supervisor Dr Jasna Rakonjac for helping me through this thesis. I would also like to thank Marjorie Russel and Peter Model from Rockefeller University for supplying materials and advice. I would also like to thank all the people in the Helipad lab who have helped me through my research. I would also like to thank my friends and family for support throughout the last few years.

Most of all I would love to thank my wife, Leonie, for all the love and support she has given me throughout my Ph.D.

# Contents

<b>Chapter 1</b> .....	Page
------------------------	------

## **Literature Review** ..... **1**

1.1 Introduction to filamentous bacteriophage.....	1
1.2 The filamentous phage genome.....	5
1.3 The Ff phage structure.....	11
1.3.1 Structure and role of pVIII, the major coat protein.....	11
1.3.2 The structure and role of pVII and pIX.....	16
1.3.3 The structure and role of pIII and pVI.....	16
1.4 The Ff phage lifecycle.....	22
1.4.1 Infection of host cell.....	22
1.4.2 Resistance to filamentous phage infection caused by pIII.....	26
1.4.3 Replication of phage within the host cell.....	26
1.4.4 The intergenic region and replication.....	31
1.4.5 The microphage-producing origin replication.....	32
1.4.6 Interference resistance and phage replication.....	35
1.4.7 The Ff phage export apparatus and phage assembly/secretion.....	39
1.5 Other viral mechanisms of entry.....	43
1.6 Aims.....	46

## **Chapter 2**.....

## **Materials and Methods**..... **47**

2.1 Bacterial strains, culture conditions and chemicals.....	47
2.2 Molecular biology methods.....	47
2.3 Construction of plasmids.....	48
2.4 Phage protocols.....	67
2.4.1 Phage stocks.....	67
2.4.2 Growth experiments for producing phage or phagemid particles (PPs) containing combinations of pIII mutants.....	68
2.4.3 Producer cell infection test.....	72
2.4.4 Concentration of virions by PEG precipitation.....	74
2.4.5 Purification of concentrated phage using Sarkosyl and Triton X-100.....	74
2.4.6 Titration of infectious phage and phagemid particles.....	75
2.4.7 Agarose gel electrophoresis of the phage and quantification.....	76
2.4.8 Protein electrophoresis and western blots.....	79
2.4.9 Growth, concentration and purification of microphage.....	79

2.5 Bootstrap Analysis ..... 83

**Chapter 3.....Page**

**Mapping of the pIII C-domain for the minimal fragment  
required for phage infection..... 86**

3.1 Introduction..... 86

3.2 Experimental system for the production of ..... 86  
composite virions ..... 86

3.3 The complete C-domain complements the assembly deficiency of the NdC83  
mutation and stabilizes the NdC93-121 virions..... 88

3.4 The NdC mutants are incorporated into the virions ..... 93

3.5 Analysis of the NdC mutant virion infectivity..... 95

3.6 A specific region of the C-domain is required for N1N2-mediated fl infection 101

3.7 Evidence for correct folding of the N1N2 domains of pIII ..... 102

3.8 Comparison of the infectivity of WTpIII/Cd positive control to complemented  
fl d3 ..... 103

3.9 Effect of N1N2 stoichiometry on opening of phage cap complex in NdC/Cd  
carrying phage ..... 104

3.10 Discussion..... 105

**Chapter 4.....**

**Search for the pIII C-terminal domain residues involved in  
infection using alanine scanning mutagenesis..... 107**

4.1 Introduction..... 107

4.2 Construction of the mutants..... 108

4.3 Phagemid particle expression system.....	108
4.4 Alanine scan mutants terminate phage assembly correctly .....	110
4.5 Analysis of infectivity of mutant phagemid particles .....	113
4.6 Summary of results .....	118
4.7 Secondary structure prediction .....	118
4.8 The ICS region is not involved in the intra-cap complex signalling.....	125
<b>Chapter 5.....</b>	<b>Page</b>
<b>Construction of an efficient microphage-producing system</b>	<b>129</b>
5.1 Introduction.....	129
5.2 Reconstruction of the microphage origin and cloning into the high copy number plasmid pCR4Blunt-TOPO. ....	130
5.3 A helper phage for efficient microphage production .....	131
5.4 Concentration and initial purification of microphage .....	134
5.5 Purification of microphage using native phage gel electrophoresis.....	137
5.6 Analysis of purified microphage.....	139
5.7 Visualisation of microphage by electron microscopy .....	143
5.8 Future uses of microphage production system for structural studies.....	149
5.9 Discussion of microphage production system .....	150
<b>Chapter 6.....</b>	<b>.....</b>
<b>Discussion .....</b>	<b>152</b>
6.1 Introduction.....	152
6.2 A new model for the involvement of the C-domain of pIII in the process of phage infection.....	152

6.3 Future experiments to evaluate the role of the pIII C-domain in phage infection	156
6.4 Construction of a microphage variant of filamentous phage for structural studies	161
6.5 Effect of pVIII on phage virion length.....	161
6.6 Future modification of the microphage production system to enhance structural studies.....	167
6.7 Modification of the microphage producing system to facilitate purification.....	167
6.8 Comparison of fl entry to other bacteriophage and viruses.....	168
<b>Chapter 7</b> .....	
<b>Conclusion</b> .....	<b>170</b>
<b>Appendix</b> .....	<b>Page</b>
Appendix 1 - General Statistic of Ff Filamentous Phage.....	173
Appendix 2 - Position of all fl <i>gIII</i> primers used in this thesis.....	174
Appendix 3 - Protein Sequence of WTpIII and NdC mutants.....	183
Appendix 4 - I-Tasser Results .....	190
Appendix 5 - Reference CD .....	200
<b>References</b> .....	<b>201</b>

<b>List of Figures .....</b>	<b>Page</b>
Figure 1. The map of the fl genome.....	7
Figure 2. Anatomy of Ff filamentous phage virion. ....	9
Figure 3. Atomic models of pVIII monomers as structured within phage virions.....	13
Figure 4. pVIII forms a shingle-like helical array around the ssDNA genome .....	14
Figure 5. Distal end of the Ff filamentous phage. ....	18
Figure 6. Three-dimensional structure of pIII N1N2 domains.....	21
Figure 7. Model of phage infection. ....	24
Figure 8. Overview of life cycle of Ff filamentous phage. ....	27
Figure 9. Secondary structure of the fl phage intergenic sequence .....	29
Figure 10. Intergenic region sequence and positions of various features.....	30
Figure 11. Secondary structure of an engineered microphage origin of replication .....	34
Figure 12. Phage secretion, termination and release from the host cell.....	37
Figure 13. The production of cell-associated phage filaments.....	41
Figure 14. Diagram of the mutant pIII proteins used in NdC mutant experiments and positions of primers .....	60
Figure 15. Relative positions of the Alanine scan primers. ....	65
Figure 16. Method for producing mutant phagemid particles.....	70
Figure 17. Flow diagram of the production of the virions carrying combination of pIII constructs.....	73
Figure 18. A diagram describing the concept of genome equivalent. ....	78
Figure 19. Optimised microphage purification protocol.....	81
Figure 20. Flow diagram of the bootstrap analysis used to predict the standard deviation of infectivity. ....	84
Figure 21. Native phage gel electrophoresis. ....	91
Figure 22. Western blots of virions .....	94
Figure 23. Infectivity of phagemid particles containing the NdC mutants.....	98
Figure 24. Infectivity of phage particles containing the NdC mutants.....	100
Figure 25. Native phage agarose electrophoresis of virions containing pIII point mutants. ....	111
Figure 26. Infectivity of virions containing pIII point mutants.....	117
Figure 27. Bioinformatics analysis of the C domain of pIII. ....	121

Figure 28. Structural model of pIII C-domain .....	123
Figure 29. Scenarios of pIII inter-subunit cooperation during entry. ....	126
Figure 30. Position of mutated codon in <i>gVIII<sup>am</sup></i> mutant used in this thesis. ....	133
Figure 31. Southern blot detection of microphage ssDNA from "high PEG" fractions. .....	135
Figure 32. Detection of the microphage by native virion agarose gel electrophoresis. ....	138
Figure 33. Purity and concentration of microphage .....	141
Figure 34. Electron micrographs of gel-purified microphage samples .....	145
Figure 35. High-resolution TEM and 3D atomic projection of phage filament.....	147
Figure 36. Proposed model one for fl entry.....	155
Figure 37. Proposed genetic screen for identification of intergenic suppressor mutations in TolQRA.....	159
Figure 38. Template-dependent spatial relations between pVIII and pIII concentration gradients in the host cells. ....	164

<b>List of Tables</b> .....	Page
Table 1: Filamentous Phage Lifestyles .....	4
Table 2. PCR primers.....	49
Table 3. Plasmid .....	54
Table 4. Phage strains .....	58
Table 5. Bacterial strains list .....	59
Table 6. Infectivity of the NdC deletion mutant phagemid particles .....	97
Table 7. Infectivity phage particles containing the NdC mutants .....	99
Table 8. Infectivity and Relative Infectivity of Alanine Point Mutations .....	115
Table 9: Infective titre of phagemid particles containing WTPIII and nested deletions of the C-domain .....	128

## List of Abbreviations

aa	-Amino acid
Amp	-Ampicillin
AP	-Alkaline phosphatase
BCIP	-5-bromo-4-chloro-3-indoxyl phosphate
bp	-Base pair
Cd	-pIII C-domain
Cm	-Chloramphenicol
Cm <sup>R</sup>	-Chloramphenicol Resistant
Cryo-EM	-Cryogenic electron microscopy
DNA	-Deoxyribose nucleic acid
DNase	-Deoxyribonuclease
dsDNA	-Double stranded deoxyribose nucleic acid
<i>E.coli</i>	- <i>Escherichia coli</i>
F	-F conjugative plasmid
F <sup>+</sup>	- <i>E.coli</i> carrying F plasmid, also termed "male"
F <sup>-</sup>	- <i>E.coli</i> not carrying F plasmid, also termed "female"
Ff	-F <sup>+</sup> specific filamentous bacteriophage of <i>E.coli</i> , including f1,fd and M13
HA	-Hemagglutinin
HBV	-Hepatitis B virus
HIV	-Human immunodeficiency virus
ICS	-Infection-competence sequence
IF	-Infective form
IPTG	-Intergenic region
IR	-Interference resistant
Kan	-Kanamycin
kb	-Kilo base
Km <sup>R</sup>	-Kanamycin Resistant
Km <sup>S</sup>	-Kanamycin Sensitive
m.o.i	-Multiplicity of infection
Nd	-pIII N-terminal domains N1 and N2

NMR	-Nuclear magnetic resonance
nt	-Nucleotide
OD	-Optical density
<i>ori</i>	-Origin of replication
PCR	-Polymerase chain reaction
PEG	-Polyethylene glycol
PONDR	-Predictors of Natural Disordered Regions
RF	-Replicative form
RNase	-Ribonuclease
Sarkosyl	-n-lauroylsarcosine, sodium salt
SDS	-Sodium dodecyl sulfate
ssDNA	-Single stranded deoxyribose nucleic acid
TBS	-Tris buffered saline
TEM	-Transmission Electron Microscopy
TFF	-Tangential Flow Filtration
WT	-Wild-type

# Chapter 1

## Literature Review

### 1.1 Introduction to filamentous bacteriophage

Filamentous phage (Inovirus) are long, thin bacteriophage that can be found infecting a wide range of different gram negative bacterial species including *Escherichia*, *Salmonella*, *Pseudomonas* and *Vibrio* (Calendar, 2005; Model & Russel, 1988). Almost all of the known species that are infected by filamentous phage are Gram-negative organisms, however there are two published reports of filamentous phage that infect Gram-positive organisms (Chopin et al., 2002; Kim & Blaschek, 1991). In total around 60 different filamentous phage strains have been described to date (ICTVdB, 2008).

All filamentous bacteriophage have certain features that are characteristic this group. They are usually between 1-2  $\mu\text{m}$  in length, are 6-7 nm in diameter and all contain a single stranded circular DNA genome (Model & Russel, 1988; Webster, 1996). In filamentous phage, the virion length is determined by the genome size. Thus the larger the genome, the longer the resulting phage becomes. Another common feature is that filamentous bacteriophage are released from their host cells without causing lysis by a combined secretion-assembly process at the cell envelope. The phage assembly/secretion machinery shares some features common to the membrane machineries used by many pathogenic bacteria for the secretion of virulence factors (type II and III secretion systems) or for the assembly of type IV pili (Coburn, Sekirov & Finlay, 2007; Collins et al., 2005; Johnson et al., 2006).

The virion coat is composed of five proteins. The dominant protein in the structure is the major coat protein (pVIII). It forms the tube-like structure (a shingle-like array of  $\alpha$  helices) that encloses the ssDNA genome. There are two sets of two different minor coat proteins which are located at either end of the bacteriophage (pVII/pIX and pIII/pVI).

The best-studied filamentous bacteriophage are Ff group that infect male (F<sup>+</sup>) strains of *Escherichia coli* (Hoffman-Berling, Durwald & Beulke, 1963; Hofschneider, 1963; Loeb, 1960). Members of the Ff group include filamentous phage f1, fd and M13, which are all almost genetically identical; their DNA sequences differ by as little as 2% (Webster, 1996). Filamentous bacteriophage found in this group infect *Escherichia coli* by the recognition of and attachment to the tip of the F conjugative pilus (F). As with all filamentous phage, infection with Ff bacteriophage does not lyse cells. However, in the case of Ff phage, the generation time of the bacterial host is extended by about 50%.

Because they produce single stranded DNA, Ff filamentous bacteriophage have been previously used as vectors for DNA sequencing (Vieira & Messing, 1982). More recently, Ff has been used for phage display technology (Barbas et al., 2004) and nanotechnology. In parallel, Ff phage have been used as a model system for some of the basic discoveries in molecular cell biology, such as rolling circle replication (Dotto, Horiuchi & Zinder, 1984), *in vitro* translation (Chan, Model & Zinder, 1975; Konings, 1973; Model & Zinder, 1974), integration of proteins into the membrane (Davis & Model, 1985), and secretion from Gram-negative bacteria (Kazmierczak et al., 1994; Linderoth, Simon & Russel, 1997).

Several types of lifecycles have been identified in filamentous phage. The major difference in the various filamentous phage lifecycles is whether or not the phage genome establishes a stable lysogeny (Waldor & Mekalanos, 1996).

The lifecycle of the Ff is probably best studied of all filamentous phage. However, the lifecycle observed in this group is not typical of the majority of currently identified filamentous phage (Table 1) (Model & Russel, 1988). Ff phage do not integrate into the host cell genome; they persist self-replicating plasmids within their hosts. Ff virions are also produced by an infected cell at a very high titre. This imposes a significant metabolic stress on host cells, increasing the cell cycle length and hence slowing cell division. This results in the formation of a region of lower cell density in the lawn of cells in comparison to regions containing non-infected cells. Hence, infected cells form visible plaques. Many filamentous phage are much less efficient in virion assembly, as a result they do not interfere with the host generation time and do not form visible plaques. The other well studied *E.coli* filamentous phage Ike and If1 (N pilus specific

filamentous phage) also do not integrate with the host genome (Khatoon, Iyer & Iyer, 1972).

Recently, due to the advent of high-throughput sequencing, many filamentous phage prophage elements have been identified within bacterial genomes (Gonzalez et al., 2002; Waldor & Mekalanos, 1996). In addition, a large number of these lysogenic filamentous phage have been linked to pathogenic strains of bacteria associated with human disease (Gonzalez et al., 2002; Mooij et al., 2007; Nasu et al., 2000; Webb, Lau & Kjelleberg, 2004). Interestingly, of the known lysogenic filamentous phage many integrate at very specific sites within the genome of host (McLeod et al., 2005).

The best-studied lysogenic filamentous phage is CTX $\phi$ . The genome of CTX $\phi$  was first identified as the pathogenicity island in *Vibrio cholerae* that coded for the CTX $\alpha\beta$  toxin, which is the major virulence factor of *V.cholerae*. Using a marker replacement to label the island with Km<sup>R</sup> it was shown that this pathogenicity island encoded a lysogenic filamentous phage (Waldor & Mekalanos, 1996). The phage receptor on the host cell was shown to be the TCP pilus. The CTX $\phi$  genome lacks the outer membrane channel and was shown to use instead the type II secretion system channel EpsD (Davis et al., 2000; Davis & Waldor, 2003). More recently, the mechanism by which CTX $\phi$  integrates into the host genome has been shown (Huber & Waldor, 2002). Interestingly, although the individual genes of CTX $\phi$  have little homology to the genes of phage within the Ff group, the genome arrangement of CTX $\phi$  is highly conserved to the genome arrangement of Ff phage. Also interestingly it seems that like both phage within the Ff group and Ike, CTX $\phi$  uses the *V.cholerae* homologue of the TolQRA complex as the phage secondary receptor (Heilpern & Waldor, 2000).

Examples and lifecycle details of some of the known filamentous phage strains are shown on Table 1.

**Table 1: Filamentous Phage Lifestyles**

	Name <sup>a</sup>	Organism	Integrated	Plaque forming	Titre	Virulence associated	Virulence factor	Reference
<b>Gram Negative</b>	f1	<i>E.coli</i>	No	Yes	High	No	N/A	(Loeb, 1960)
	Fd	<i>E.coli</i>	No	Yes	High	No	N/A	(Marvin & Hoffmann-Berling, 1963)
	M13	<i>E.coli</i>	No	Yes	High	No	N/A	(Hofschneider, 1963)
	Ike	<i>E.coli</i>	No	Yes	High	No	N/A	(Khatoon et al., 1972)
	If1	<i>E.coli</i>	No	Yes	High	No	N/A	(Meynell & Lawn, 1968)
	If2	<i>E.coli</i>	No	Yes	High	No	N/A	(Meynell & Lawn, 1968)
	CTX $\phi$	<i>V.cholerae</i>	Yes	No	Low	Yes	ctx $\alpha\beta$	(Waldor & Mekalanos, 1996)
	pf1	<i>P. aeruginosa</i>	No	Yes	High	No	N/A	(Takeya & Amako, 1966)
	pf4	<i>P. aeruginosa</i>	Yes	Yes	Medium	Yes	None found	(Webb et al., 2004)
	pf5	<i>P. aeruginosa</i>	Yes	Not indicated	Low	Yes	None found	(Mooij et al., 2007)
	CUS1	<i>E.coli</i> O18:K1:H7	Yes	No	Low	Yes	Unkown	(Gonzalez et al., 2002)
	CUS2	<i>Y.pestis</i> biovar <i>orientalis</i>	Yes	No	Low	Yes	Unkown	(Gonzalez et al., 2002)
	F237	<i>V.parahaemolytias</i> O3:K3	No	Yes	Not Indicated	Yes	Unkown	(Nasu et al., 2000)
<b>Gram Positive</b>	CAK1	<i>C. acetobutylicum</i>	No	Not Indicated	Not Indicated	No	No	(Kim & Blaschek, 1991)
	B5	<i>P. freudenreichii</i>	No	Yes	Not Indicated	No	No	(Chopin et al., 2002)

a) This is not an extensive list of filamentous phage isolates. It is only a list of examples of the various filamentous phage lifestyles.

## 1.2 The filamentous phage genome

The genome of  $\phi$ 1 was the first genome to be completely sequenced in New Zealand and second ever genome sequenced (after  $\phi$ X174) (Beck & Zink, 1981; Hill & Petersen, 1982; van Wezenbeek, Hulsebos & Schoenmakers, 1980). The  $\phi$ 1 genome contains 11 genes and an intergenic sequence, which accommodates the origin of replication and the packaging or morphogenetic signal (Figure 1).

All genes are transcribed in the same direction and are organised in four groups, according to their function (Figure 1). The first group are genes that code for the proteins involved in the replication of the genome: *gII*, *gX* and *gV*. The second group are genes that code for the proteins of the virion: *gVII*, *gIX*, and *gVIII*. The third group are genes arranged as an operon in which 2 proximal cistrons (*gIII*, *gVI*) form the distal cap of the virion, whereas the distal cistron (*gI/gXI*) encodes for the inner membrane component of the assembly machinery. The fourth group is represented by a single gene (*gIV*) that encodes for the outer membrane component of the assembly machinery. It is separately transcribed and is located just downstream of *gI*. Genes *gX* and *gXI* are translated from internal AUG start codons within *gII* and *gI*, respectively, giving rise to N-terminally truncated derivatives of pX and pXI (Model & Russel, 1988; Rapoza & Webster, 1995).

The expression of phage genes is tightly controlled and varies according to gene. Large amounts of the major coat protein pVIII and single stranded DNA binding protein pV are made, compared with proteins pI, pIII, pIV, pVI, pVII, pIX, pX and pXI which are made in small amounts (Blumer, Ivey & Steege, 1987). Transcription is initiated from multiple promoters, but only terminated at a few sites (Chan et al., 1975; Okamoto et al., 1977). This results in numerous overlapping transcripts. Transcription only occurs from the supercoiled dsDNA form of the genome (Jacob & Hofschneider, 1969), where the negative (-) strand acts as the template strand (Jacob & Hofschneider, 1969).

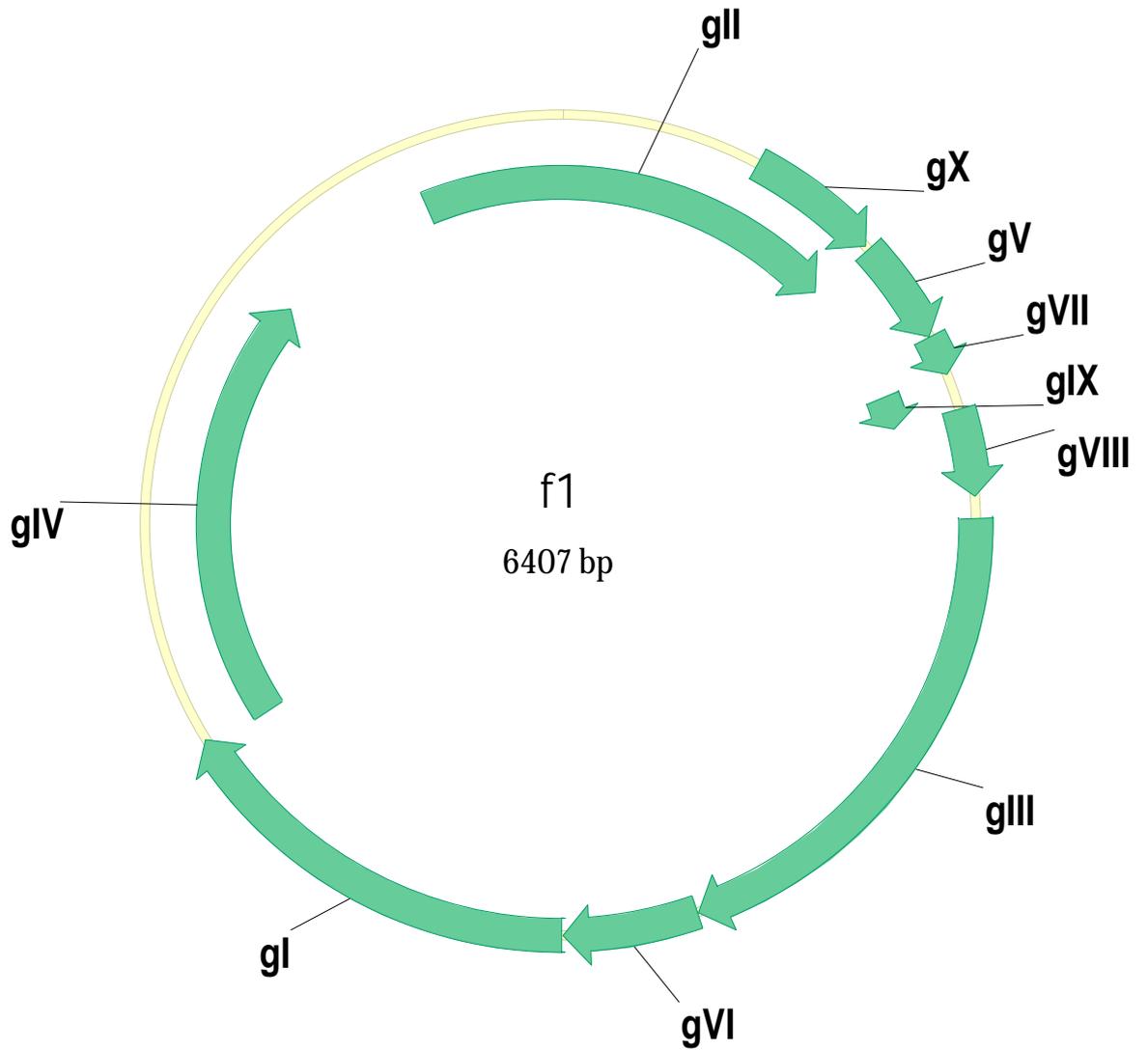
Early genetic studies determined that many mutations that either drastically decrease phage gene expression or inactivated phage proteins are lethal to the host cell. This is especially the case when mutations involve the proteins required for the formation of

the assembly/export complex (pI/pXI:pIV), or the initiation of phage assembly (pVII and pIX) (Pratt, Tzagoloff & Beaudoin, 1969; Pratt, Tzagoloff & Erdahl, 1966).

Therefore, Ff gene expression without phage assembly and export is lethal.

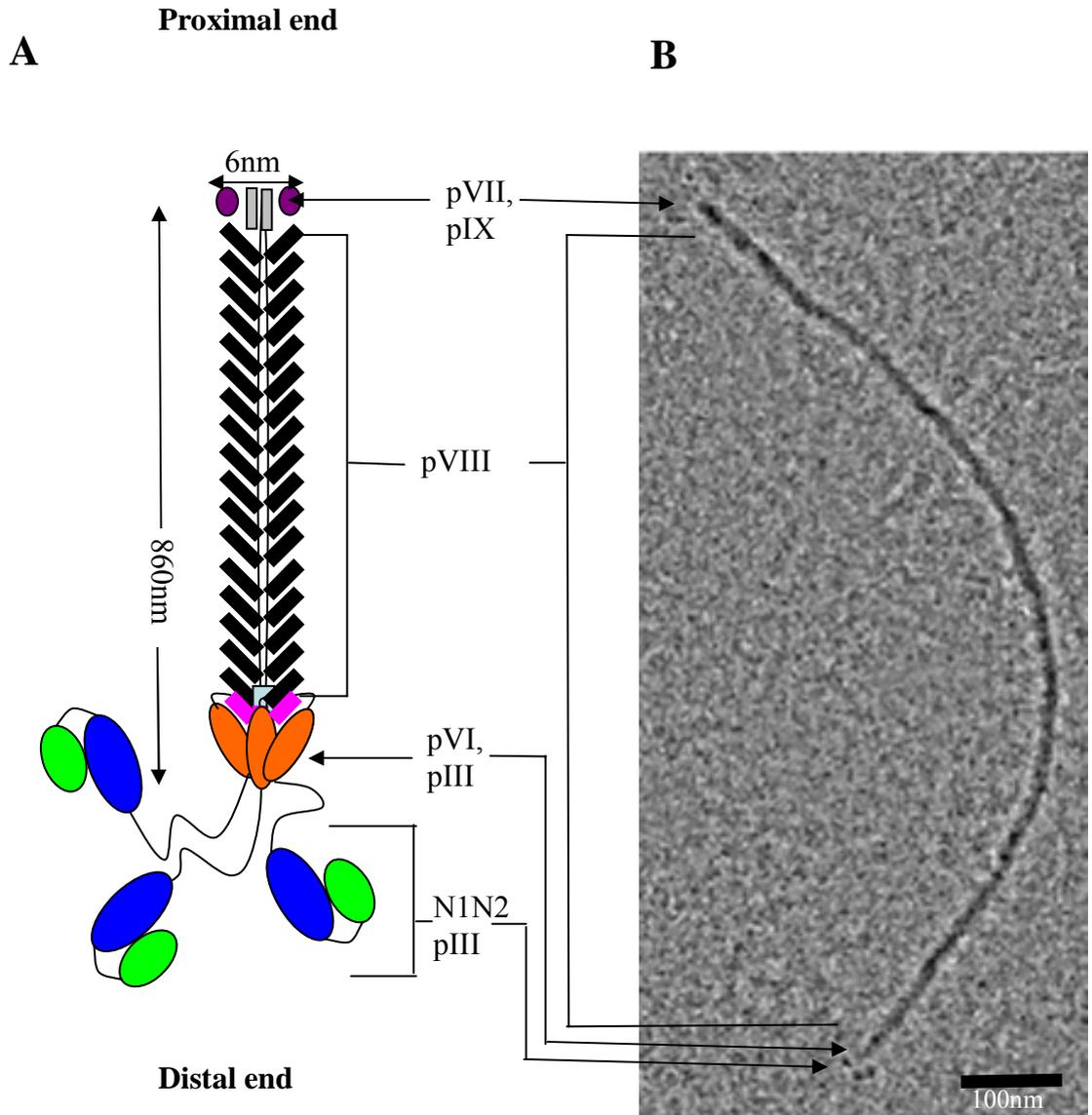
Overexpression of some genes (especially pVIII) is also lethal to the host cell.

Therefore, any attempts to modify Ff phage gene expression have to be accompanied by the compensatory changes (or mutations) to prevent host death.

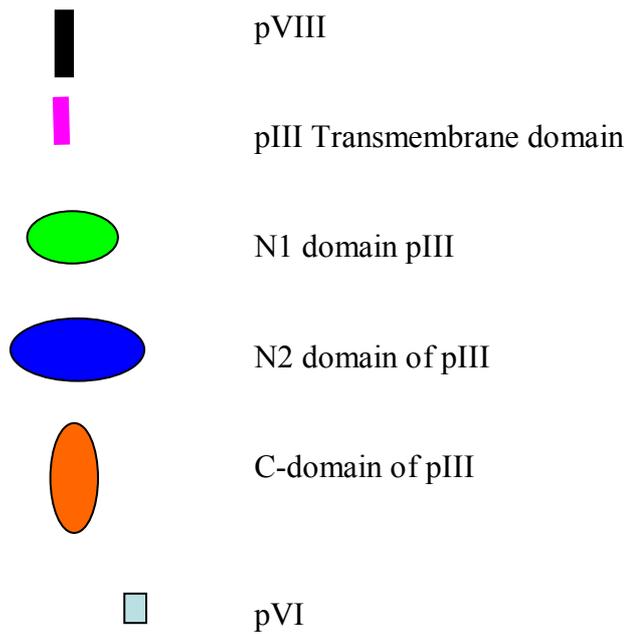


Gene	Protein Function	No of Amino acids	Protein MW
II	DNA replication	410	46137
X	DNA replication	111	12672
V	ssDNA binding protein	87	9682
VIII	Major coat protein	50	5235
III	Minor coat protein	406	42522
VI	Minor coat protein	112	12342
VII	Minor coat protein	33	3599
IX	Minor coat protein	32	3650
I	Assembly cytoplasmic membrane	348	39502
XI	Assembly cytoplasmic membrane	108	12424
IV	Assembly outer membrane	405	43476

**Figure 1. The map of the f1 genome.** The position of genes is indicated and a table of the gene products is included below the map. Molecular weights of mature protein products are given; these do not include amino-terminal signal sequence. The diagram and table was adapted from (Webster, 1996). A note on Ff genes and protein symbols: Genes are symbolized by a lowercase "g" as a prefix to the appropriate roman number (e.g. *gIII*). Phage protein products are symbolized by a lowercase "p" as a prefix (e.g. pIII). However, some articles may also refer to the genes and proteins in Ff phage using the letters A-K, or Arabic numerals.



**Key**



**Figure 2. Anatomy of Ff filamentous phage virion.** **A** Schematic representation of the virion proteins; **B** Electron Micrograph of the Ff phage. Minor coat proteins pVII and pIX form capping structure at the proximal end of the virion. Major coat protein pVIII forms the main cylinder of the virion. Minor coat proteins pIII and pVI form the capping structure at the distal end of the virion. Phage is a filament 860 nm in length and 6 nm in diameter. TEM (B) is from (Gray, Brown & Marvin, 1979)

### **1.3 The Ff phage structure**

Ff virions all have the same structure (Figure 2). The virion coat consists of five proteins (pVIII, pIII, pVI, pVII and pIX). pVIII is the major coat protein and forms the majority of the virion coat. The other proteins can be separated into two different groups; pIII and pVI are at the distal end of the virion where the assembly is terminated and pVII and pIX at the proximal end where assembly is initiated.

#### **1.3.1 Structure and role of pVIII, the major coat protein**

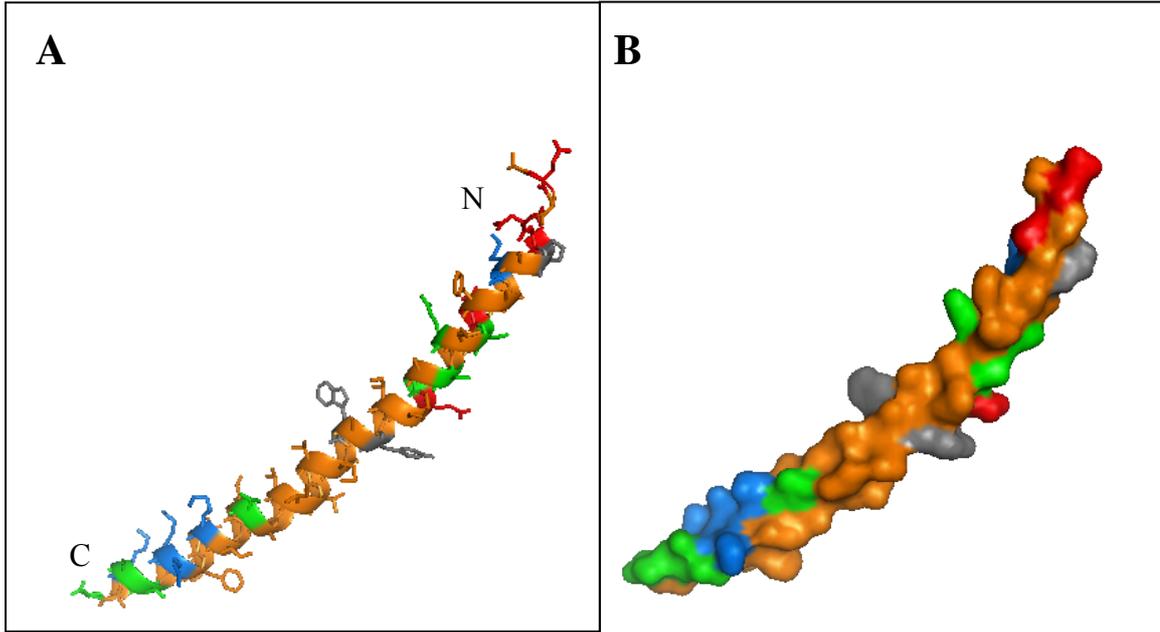
pVIII forms the majority of the virion coat. Its structure in solution and within the virion has been solved using X-ray fibre diffraction, Cryo EM and NMR to high resolution (Liu & Day, 1994; Marvin, 1998; Marvin et al., 1994; Overman & Thomas, 1995; Williams et al., 1995) (Figure 3).

pVIII is a small protein of only 50 amino acids. It consists of two  $\alpha$ - Helices (N-terminal helix and a C-terminal helix) separated by a short flexible hinge (Figure 3). pVIII also has an N-terminal signal sequence for targeting to the SecYEG/YidC translocator (Samuelson et al., 2000). This signal sequence is cleaved off upon insertion into the bacterial membrane (Russel & Model, 1982).

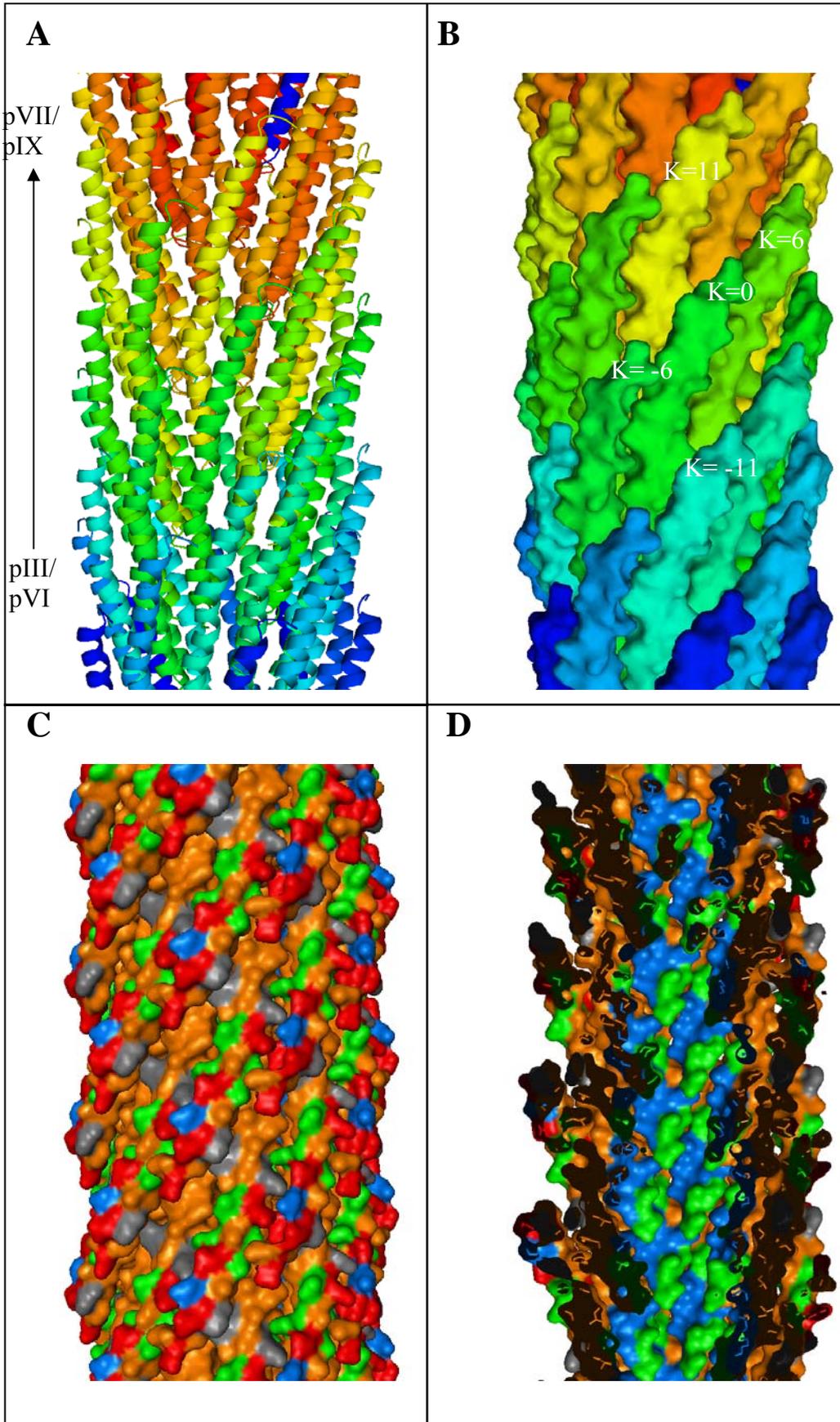
In the virion, the amphiphatic N-terminal  $\alpha$  helix is on the exterior of the structure. The hydrophilic side of the N-terminal helix faces the exterior of the virion, whereas the hydrophobic face of this helix covers the hydrophobic C-terminal helices of two different pVIII subunits assembled into the virion 6 and 11 steps before the observed pVIII subunit (Figure 4B). The C-terminal helix of a pVIII is covered by the pVIII subunit assembled into the virion 6 and 11 positions after the observed subunit (Figure 4B). This arrangement of the pVIII subunit forms a shingle-like helical array that encapsulates the ssDNA core. Each subunit of pVIII is tilted by 20° relative to the main axis of the virion and there are 5 subunits of pVIII per turn of the pVIII helical array (Figure 4)(Nambudripad et al., 1991). Prior to assembly into the virion, the hydrophobic

C-terminal  $\alpha$  helix serves as a membrane anchor (Papavoine et al., 1998; Russel & Model, 1982).

The C-terminal tail (downstream of the C-terminal helix) is positively charged. These positively charged residues are exposed to the cytosol prior to assembly, whereas in the virion they interact with the phage ssDNA (Figure 3, 4). It has been modelled that the ssDNA genome of filamentous phage packs into the virion with the phosphates at the centre of the structure and the DNA bases facing outwards, unpaired (Day et al., 1988; Day, Wiseman & Marzec, 1979). The C-terminus of the pVIII subunits complements the electrostatic charge of the phosphates; this unusual ssDNA helix depends on the number of positive charges available from the pVIII C-terminal tail to form interactions with the bases of the ssDNA genome (Hunter, Rowitch & Perham, 1987).



**Figure 3. Atomic models of pVIII monomers as structured within phage virions A.** Model of a pVIII monomer showing hydrophobic residues (green), hydrophilic residues (orange), positively charged residues (red), and negatively charged residues (blue). N and C termini of pVIII are indicated. **B.** Spacefilling model of pVIII (Corresponding to the ribbon model in A). Images are obtained from coordinates of the RSCB PDB accession number 2cOw (Marvin et al., 2006) using PyMol (DeLano, 2006)



**Figure 4. pVIII forms a shingle-like helical array around the ssDNA genome.**

**A.** A ribbon representation of the pVIII arrangement within the filamentous phage virion. pVIII within the virion forms a shingle like array of helices. **B.** A spacefilling model (in the same orientation as the image in A) of major coat protein pVIII within the phage virions structure showing separate pVIII subunits. Number indicates order in which the subunit are added into virion structure relative to K=0 subunit. K=-6 and K=11 are the nearest neighbours of K=0 in the virion structure. **C.** Spacefilling model of the virion showing hydrophobic residues (green), hydrophilic residues (orange), positively charged residues (red) and negatively charged residues (blue). **D.** A longitudinal section showing the hydrophobic and negatively charged interior channel formed by pVIII around ssDNA genome. Properties of the residues are colour-coded as in C. Images are obtained from coordinates of the RSCB PDB accession number 2cOw (Marvin et al., 2006) using PyMol (DeLano, 2006) .

### **1.3.2 The structure and role of pVII and pIX**

Proteins pVII and pIX are incorporated into the virion at the initiation step of assembly and are the first to be extruded from the cell. pVII/pIX complex forms the proximal end of the virion (Figure 2) (Endemann & Model, 1995; Grant et al., 1980). Both are small hydrophobic proteins of only 32 (pVII) and 33 (pIX) amino acids. The structure of these two proteins has not been solved and their arrangement in the virion has not been determined. Immuno-electron microscopy studies have shown that pIX binds to cognate antibodies when within the virion structure, while pVII does not (Endemann & Model, 1995). This suggests that pVII is buried in the virion structure and is inaccessible to antibodies, while pIX is exposed to the surface of the structure.

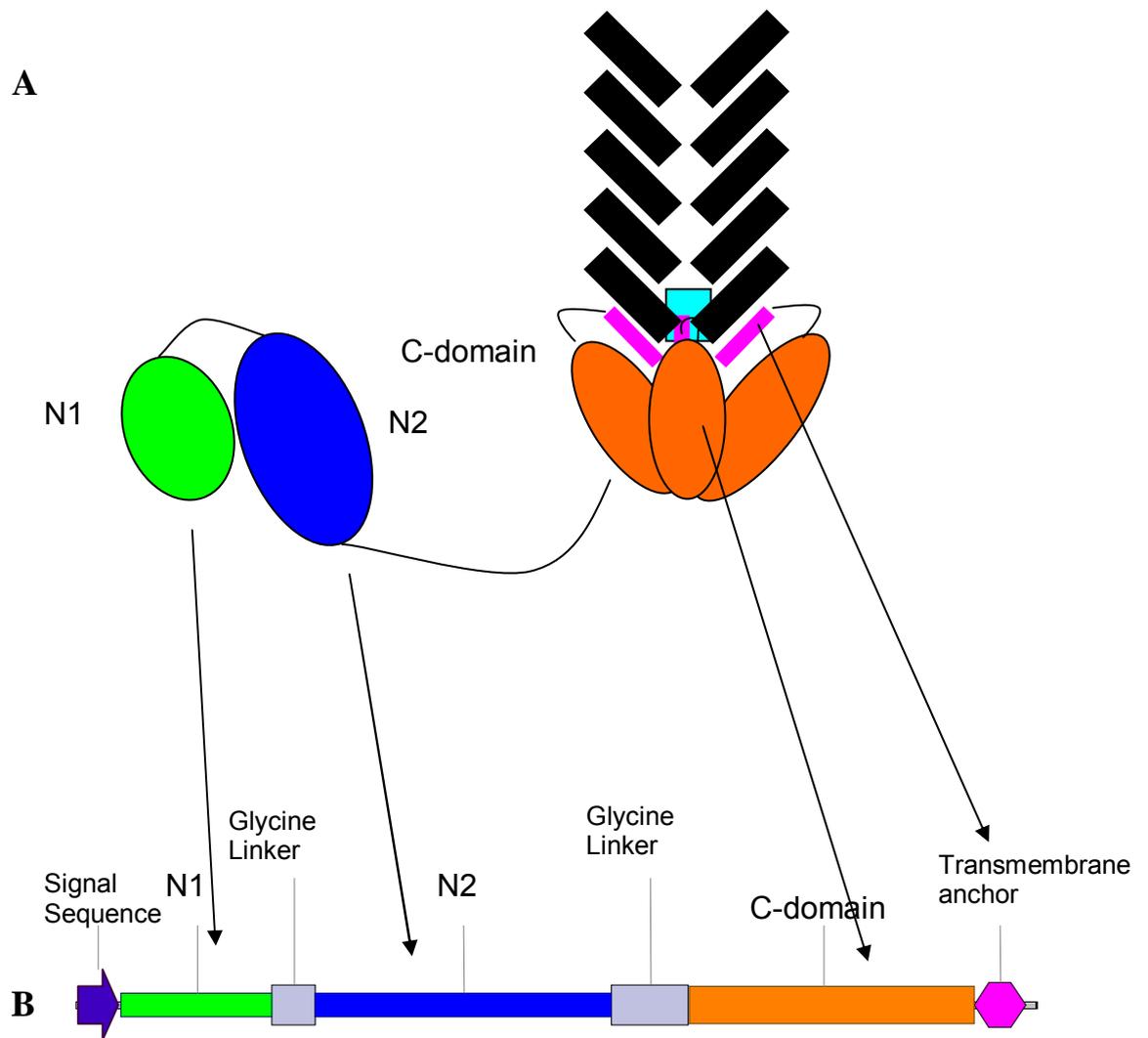
### **1.3.3 The structure and role of pIII and pVI**

Proteins pIII and pVI form the distal end of the virion (Figures 2, 5) and will be referred to in this thesis as the phage or virion cap complex. These two proteins are required for structural stability of the virion and termination of assembly (Rakonjac, Feng & Model, 1999; Rakonjac & Model, 1998). pVI is a relatively small, largely hydrophobic protein (112 aa) and before assembly into the phage, it is a membrane protein (Endemann & Model, 1995). During phage assembly, it is transferred from the membrane into the virion to form a very tight complex with pIII (Endemann & Model, 1995). This phage cap complex is so tight that even after virion disruption using surfactants (deoxycholate or Triton X-100) in combination with organic solvents (chloroform) pVI and pIII remain in a complex (Gailus & Rasched, 1994).

The estimation of number of pIII and pVI subunits within the phage cap complex has been an area of debate within filamentous phage literature. The first attempts at estimating the copy number of pIII and pVI was carried out using gel-filtration of SDS disassociated virions (Goldsmith & Konigsberg, 1977; Woolford, Steinman & Webster, 1977). Based on the apparent molecular mass of pIII it was reported that there was 3-4 copies of pIII within each virion. This seemed confirmed when Gray et al. (1979) visualised the N1N2 domains of pIII under electron microscopy. On the electron micrographs presented in that paper, three areas of density theorised to be

corresponding to the N1N2 domain of pIII were visualised. From this result, it was suggested that there must be three copies of pIII within the phage cap complex. However, Grant et al. (1981) estimated that there were five copies of pIII within the phage cap complex based on the calculated molecular weight (~46 kDa) which is very different from that observed in Laemmli SDS-PAGE system (68 kDa). Lee et al. (2002) has confirmed this conclusively, in this paper it was shown that there were 5 subunits of pIII within the phage cap complex. By using a phage display system that displayed ZnS nanocrystals on pIII it was shown by TEM that five nanocrystals could bind to the tip of the phage. This conclusively shows that there must be five copies of pIII within the phage cap complex. As for pVI, in Grant et al (1981) it was shown that pVI and pIII are within the phage cap complex at a ratio of 1:1. Therefore, if pIII is present at five copies within the phage cap complex, pVI must also be present in five copies.

At 406 amino acids in length (424aa including signal sequence), pIII is large in comparison to other four virion proteins. The domain organisation of pIII includes two N-terminal domains (N1 and N2) separated by a glycine linker and a C-terminal domain (C-domain) which is separated from the N2 domain by another glycine-rich linker (Figure 5). At the C-terminal end of the C-domain there is a hydrophobic membrane anchor (15aa) which has been well characterised (Davis & Model, 1985).



**Figure 5. Distal end of the Ff filamentous phage. A.** Schematic representation of the distal end of Ff filamentous phage, showing domain organisation of the phage cap complex. pVIII (Black rectangle) and pVI (Light blue rectangle) are also shown. **B.** pIII protein domain arrangement in the primary structure. The signal sequence is cleaved off during membrane translocation of the periplasmic portion of pIII. Arrows align spatial model to the primary sequence arrangement of pIII domains.

After translation, the pIII protein is targeted to the SecYEG translocator in a SecA-dependent manner. It is then inserted into the cytoplasmic membrane and is anchored in membrane by its C-terminal anchor domain. The N-terminal domains and most of the C-terminal domain are in the periplasm, leaving only five residues of the C-terminus in the cytoplasm (Boeke & Model, 1982; Davis, Boeke & Model, 1985).

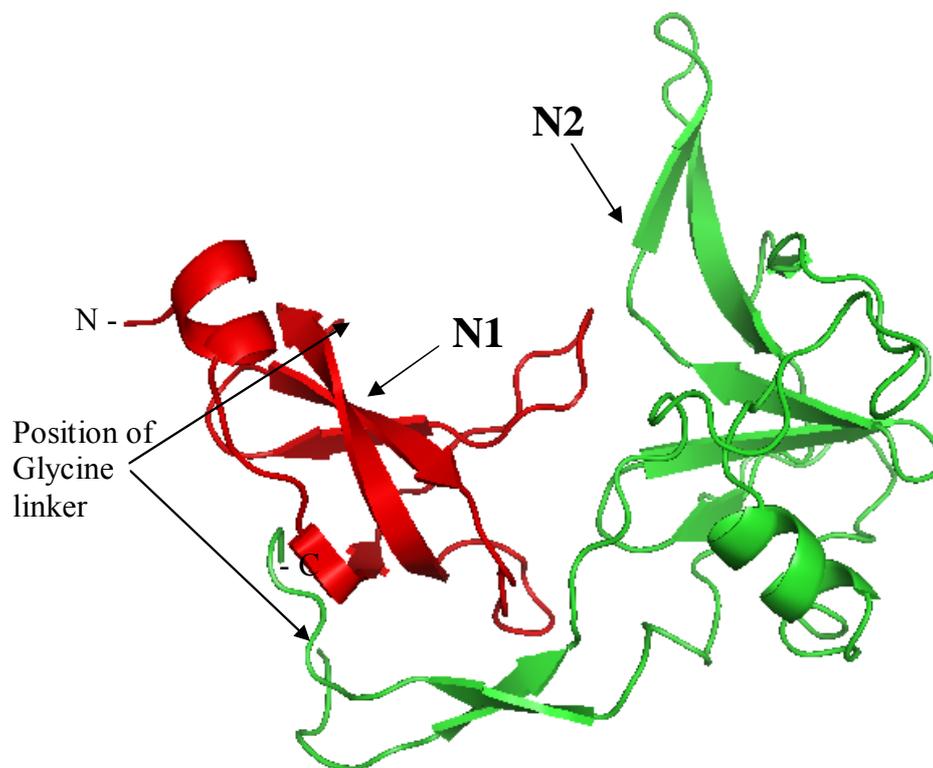
The structure of N1 and N2 domains has been determined using X-ray crystallography and NMR (Holliger, Riechmann & Williams, 1999; Lubkowski et al., 1998) (Figure 6). This was achieved because these two domains can fold independently of the C-domain. An X-ray crystallography structure has also been solved for a complex of the N1 domain and the phage co-receptor; the domain III of the TolA protein (Lubkowski et al., 1999; Reichmann & Holliger, 1997). These analyses show that the N1/TolA co-receptor binding site is on the interdomain face between N1 and N2, and thus is usually covered by N2 domain prior to F pilus binding (Deprez et al., 2005). It has been shown that the N2 domain serves as the F pilus binding site during phage infection (Caro & Schnös, 1966; Stengele et al., 1990). The kinetics and mechanism of the N-terminal domain folding during protein production has been studied in detail (Martin & Schmid, 2003a; Martin & Schmid, 2003b). The kinetics and structural changes within the N1N2 domains during binding to the respective host receptors; TolA and the tip of the F-pilus, has also been studied (Eckert et al., 2007; Eckert & Schmid, 2007).

The three-dimensional structure of the C-domain is yet to be determined. However, a study of C-terminal domain portion immediately upstream of the membrane anchor using alanine scanning mutagenesis has been carried out (Weiss et al., 2003). This study identified the residues required for the incorporation of pIII into the virion. An antibody specific for the C-terminal 10 residues of pIII (FANILRNKES, R164) cannot bind to pIII when it is incorporated into the virion (Rakonjac, unpublished research). Therefore, this C-terminal peptide must be buried within the virion cap, containing the pIII C-domain and pVI. The C-domain of pIII also contains a di-sulphide bond between Cystiene 354 and Cystiene 371 and this bond is essential for infectivity (Kremser & Rasched, 1994).

It has been found that the final 93 residues of the C-domain of pIII are essential for the release of the phage from the cell (Rakonjac et al., 1999; Rakonjac & Model, 1998). If

*gIII* (gene III) is deleted from the phage genome or if a fragment comprising of the C-terminal 83 residues is used to complement a *gIII* deletion, then phage are not released from the cell and instead form cell-associated filaments on the cell surface. Any virions released by cell infected under these conditions are derived from cell-associated filaments that have been broken off host cell by shear forces. These filaments do not contain pIII and thus are non-infectious. In addition, these broken off cell-associated filaments are unstable and can be disassembled in the detergent Sarkosyl. Likewise, phage terminated with C-domain fragments shorter than the C-terminal 132 residues of pIII are also unstable in the detergent Sarkosyl (Rakonjac et al., 1999). Virions terminated with C-domain fragment longer than the C-terminal 132 residues are stable in the detergent Sarkosyl.

The pIII C-domain is essential for phage infection. In a study performed by Bennett N and Rakonjac J (2006) it was shown that if a termination-incompetent C-domain of pIII (C83) was covalently linked to functional N1N2 receptor-binding domains (NdC83) and was then used to complement a *gIII* deletion phage, the resulting virions were non-infectious, despite their ability to bind to the host cell receptors. This means that the C-domain must have a role in the infection process.



**Figure 6. Three-dimensional structure of pIII N1N2 domains.** The ribbon representation of the N1 domain (Red) and the N2 domain (Green). The glycine linker located between the C-terminus of the N1 domain and the N-terminus of the N2 domain is flexible and hence is not visible in this structure. The positions of the termini of the glycine linker are indicated. This figure is derived from the from the RCSB PDB database accession number 1g3p which is based on the work presented in (Lubkowski et al., 1998). Image obtained using PyMOL (DeLano, 2006).

## 1.4 The Ff phage lifecycle

### 1.4.1 Infection of host cell

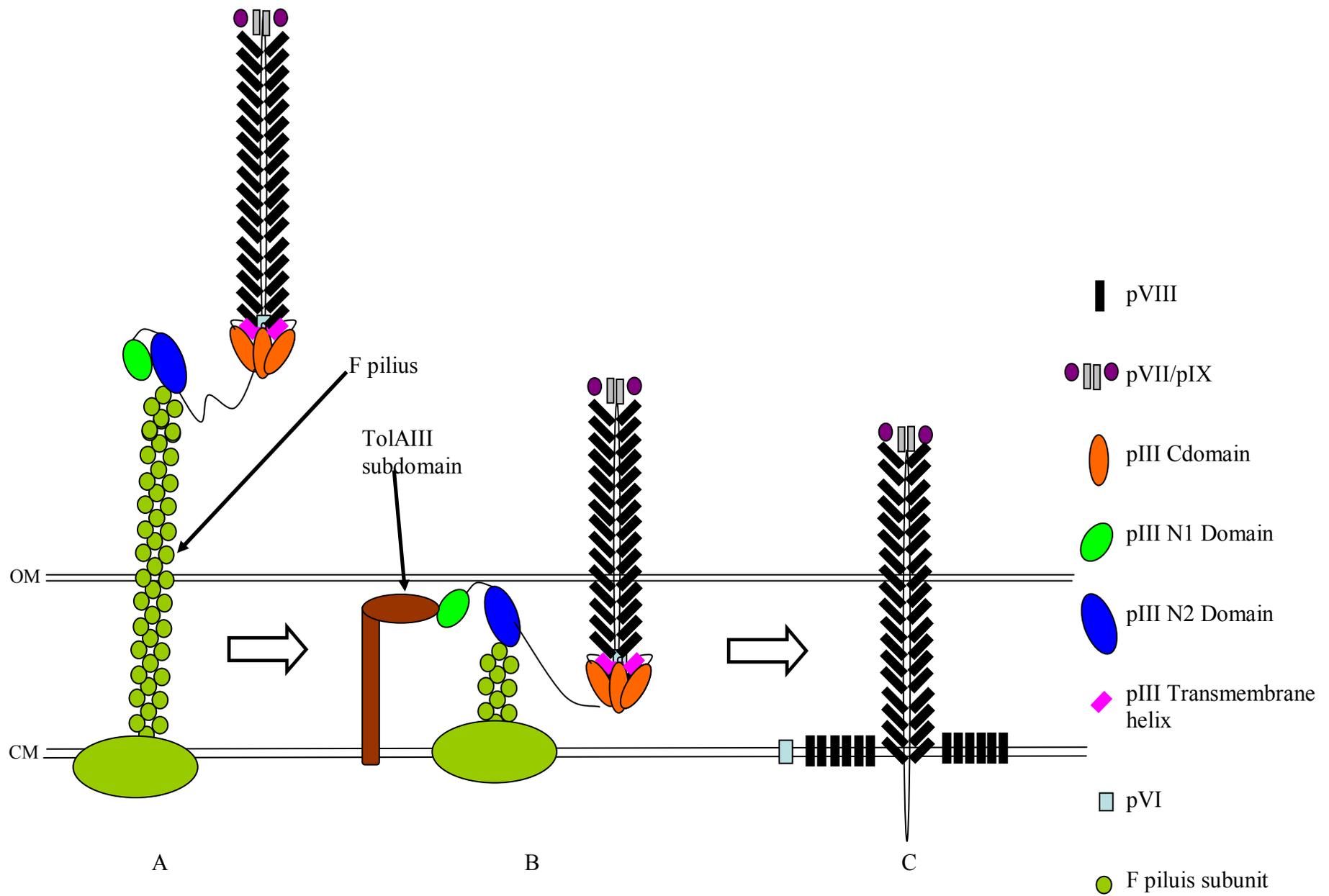
Ff filamentous bacteriophage begins the process of infecting an *E. coli* host by binding to the tip of the F pilus (Caro & Schnös, 1966) (Figure 7) via the N2 domain of pIII. The F pilus is the primary receptor for Ff phage, however it is not the only protein required by the phage to infect the bacteria. Ff phage also requires interactions to occur between the N1 domain and the TolA protein, which is part of the cytoplasmic membrane complex TolQRA. The N1 domain of pIII binds *via* the periplasmic domain III of TolA. If the F pilus is not present due to the host cell lacking the F conjugative plasmid (termed F<sup>-</sup> or “female” *E. coli*), the infection efficiency decreases by several orders of magnitude. If the cytoplasmic membrane proteins Tol Q, R or A are mutated or missing then Ff phage can not infect the *E. coli* (Bradley & Whelan, 1989; Smilowitz, 1974a) suggesting that either the phage forms contacts with TolQ and TolR in addition to TolA, or that the intact complex is required for infection.

The N2 domain of pIII binds to the tip of the F pilus (Figure 5). The F pilus binding site on the N2 domain has been identified by mutagenesis (Deng & Perham, 2002). However, the binding site for N2 on the F pilus tip is yet to be identified. Binding to the F pilus causes a conformational unfolding reaction (Eckert & Schmid, 2007) to occur between the N1 and N2 domains mediated by the cis-trans isomerization of Pro213 within the N2 domain (Eckert et al., 2007). This releases the N1 domain from the N2 domain, exposing the TolA binding site on the N1 domain (Eckert et al., 2007; Eckert & Schmid, 2007; Lubkowski et al., 1999; Reichmann & Holliger, 1997). The N2 domain binding to the F pilus tip also causes the F pilus to retract, bringing the phage into closer contact with the cell, allowing the interaction between N1 and TolA. It is not clear how the pIII N1 domain, and the rest of the phage, crosses the outer membrane to gain access to the TolAIII domain in the periplasm. The N1 domain binds to the periplasmic/outer membrane-associated domain III of TolA (Lubkowski et al., 1999). This interaction is thought to cause a conformational change leading to the "opening" of the phage structure, allowing membrane insertion, virion coat integration into the host cell

membrane (Smilowitz, 1974b; Trenkner, Bonhoeffer & Gierer, 1967) and ssDNA uptake.

There is evidence that the N1 and N2 domains must be covalently linked to the functional C-domain at a specific distance for membrane insertion to occur. Evidence presented in the papers by Krebber et al. (1997) and Spada et al. (1997) showed that if a large protein domain ( $\beta$ -lactamase) was placed between the N2 domain and the C-domain, then infectivity was decreased hundred fold compared with the wild type phage. The data from Krebber et al. (1997) also showed that a N1N2 complex non-covalently bound to the C-domain results in four orders of magnitude reduction in infectivity compared to that of the WT phage. The authors replaced the N1N2 domains of pIII with a ligand-binding protein-C-domain fusion and then expressed the ligand independently as a fusion to N1N2. The idea was that when the ligand binding protein binds its ligand then the phage should become infective. However when this was tried infectivity was low (0.0001% that of the wild-type phage).

Direct evidence of involvement of the C-domain in infection was shown in Bennett & Rakonjac (2006). This paper found that that infection did not occur if the N1N2 domains of pIII were covalently attached to a termination incompetent C-domain, even though a functional C-domain was also present within the phage cap complex. This means that firstly the C-domain must play an active role during the process of phage infection and secondly that the C-domain mediated entry *in cis* with the N1N2 domains covalently bound to it. An intact C-domain cannot mediate infection if it is not bound covalently to the receptor binding N1N2 domains.



**Figure 7. Model of phage infection.** **A.** Phage binds to the tip of the F pilus through the N2 domain of pIII. F pilus retracts and N1 domain is released. **B.** N1 domain binds to TolQRA complex via the TolAIII domain. The phage structure "opens" via a conformational change of the C-domain allowing membrane insertion of phage. **C.** Coat proteins integrate into the inner (cytoplasmic) membrane, ssDNA genome transverses the membrane and enters the host cell.

### **1.4.2 Resistance to filamentous phage infection caused by pIII**

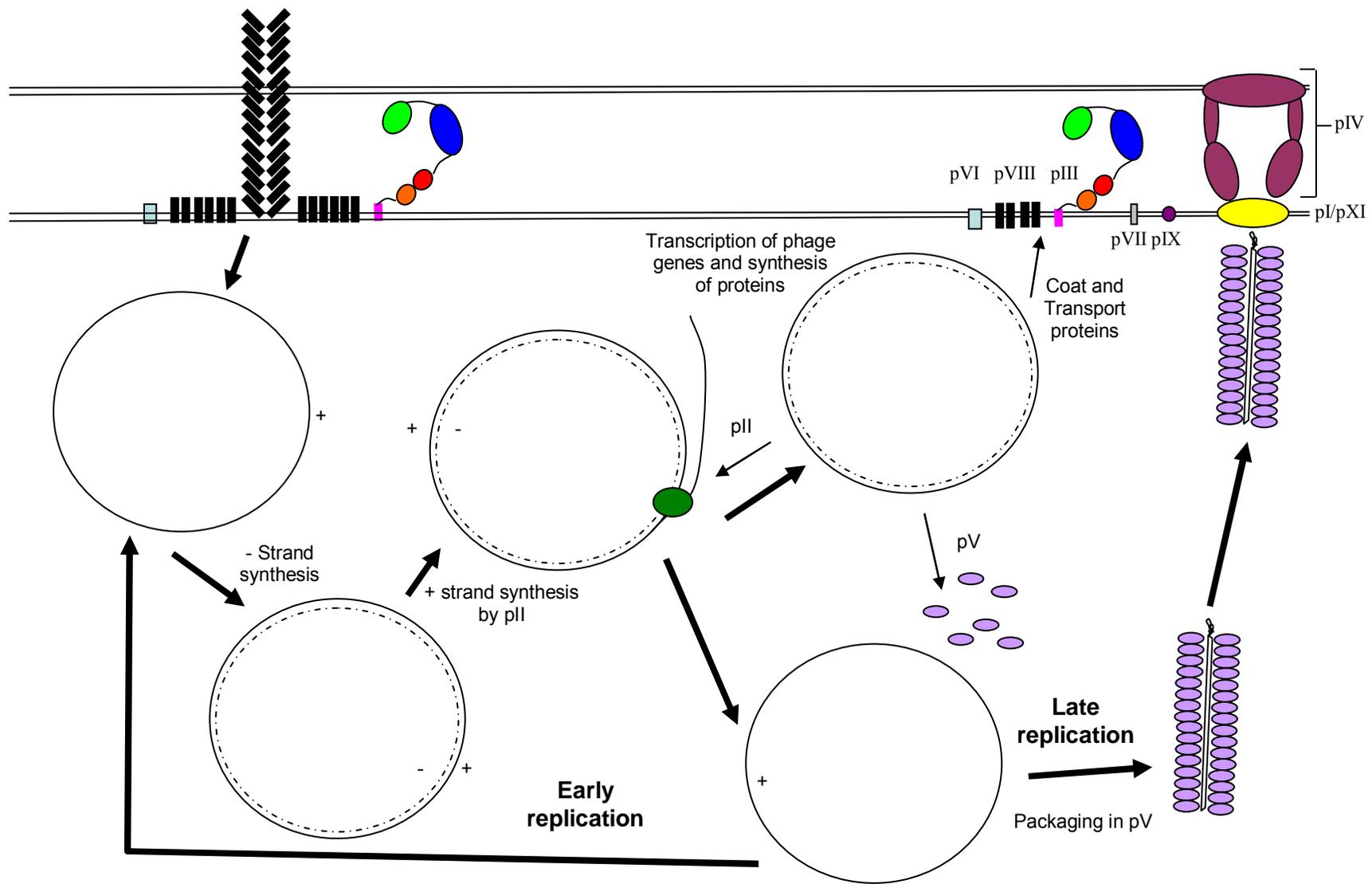
Cells can become resistant to Ff filamentous phage infection in many ways. One of these ways is if pIII N1 or N2 domains (or both) are expressed in the host cells. This is particularly a problem when producing phage stocks of  $\Delta gIII$  helper phage.

Resistance by the *in-trans* expression of pIII occurs because the N1 and N2 domains of pIII block the TolA receptor and cause the retraction of the F pilus, respectively, rendering them unavailable for infection (Boeke, Model & Zinder, 1982). To resolve this effect, *in-trans* pIII expression systems need to be tightly controlled so that there is no pIII expression before phage infection, but yet is highly inducible upon phage infection.

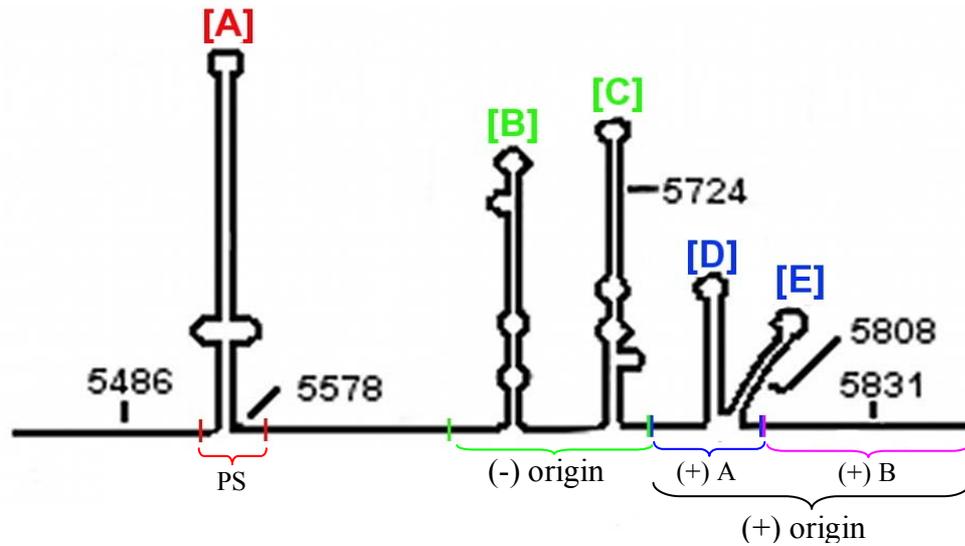
### **1.4.3 Replication of phage within the host cell**

Once the Ff phage ssDNA genome positive strand enters the cytoplasm of the host cell, the complementary negative strand is synthesised by the host DNA replication machinery (Figure 8). The dsDNA form of the Ff phage genome is referred to as the replicative form (RF), whereas the positive strand (+) ssDNA form of the Ff phage genome is referred to as the infective form (IF). The replicative form of the genome serves as a template to synthesise the DNA positive strand and for transcription of the phage genes. During early infection newly synthesized (+) strands are recycled to create more dsDNA replicative forms. Later in the infection, positive (+) strands are coated with pV and serve as a packaging substrate in the phage assembly (Model & Russel, 1988).

Newly synthesised phage proteins are targeted to their predetermined locations: pII, pV and pX remain in the cytoplasm and pI, pIII, pIV, pVI, pVII, pVIII, pIX, and pXI are targeted to membranes (Endemann & Model, 1995; Russel & Kazmierczak, 1993; Smilowitz, Carson & Robbins, 1972; Trenkner et al., 1967).



**Figure 8. Overview of life cycle of Ff filamentous phage.** The ssDNA genome is transported across the cellular membrane and released into cytoplasm of the cell. Synthesis of the negative strand is initiated at the negative strand origin of replication by RNA polymerase, which generates an RNA primer and then becomes released from the template (Zenkin et al., 2006). Host DNA polymerase III uses this primer to replicate the negative strand. Positive strand synthesis uses a rolling circle mechanism and is initiated by pII at the positive origin of replication. During the initial period of viral infection, positive strands are used as templates for synthesis of the negative strands, resulting in an increase in copy number of the dsDNA replicative form (RF). The RF serves as a template of phage proteins. Phage proteins II, V and X remain in cytoplasm and mediate genome replication and formation of the packaging substrate. Proteins pI, pIV and pXI form a transport complex spanning the inner and outer membrane. Virion proteins pVII, pIX, pVIII, pVI, pIII are inserted into the membrane in preparation for phage assembly. Later in the infection, positive strands are coated in the phage encoded single stranded DNA binding protein pV to form the packaging substrate and brought to the cell membrane assembly/export complex for assembly and export.



**Figure 9. Secondary structure of the f1 phage intergenic sequence** [A], Packaging signal (PS); [B] and [C], (-) origin, negative strand origin of replication; [D] and [E], (+) A, the A region of the positive strand origin of replication, (+) B, The B region of positive strand origin of replication. Numbers indicate nucleotide positions according to f1 sequence (Hill & Petersen, 1982). Diagram is adapted from (Speckthreie et al., 1992).

```

                                Packaging Signal
                                ~~~~~
5501  CGGCGCATTAGCGCGGCGG GTGTGGTGGT TACGCGCAGC

                                Packaging Signal
                                ~~~~~
5541  GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT

Packaging Signal                Phage negative origin
~~~~~                            ~~~~~
5581  TCGCTTTCTT CCCTTCTTT CTCGCCACGT TCGCCGGCTT

                                Phage negative origin
                                ~~~~~
5621  TCCCCGTCAA GCTCTAAATC GGGGGCTCCC TTAGGGTTC

                                Phage negative origin
                                ~~~~~
5661  CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAACTTG

                                Phage negative origin
                                ~~~~~
5701  ATTAGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA

                                Positive origin part A
                                ~~~~~
                                pII Nicking Site
                                ↓
5741  GACGGTTTTT CGCCCTTGA CGTTGGAGTC CACGTTCTTT

                                Positive origin part B
                                ~~~~~

                                Positive origin part A
                                ~~~~~
5781  AATAGTGGAC TCTTGTTCCA AACTGGAACA ACACTCAACC

                                Positive origin part B
                                ~~~~~
5821  CTATCTCGGT CTATTCTTTT GATTTATAAG GGATTTTGCC

                                Positive origin part B
                                ~~~~~
5861  GATTCGGCC TATTGGTT

```

**Figure 10. Intergenic region sequence and positions of various features.** Colour coding of feature labels corresponds to that of loops and regions indicated in Figure 7.

#### 1.4.4 The intergenic region and replication

The intergenic region (between *gIV* and *gII*) contains the positive and negative origins of replication, and the packaging signal or morphogenetic signal (Figures 9 and 10). It is the site at which the replication of the phage genome is initiated.

The synthesis of the negative (-) strand starts from the negative (-) strand origin of replication (Figure 7). The conversion of the single stranded (+) DNA infective form into a double stranded DNA replicative form does not require any phage genes. The negative origin serves as a starting site for the host cells RNA polymerase to synthesize a primer (Gray et al., 1978). It then stalls and disassociates from the template (Zenkin et al., 2006). DNA polymerase III then uses this primer to synthesize the (-) DNA strand. The negative strand origin is not absolutely required for phage replication (Kim, Hines & Ray, 1981); the RNA primer for DNA replication can be synthesized at other locations in the genome, albeit with a lower efficiency (Kim et al., 1981). Thus, phage carrying mutation in the negative origin have very low infectivity (Kim et al., 1981).

The positive (+) strand origin is absolutely required for phage replication and packaging (Figure 9). The product of the replication cycle initiated from this origin of replication is the circular ssDNA that is packaged into the virion. Phage proteins pII and pX are required for the (+) strand synthesis. Protein pV has two functions in the Ff phage lifecycle, one of which is to coat ssDNA phage genomes in the cytoplasm for packaging and export. The other purpose is to negatively regulate the expression of pII (Michel & Zinder, 1989). This regulatory loop serves to coordinate ssDNA production and packaging.

The positive origin can be divided into two regions designated A and B (Figures 9 and 10). The A region (5769-5819) is the core region for positive strand synthesis. It contains the pII nicking site (5780) and the sequence required for initiation and termination of positive strand synthesis. Point mutations in the A region of the positive origin reduce positive strand synthesis typically by 10 000 fold (Dotto, Enea & Zinder, 1981; Dotto et al., 1984; Dotto & Zinder, 1983; Dotto & Zinder, 1984). The B region extends 100bp downstream of the A region. The B region is also required for initiation

of positive strand synthesis. Mutations in the B region reduce positive strand synthesis by about 100 fold (Dotto et al., 1984; Dotto, Horiuchi & Zinder, 1982; Dotto & Zinder, 1984; Zinder & Horiuchi, 1985).

The phage-encoded protein pII nicks the positive origin at the position 5780 (Meyer & Geider, 1979), leaving a free 3' end that serves as a primer for the host DNA polymerase III which synthesises the new (+) strand (Meyer & Geider, 1982). The old (+) strand is displaced as the new one is synthesised. When the replication cycle of genome is complete, pII cuts and ligates the ends of the displaced (+) strand. This displaced (+) strand can then be either coated in pV for phage export (Salstrom & Pratt, 1971) or converted, using the host cell DNA replication machinery and the negative origin of replication, into dsDNA (RF form).

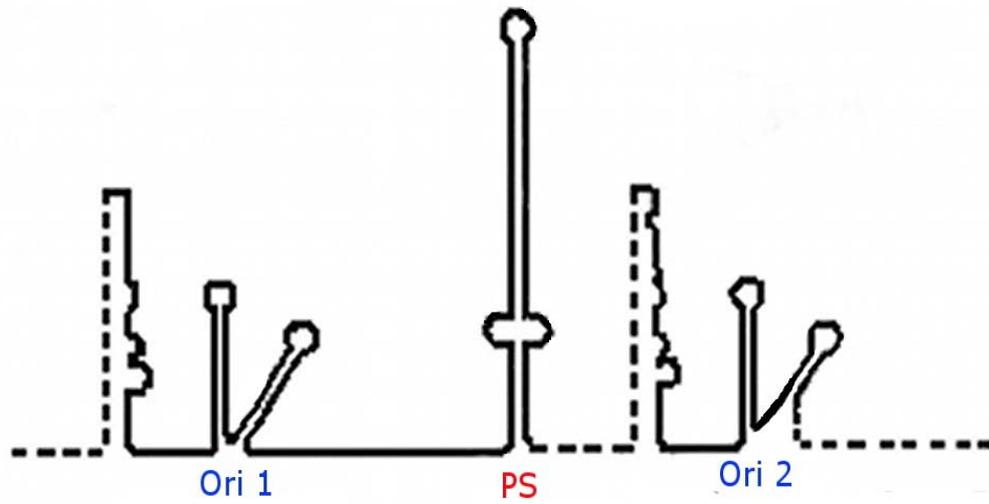
The packaging or morphogenetic signal is a hairpin loop (Figure 9). It acts as a targeting signal for the pI/pXI/pIV phage export complex as well as assisting the minor proteins pVII and pIX in identifying phage ssDNA genomes so that they can be packaged into the phage virions and exported (Russel, 1993). Only the packaging signal on the positive strand is recognised by the pI/pXI/pIV phage export complex and packaged (Zinder & Horiuchi, 1985). It has been noted however, that the export system will package phage and other plasmids that replicate via a rolling circle mechanism, even though they lack the packaging signal. However the efficiency of this occurring is low, at approximately 0.1% of the normal rate (Russel & Model, 1989; Specthrie et al., 1992).

#### **1.4.5 The microphage-producing origin replication**

A microphage (or defective infective particles) is defined as a phage virion or particle which is approximately 20-fold shorter than the wild-type phage. In Ff filamentous bacteriophage, shorter than full length virions are sometimes naturally produced, often after about 40 passages of the phage through host cells in the absence of clonal (plaque) purification. Microphage production arises when the positive origin of replication is duplicated in the same genome (La Farina et al., 1987). The (+) strand replication is initiated at the upstream origin (*oriI*). When the replication fork reaches the next

positive origin (*ori 2*), it provides a termination signal where pII makes another cut and then ligates the two ends. This creates a small circle of ssDNA between *ori1* and *ori2* that is then packaged into the virion. Given that the virion length is determined by the size of packaged ssDNA, the resulting virions are very short. Microphage production is inefficient and only represent only a small fraction (0.1%) of the mass of the total phage produced (La Farina et al., 1987).

An engineered system for producing microphage was presented in Specthrie et al (1992). This system uses a plasmid containing a specifically modified IG sequence as a replication template to generate a 200nt circular ssDNA as a packaging substrate to produce microphage particles. The IG of the microphage-producing plasmid contains two positive origins, *ori1* and *ori2*, separated by a packaging signal, placing them approximately 200 nt apart (Figure 11). The *ori1* is a wild-type positive origin and can serve as both an initiator and terminator of positive strand replication, whereas the *ori2* positive origin has been mutated so that it can only serve as a terminator of rolling circle replication, but not a point of initiation (Specthrie et al., 1992). When pII binds to the *ori1* it nicks the (+) strand and replication starts. DNA replication proceeds through the packaging signal and reaches the next (mutated) positive origin which serves as a terminator. pII then catalyses the (+) strand transfer reaction to form a 200bp ssDNA circle which is packaged into microphage virions. This system requires a helper phage to provide the necessary proteins for virion replication and assembly. Microphage may be useful for structural studies of the filamentous phage termini due to their reduced length to width ratio compared to WT virions.



**Figure 11. Secondary structure of an engineered microphage origin of replication**  
*Ori 1*, Wild-type (full length) filamentous phage positive origin; PS, filamentous phage packaging signal; *Ori 2*, modified filamentous phage positive origin that supports rolling circle replication termination but not rolling circle replication initiation (Speckhrie et al., 1992).

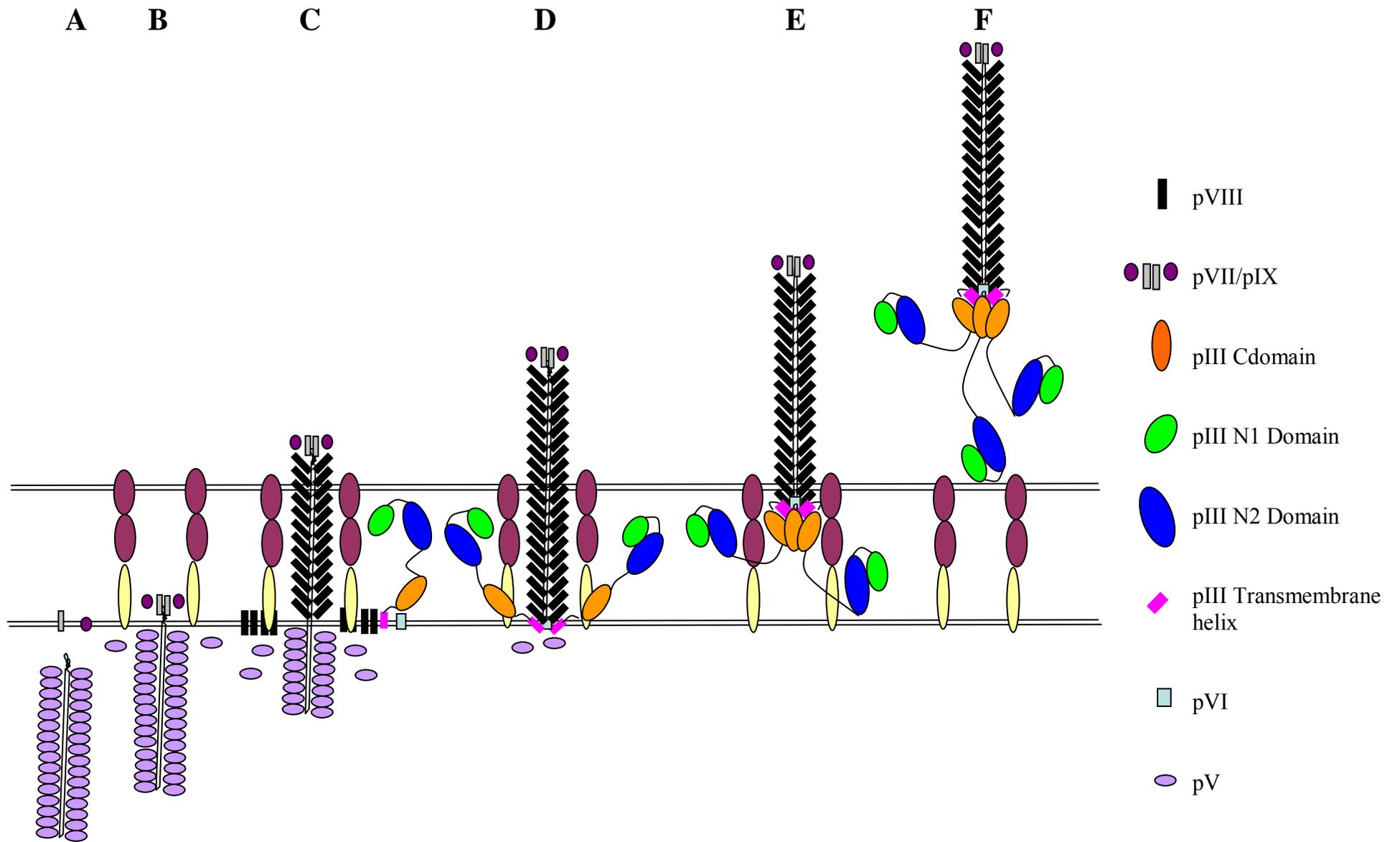
#### **1.4.6 Interference resistance and phage replication**

Because filamentous phage produce ssDNA, they have been used extensively in the past for DNA sequencing. Special vectors were constructed for easy cloning and ssDNA production (Vieira & Messing, 1982; Zinder & Boeke, 1982) called phagemid plasmids. These vectors contain both a plasmid origin of replication and filamentous phage intergenic sequence containing a filamentous phage origin of replication. To produce ssDNA templates for DNA for sequencing, the fragments of interest were inserted into the vector and cloned into a host cell. The host cell was then infected with a helper phage, to induce replication from the filamentous phage origin of replication on the phagemid vector and thus induce production of ssDNA and packing into the virion. This system produced two types of virions, first containing the phagemid ssDNA and second containing the helper phage ssDNA. Produced virions could then be used to purify ssDNA for sequencing. With the advent of PCR and high-throughput methods of sequencing this method is no longer used. Currently phagemid vectors are used in phage display technology. Phage display is one of the major combinatorial methods used for affinity screening of peptides, antibodies or cDNA libraries for clones that bind to a ligand of interest (Barbas III et al., 2001; Cramer et al., 1994; Lowman et al., 1991; Makowski, 1993; McCafferty et al., 1990; O'Neil & Hoess, 1995; Smith, 1985; Zwick, Shen & Scott, 1998).

When phagemid vectors were first used, the titre of helper phage and phagemid virions produced was low due to a phenomenon called interference. Interference in phage replication is a phenomenon in which one phage origin of replication (phagemid or microphage origin) affects the replication of another phage origin of replication (helper phage). In filamentous phage, interference can be seen as a hundred-fold reduction in phage and phagemid virion titres. This is because in filamentous phage, early infection is the point where the copy number of phage genomes in the cell is set (Model et al., 1982). The process of setting the copy number of phage genomes within a cell involves phage protein pII. Because the phagemid plasmid is already present in multiple copies at the time the host cell is infected with helper phage, pII will preferentially bind to the phagemid origin of replication. This means that phagemid, which do not carry any of the phage genes, become counted as phage genomes. This reduces the number of double

stranded helper phage DNA templates available for phage gene transcription, meaning that less phage proteins are produced. Overall this results in a lower titre of both phage and phagemid (microphage) virions (Enea & N.D Zinder, 1982; Model & Russel, 1988).

To overcome this limitation, interference-resistant helper phage have been isolated (Enea & N.D Zinder, 1982; Vieira & Messing, 1982). The interference-resistant phage used in this thesis all carry the IR1 mutations conferring interference resistance. IR1 is a notation for a pair of mutations; one within the pII regulatory sequence and the other within pII as defined by Enea and Zinder (1982), which in combination results in preferential binding of pII to the helper phage origin of replication.



**Figure 12. Phage secretion, termination and release from the host cell.**

**A.** The positive strand is introduced into the export machinery starting from the packaging signal (proximal end). **B.** The packaging signal forms interactions with the export machinery proteins pI and pXI and virion proteins pVII, pIX and pVIII. The order in which this occurs is unknown. **C.** After the initiation of assembly the phage is extruded through the cytoplasmic membrane as more pVIII monomers are added. As this happens pV disassociates from the (+) strand and stays in the cytoplasm. The process of export requires ATP hydrolysis, proton-motive force and thioredoxin(Feng, Russel & Model, 1997). **D.** Once the genome has been fully covered by pVIII, the minor coat proteins pIII and pVI are added to the structure. **E.** It has been proposed that pIII undergoes a structural change in the C-domain capping the phage structure and releasing the virion from the cell membrane (Rakonjac et al., 1999). **F.** The virion can now be fully released from the cell and into the growth medium.

### 1.4.7 The Ff phage export apparatus and phage assembly/secretion

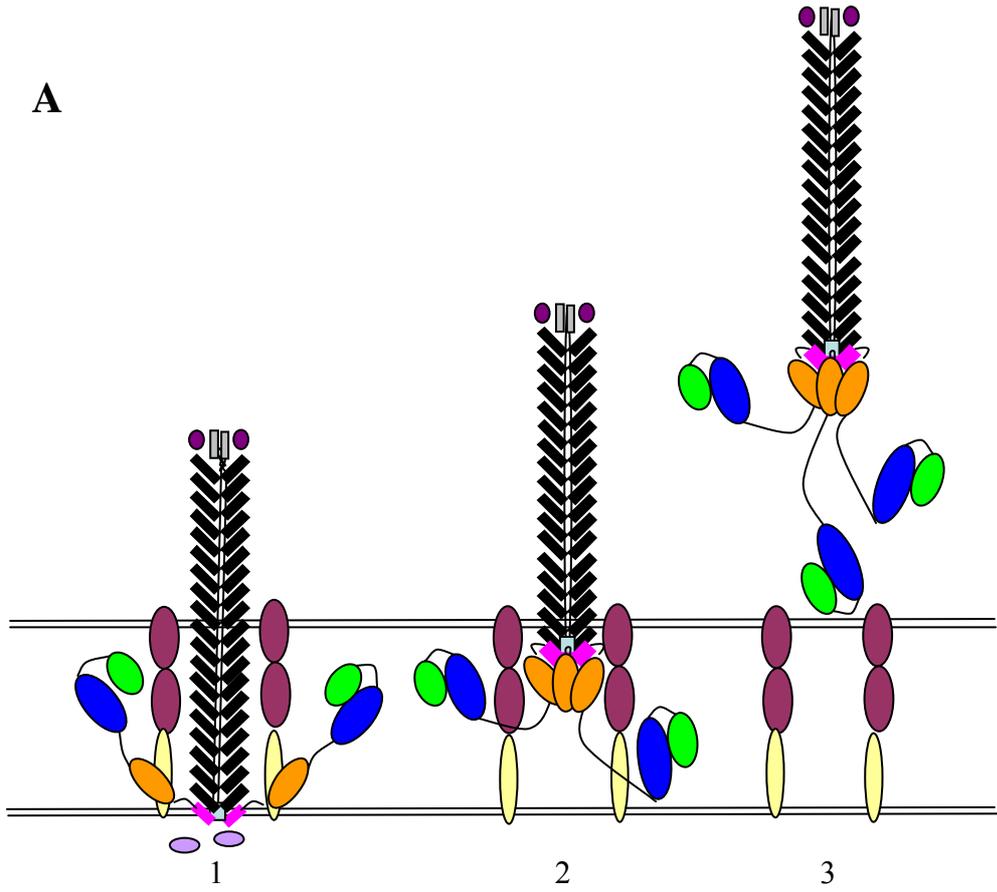
Ff assembly and export from the cell resembles the secretion of virulence factors or the assembly of the bacterial cell surface structures in Gram-negative bacteria. Encoded by the phage genome are the two genes, *gI* and *gIV*, that form the phage assembly and exporting machinery. One gene (*gI*) encodes two proteins, pI and pXI, which form the cytoplasmic (inner) membrane portion of the phage assembly complex. The structure of the inner membrane component of the phage assembly complex is unknown (Figure 12). The virions are assembled at the cytoplasmic membrane and at the same time are extruded through a large gated channel in the outer membrane which is encoded by phage gene *gIV*. This channel is composed of 14 pIV subunits (Linderoth et al., 1997; Marciano, Russel & Simon, 2001; Opalka et al., 2002). Its quaternary structure is a large ring with an internal diameter of 8nm. When analysed by cryo-electron microscopy, it was found that the channel is blocked by a septum like density. pIV belongs to a large family of outer membrane channels termed secretins, which are common outer membrane components of the type II and III secretion systems and the type IV pilus assembly systems in Gram-negative bacteria (Craig, Pique & Tainer, 2004; Galan & Collmer, 1999; Johnson et al., 2006; Kubori et al., 1998).

The intracellular substrate for the production of virions is ssDNA coated with ssDNA binding protein pV. Coating of the positive strand ssDNA genomes in ssDNA binding protein pV leaves only the packaging signal exposed. The uncoated packaging signal serves to target the ssDNA/pV complex to the phage assembly complex (Figure 12A). The packaging signal (proximal end) is recognized by the phage assembly complex in the inner membrane where it interacts with pVII, pXI, and pVIII (Figure 12B) to initiate assembly. After initiation the ssDNA genome is passed through the pI/pXI complex with pV being removed and replaced by major virion coat protein pVIII (Figure 12C). When the phage ssDNA genome is fully packaged in pVIII, minor coat proteins pVI and pIII are added to the phage virion (Figure 12D) causing the termination of assembly and the release of the phage from the cell (Figure 12E, F).

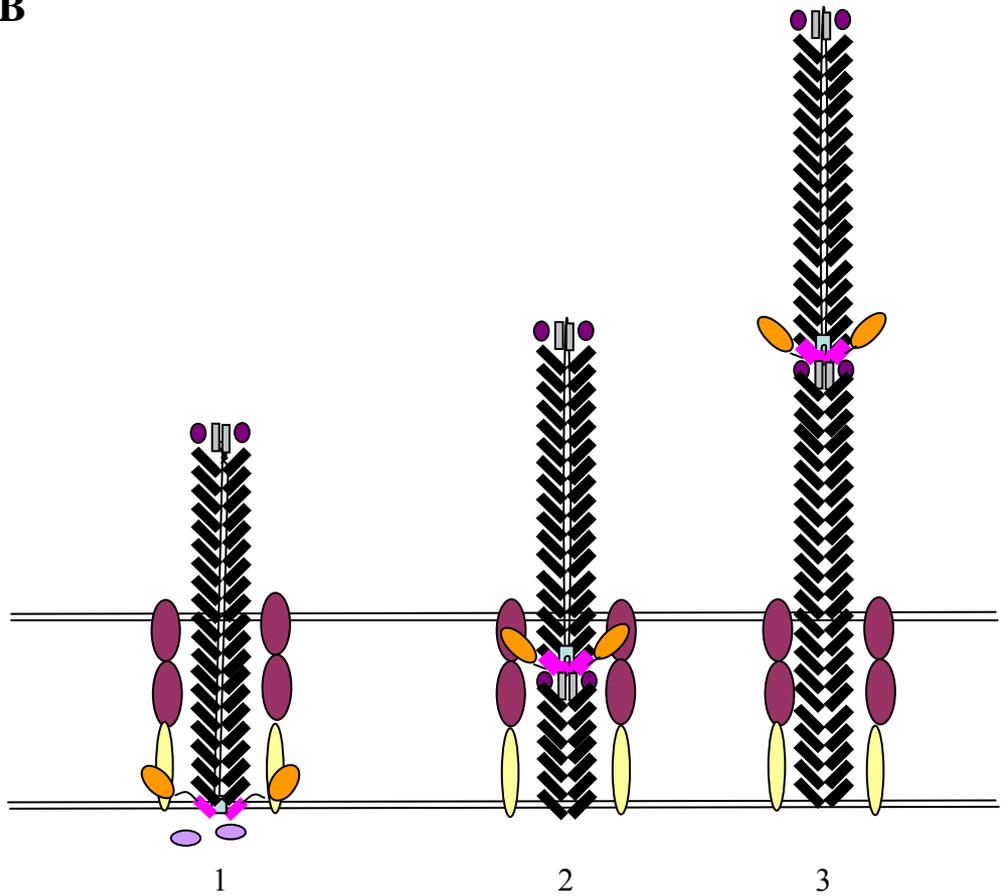
Genetic evidence suggests that the C-terminal region of pIII is involved in the termination of elongation and release of the phage (Crissman & Smith, 1984; Nelson,

Friedman & Smith, 1981; Rakonjac et al., 1999; Rakonjac & Model, 1998; Zacher et al., 1980). In Rakonjac J et al. (1999) it was shown that if the C-domain was truncated from 154 to only the final 93 residues, then termination of the phage could still occur. If the truncation was to the final 83 residues of the C-domain, phage termination did not occur. In this case, the phage virion was not released from the cell. However, elongation did continue with the virions becoming long cell-associated filaments, and the 83-residue C-terminal fragment was incorporated into these cell-associated filaments (Figure 13). A model that a conformation change of the C domain during pIII addition to the virion was responsible for phage termination and release from the host cell was proposed in Rankonjac. J et al (1999). Given that the structure of the C-terminal region of pIII is unknown, it is also unknown how the pIII C-domain structure changes, upon incorporation into the virion, to mediate termination of assembly and release of the virion.

**A**



**B**



**Figure 13. The production of cell-associated phage filaments. A.** Normal phage assembly termination. 1, ssDNA genome is fully packaged, pIII and pVI are incorporated into virion filament; 2, The C-domain of pIII undergoes a conformational change, resulting in termination phage assembly; 3, Virion is released from the cell. **B.** Phage assembly in the presence of C-domain truncated to the 83 C-terminal fragment. 1. Truncated pIII is incorporated into the virion; 2, C-domain does not undergo a conformational change. Virion does not terminate, a new genome enters the phage assembly complex; 3, Virion extension continues without termination resulting in cell-associated filaments.

## 1.5 Other viral mechanisms of entry

The mechanism of filamentous phage infection is distinct from all other bacteriophage. Unlike the classical known tailed phage, e.g.  $\lambda$  and T-even series, whose virion remains extracellular after the genome-entry into the host cell (Hershey & Chase, 1952), the major virion protein of the Ff virion integrates into the host cellular membrane as the ssDNA genome traverses the membrane (Smilowitz, 1974b; Trenkner et al., 1967) .

Tailed phage, like  $\lambda$  and the T even series, bind to the host cellular receptors on the surface of the cell. They then inject their genome across the host cellular envelope. This is accomplished by either a syringe-like contractile tail (Hershey & Chase, 1952; Kanamaru et al., 2002) or a non-contractile tail that releases proteins which then in turn form a channel through the cell membrane. In all these phage, after genome transfer has occurred, the empty shell of the phage virion coat remains on the outside of the infected host cell. In these phage, the transport of DNA from the phage to the host cell during infection occurs due to mechanical force provided by the high internal pressure within the phage head which are developed during phage genome packaging into the virion (Smith et al., 2001) .

Another interesting and rare mechanism of entry in bacteriophage is the mechanism employed by *Pseudomonas* phage  $\phi 6$ , a member of the Cystovirus family. This phage has a protein-lipid outer layer (Hu et al., 2008). During phage infection, the bacteriophage outer layer fuses with the host's outer membrane, releasing a nucleocapsid into the periplasmic space. This process is facilitated by phage protein p6 (Bamford, Palva & Lounatmaa, 1976; Bamford, Romantschuk & Somerharju, 1987). The nucleocapsid then traverses the periplasmic space after peptidoglycan digestion by virion-encapsidated peptidoglycan-digesting enzymes (Caldentey & Bamford, 1992). The virus then crosses the cellular membrane by membrane invagination and formation of an intracellular vesicle (Romantschuk, Olkkonen & Bamford, 1988). This makes  $\phi 6$  unique as it uses the penetration methods similar to both enveloped (membrane fusion) and non enveloped (invagination) eukaryotic viruses (Colman & Lawrence, 2003).

The  $\phi 6$  mechanism of reintegration into the host cellular membrane, is similar to the one used by eukaryotic membrane-enveloped viruses e.g. HIV, HBV and Influenza (Colman & Lawrence, 2003). Enveloped viruses of eukaryotes enter host cells by binding to and then fusion with the host cellular membrane. In eukaryotic viruses membrane fusion is mediated by a hydrophobic  $\alpha$  helix or loop known as a fusion peptide.

In HIV, this fusion peptide is part of the gp41 viral envelope protein (Gallaher, 1987). Interaction of the receptor-binding subunit of gp120 (which is part of the gp120/gp41 complex) with the primary (CD4) and secondary (CCR5) host receptors, causes both gp120 and gp41 to undergo a major conformational change. The ultimate outcome of these changes is the exposure of the hydrophobic fusion peptide of gp41 to the lipid bilayer and insertion of the peptide into the host membrane (Brasseur et al., 1988). Upon insertion into the membrane the gp41 protein forms a trimer, with N- and C-terminal  $\alpha$  helices integrated into the viral and host membrane respectively. This is called a 3-helix intermediate. Insertion into the membrane causes the N-terminal and C-terminal  $\alpha$  helices to undergo another structural rearrangement to form a six helix bundle. This first distorts and then fuses the host cell and viral membranes, releasing the viral capsid into the host cell cytoplasm (Root, Kay & Kim, 2001; Tan et al., 1997; Weissenhorn et al., 1997).

Influenza enters cells via a similar mechanism involving the haemagglutinin protein (HA) (Wiley & Skehel, 1987). During viral entry into host cell, the virus binds to host cellular receptors and is internalized by endocytosis. As the pH of the endosomes is lowered, HA is activated (Maeda & Ohnishi, 1980). This activation involves a large structural rearrangement of HA (Bullough et al., 1994; Korte et al., 1999) exposing a hydrophobic fusion peptide which is then inserted into the host cell membrane. This eventuates in the fusion of the viral and host cell membranes.

It can be noted that the host receptor-binding proteins of viruses are generally not recycled into new virions. It is thought that this is because the conformational changes that expose the fusion peptide and later mediate membrane fusion are irreversible (Lu, Blacklow & Kim, 1995; Stiasny et al., 2002).

It is possible that filamentous phage infection is mediated by a similar mechanism to these eukaryotic viruses. The C-domain of pIII contains a hydrophobic anchor sequence, which is not cleaved from the protein during the process of assembly termination and thus could be available for reinsertion into host cell membrane during infection. This would be followed by the reintegration of the major coat protein into the host cellular membrane as the ssDNA genome is transported into the host cell (Figure 12) (Smilowitz, 1974b; Trenkner et al., 1967).

## 1.6 Aims

1. The first aim of this thesis is to understand better the involvement of the C-domain of pIII in phage infection, particularly to investigate whether the requirements for entry match those for phage release. To identify the fragment of pIII required for infection, a series of internal deletion mutants of the pIII C-domain was constructed and characterised for their ability to mediate infection. Once the minimal fragment of the C-domain that could mediate the process of infection has been defined, a mutagenic scan of the pIII C-domain was initiated and partially completed to identify residues that are important for the process of phage infection.

2. No structural information about the pIII-containing virion cap complex is available. This is mainly due to the high length to width ratio of the fl virion (1000 nm x 6 nm). Thus to make the virion cap complex more amenable to structural analyses, extremely short virions, called microphage, were produced. Given that the pre-existing system for microphage production was too inefficient to be practical, the second aim of this thesis was to design an efficient system for production and purification of the microphage. This was achieved by redesigning the original production plasmid and by developing a novel protocol for purification of the microphage.

## **Chapter 2**

### **Materials and Methods**

#### **2.1 Bacterial strains, culture conditions and chemicals**

All bacterial strains used in this thesis are listed in Table 2. The phage strains used in this thesis are listed in Table 3. All plasmids used in this thesis are listed in Table 4. All primers used in this thesis are from Invitrogen and are listed in Table 5. *E. coli* was cultivated in 2xYT medium (BD) at 37°C. The antibiotics used were Ampicillin (Amp), 100 µg/mL, Chloramphenicol (Cm), 25 µg/mL, and Kanamycin (Kan), 30 µg/mL. Liquid cultures were aerated by shaking (200 rpm) unless otherwise indicated. All agar plates were made using 1.2% Bactoagar from BD. All analytical grade chemicals were sourced from Sigma (USA), Merck KGaA (Germany), Invitrogen (USA) or UNIVAR.

#### **2.2 Molecular biology methods**

All restriction enzymes were purchased from New England Biolabs (NEB), Roche or Invitrogen. All high-purity plasmid DNA purification was carried out using either the high pure Roche plasmid kit or Qiagen Midi prep kit. All high throughput plasmid preps were carried out by using the alkaline lysis of cells, precipitation of protein using potassium acetate (3M, pH 5.5), followed by ethanol precipitation of the plasmid DNA. Gel slice purification was carried out either by the protocol outlined by centrifugation through siliconized glass wool (Heery, Gannon & Powell, 1990) followed by concentration by ethanol precipitation or by using the Roche High pure PCR purification kit, Roche Agarose gel DNA Extraction Kit or the Promega Wizard SV Gel and PCR purification kits.

For all PCR amplifications that were destined for cloning, the DNA polymerase used was either PWO (Roche) or PRIMESTAR (Takara). These are both proofreading polymerases with an up to twentyfold increased fidelity in comparison to Taq polymerase. Taq polymerase (Roche) was used for all analytical PCR reactions. All

routine molecular biology protocols were carried out either according to the Molecular Cloning Manual (Sambrook, Fritsch & Maniatis, 1989) or manufacturers' instructions.

### **2.3 Construction of plasmids**

Unless otherwise stated, the *E. coli* K12 strain TG1 was used for cloning. The exception to this is cloning into plasmids pCR Blunt II or pCR4 Blunt (Invitrogen) for which the host was *E. coli* K12 strain Top10.

Plasmids pNJB21, pNJB22, pNJB23, pNJB24 and pNJB25 are pCR Blunt II containing NdC93, NdC111, NdC121, NdC132, and NdC141 inserts, respectively (Figure 14A). They were constructed in a two-step process using ligation-mediated PCR. The N1N2 terminal domain fragment (Nd), composed of complete N1 and N2 domains, including both glycine rich linkers, and present in all members of the deletion series, was amplified using the primer pair NJB1 and NJB9 with plasmid pJARA200 used as the template. The C-domain deletion series was constructed using a single reverse primer (NJB2) and a series of forward primers: NJB4 (C141), NJB5 (C132), NJB6 (C121), NJB7 (C111), and NJB8 (C93) the template was again pJARA200. The PCR fragments encoding for the Nd fragment and the truncated C-domain series were gel-purified and phosphorylated (Optikinase, USB). Each C-domain-encoding fragment was then ligated to the amplified sequence encoding the Nd at an equal molar ratio. The ligated fragments were then PCR-amplified using the flanking primers NJB1 and NJB2. The resulting products were gel-purified and ligated into the pCR Blunt II vector (Invitrogen). Clones were then sequenced using the M13 forward and M13 reverse primers. Originally, the constructs were designed for cloning into pBAD hisB (Invitrogen) under the control of the *araP* promoter. However, all attempts to clone mutant *gIII* constructs into this plasmid were unsuccessful, due to higher residual expression, in conjunction with possible channel-forming tendency of pIII (Glaser-Wuttke, Keppner & Rasched, 1989) and already observed toxicity of C-domain fragments 141, 132 and 121 (Rakonjac, unpublished research).

**Table 2. PCR primers**

Name	Sequence	Restriction Site	5'/3' partner	Target	Note
<b>Sequencing primers</b>					
M13 Forward	GTA AAA CGA CGG CCA G	N/A	Sequencing	pCR Blunt4, pCR Blunt II	Reverse primer for sequencing TOPO cloning site
M13 Reverse	CAG GAA ACA GCT ATG AC	N/A	Sequencing	pCR Blunt4, pCR Blunt II	Forward primer for sequencing TOPO cloning site
T7	ATT AAC CCTCAC TAA AGG GA	N/A	Sequencing	pCR Blunt4, pCR Blunt II	Reverse primer for sequencing TOPO cloning site
T3	TAA TAC GAC TCA CTA TAG GG	N/A	Sequencing	pCR Blunt4, pCR Blunt II	Forward primer for sequencing TOPO cloning site
<b>Microphage primers</b>					
NB26	AGA CGT TTT CCA GTT TGG AAC AAG	N/A	NB28	pLS7	Forward primer of amplifying microphage origin out of pLS7
NB27	GTC CAA GCT TGC CCG AGA TAG	N/A			Mutagenic primer for repairing microphage origin in pCR Blunt
NB28	CCT ATA AAA ATA GGC GTA TCA CGA G	N/A	NB26	pLS7	Reverse primer of amplifying microphage origin out of pLS7
<b>NdC deletion series primers</b>					
NJB1	CCG GGA ATT CTT AAG ACT CCT TAT TAC	N/A	NJB2	pIII	Flanking forward primer for ligation mediated PCR for deletion series
NJB2	CAT GGC ATG GCT AAA AAA TTA TTA TTC G	N/A	NJB1 and NJB3	pIII	Flanking reverse primer for ligation mediated PCR for deletion series
NJB3	TTC ATA ATC AAA ATC ACC GGA ACC A	N/A	NJB1	pIII	N terminal fragment reverse primer for ligation mediated PCR deletion series

Name	Sequence	Restriction Site	5'/3' partner	Target	Note
NJB4	GCT AAT AAG GGG GCT ATG	N/A	NJB2	pIII	Forward ligation mediated primer for amplification of 141 encoding fragment
NJB5	GCC GAT GAA AAC GCG C	N/A	NJB2	pIII	Forward ligation mediated primer for amplification of 132 encoding fragment
NJB6	GGC AAA CTT GAT TCT GT		NJB2	pIII	Forward ligation mediated primer for amplification of 121 encoding fragment
NJB7	GGT GCT GCT ATC GAT GG	N/A	NJB2	pIII	Forward ligation mediated primer for amplification of 111 encoding fragment
NJB8	GGT GCT ACT GGT GAT TTT GC	N/A	NJB2	pIII	Forward ligation mediated primer for amplification of 93 encoding fragment
NJB9	CTC ATA ATC AAA ATC ACC GGA ACC A	N/A	NJB1	pIII	N-terminal fragment reverse primer. Changed final codon from TTC to CTC. Both codons are for glutamic acid.
NJB19	CCC AAG CTT TTA AGA CTC CTT ATT AC	<i>HindIII</i>	NJB20	pIII	Reverse flanking primer for cloning deletion series into pJARA200
NJB20	GCT TTC CAT TCT GGC T	<i>BamHI</i>	NJB19	pIII	Forward flanking primer for cloning deletion series into pJARA200
<b>Alanine scan series primers ligated mediated</b>					
NJB3000	AAG GCC GGA AAC GTC ACC	N/A	NJB3010	pJARA200	Reverse primer for amplifying N-terminal fragment for alanine mutants 93-97
NJB3001	GTC ACC AAT GAA ACC GTC GA	N/A	NJB3010	pJARA200	Reverse primer for amplifying N-terminal fragment for alanine mutants 98-101
NJB3002	ACC GTC GAT AGC AGC ACC G	N/A	NJB3010	pJARA200	Reverse primer for amplifying N-terminal fragment for alanine mutants 102-105
NJB3003	AGC ACC GTA ATC AGT AGC GA	N/A	NJB3010	pJARA200	Reverse primer for amplifying N-terminal fragment for alanine mutants 106-109
NJB3004	AGT AGC GAC AGA ATC AAG TTT G	N/A	NJB3010	pJARA200	Reverse primer for amplifying N-terminal fragment for alanine mutants 110-113

Name	Sequence	Restriction Site	5'/3' partner	Target	Note
NJB3005	ATC AAG TTT GCC TTT AGC GTC A	N/A	NJB3010	pJARA200	Reverse primer for amplifying N-terminal fragment for alanine mutants 114-117
NJB3006	TTT AGC GTC AGA CTG TAG CG	N/A	NJB3010	pJARA200	Reverse primer for amplifying N-terminal fragment for alanine mutants 118-121
NJB3093	GCT AAT GGT AAT GCG GCT ACT G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 93 Ala mutant
NJB3094	GCT AAT GGT GCG GGT GCT ACT	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 94 Ala mutant
NJB3095	GCT AAT GCG AAT GGT GCT ACT	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 95 Ala mutant
NJB3096	GCT GCG GGT AAT GGT GCT AC	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 96 Ala mutant
NJB3097	GGG AAT GGT AAT GGT GCT ACT	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 97 Gly mutant
NJB3098	GTT TCC GGC GCG GCT ATT G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 98 Ala mutant
NJB3099	GTT TCC GCG CTT GCT ATT GGT		NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 99 Ala mutant
NJB3100	GTT GCG GGC CTT GCT ATT GG	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 100 Ala mutant
NJB3101	GCG TCC GGC CTT GCT ATT G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 101 Ala mutant
NJB3102	TTC ATT GGT GCG GTT TCC GG	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 102 Ala mutant
NJB3103	TTC ATT GCG GAC GTT TCC GG	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 103 Ala mutant

Name	Sequence	Restriction Site	5'/3' partner	Target	Note
NJB3104	TTC GCG GGT GAC GTT TCC G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 104 Ala mutant
NJB3105	GCG ATT GGT GAC GTT TCC GG	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 105 Ala mutant
NJB3106	GCT ATC GAT GCG TTC ATT GGT	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 106 Ala mutant
NJB3107	GCT ATC GCG GGT TTC ATT GGT	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 107 Ala mutant
NJB3108	GCT GCG GAT GGT TTC ATT GGT	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 108 Ala mutant
NJB3109	GGG ATC GAT GGT TTC ATT GGT	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 109 Gly mutant
NJB3110	GAT TAC GGT GGG GCT ATC GA	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 110 Gly mutant
NJB3111	GAT TAC GCG GCT GCT ATC GA	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 111 Ala mutant
NJB3112	GAT GCG GGT GCT GCT ATC G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment of 112 Ala mutant
NJB3113	GCG TAC GGT GCT GCT ATC G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 113 Ala mutant
NJB3114	TCT GTC GCT GCG GAT TAC G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 114 Ala mutant
NJB3115	TCT GTC GGG ACT GAT TAC GG	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 115 Gly mutant
NJB3116	TCT GCG GCT ACT GAT TAC GG	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 116 Ala mutant

<b>Name</b>	<b>Sequence</b>	<b>Restriction Site</b>	<b>5'/3' partner</b>	<b>Target</b>	<b>Note</b>
NJB3117	GCG GTC GCT ACT GAT TAC G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 117 Ala mutant
NJB3118	GGC AAA CTT GCG TCT GTC G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 118 Ala mutant
NJB3119	GGC AAA GCG GAT TCT GTC G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 119 Ala mutant
NJB3120	GGC GCG CTT GAT TCT GTC G	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 120 Ala mutant
NJB3121	GCG AAA CTT GAT TCT GTC GCT	N/A	NJB3011	pJARA200	Forward mutagenic primer for amplification of the C-terminal fragment 121 Ala mutant

**Table 3. Plasmid**

Name	Backbone	Description	Resistance	Origin	Reference
pLS7	pBR322	Contains Microphage origin of replication	Amp	<i>ColE1</i> (pBR322)	(Specthrie et al., 1992)
pJARA200	pJARA200	WT pIII under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	(Rakonjac & Model, 1998)
pNJB4	pBADhisH	NdC83 under the control of the <i>ara</i> promoter	Amp	<i>ColE1</i> (pBR322)	(Bennett & Rakonjac, 2006)
pJARA24	pGZ119EH	<i>tac-pelB-C154</i>	Cm	<i>ColD</i>	(Rakonjac et al., 1999)
pJARA 34	pGZ119EH	<i>tac-pelB-C148</i>	Cm	<i>ColD</i>	(Rakonjac et al., 1999)
pJARA 44	pGZ119EH	<i>tac-pelB-C141</i>	Cm	<i>ColD</i>	(Rakonjac et al., 1999)
pJARA 54	pGZ119EH	<i>tac-pelB-C132</i>	Cm	<i>ColD</i>	(Rakonjac et al., 1999)
pJARA 64	pGZ119EH	<i>tac-pelB-C121</i>	Cm	<i>ColD</i>	(Rakonjac et al., 1999)
pJARA 74	pGZ119EH	<i>tac-pelB-C111</i>	Cm	<i>ColD</i>	(Rakonjac et al., 1999)
pJARA 84	pGZ119EH	<i>tac-pelB-C93</i>	Cm	<i>ColD</i>	(Rakonjac et al., 1999)
pJARA 94	pGZ119EH	<i>tac-pelB-C83</i>	Cm	<i>ColD</i>	(Rakonjac et al., 1999)
pJARA140	pBR322	WTpIII under the control of the <i>psp</i> promoter	Amp	<i>ColE1</i> (pBR322)	JR, unpublished

Name	Backbone	Description	Resistance	Origin	Reference
pJARA220	pGZ119EH	<i>pspP</i>	Cm	<i>ColD</i>	JR, unpublished
pC3S	pBR322	Phagemid Vector	Cm	<i>ColE1</i> (pBR322)	JR, unpublished
pNJB20	pCR Blunt II	Contains NdC141	Amp	<i>ColE1</i> (pBR322)	This study
pNJB21	pCR Blunt II	Contains NdC132	Amp	<i>ColE1</i> (pBR322)	This study
pNJB22	pCR Blunt II	Contains NdC121	Amp	<i>ColE1</i> (pBR322)	This study
pNJB23	pCR Blunt II	Contains NdC111	Amp	<i>ColE1</i> (pBR322)	This study
pNJB24	pCR Blunt II	Contains NdC93	Amp	<i>ColE1</i> (pBR322)	This study
pNJB30	pJARA200	pIII NdC141 under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB31	pJARA200	pIII NdC132 under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB32	pJARA200	pIII NdC121 under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB33	pJARA200	pIII NdC111 under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB34	pJARA200	pIII NdC93 under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB35	pJARA200	pIII NdC83 under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study

Name	Backbone	Description	Resistance	Origin	Reference
pNJB50	pGZ119EH	pIII C-domain under the control of the <i>psp</i> promoter	Cm	<i>ColD</i>	This study
pNJB51	pGZ119EH	Phagemid vector derived from pJARA220	Cm	<i>ColD</i>	This study
pYW01	pGZ119EH	Phagemid expressing the C-domain of pIII under the control of <i>psp</i> promoter	Cm	<i>ColD</i>	(Jankovic, 2008)
pNJB093	pJARA200	pIII G332A under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB097	pJARA200	pIII A328G under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB106	pJARA200	pIII G319A under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB107	pJARA200	pIII D318A under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB108	pJARA200	pIII I317A under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB109	pJARA200	pIII A316G under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB110	pJARA200	pIII A315G under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB111	pJARA200	pIII G314A under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB115	pJARA200	pIII A310G under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB116	pJARA200	pIII V309A under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB117	pJARA200	pIII S308A under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study

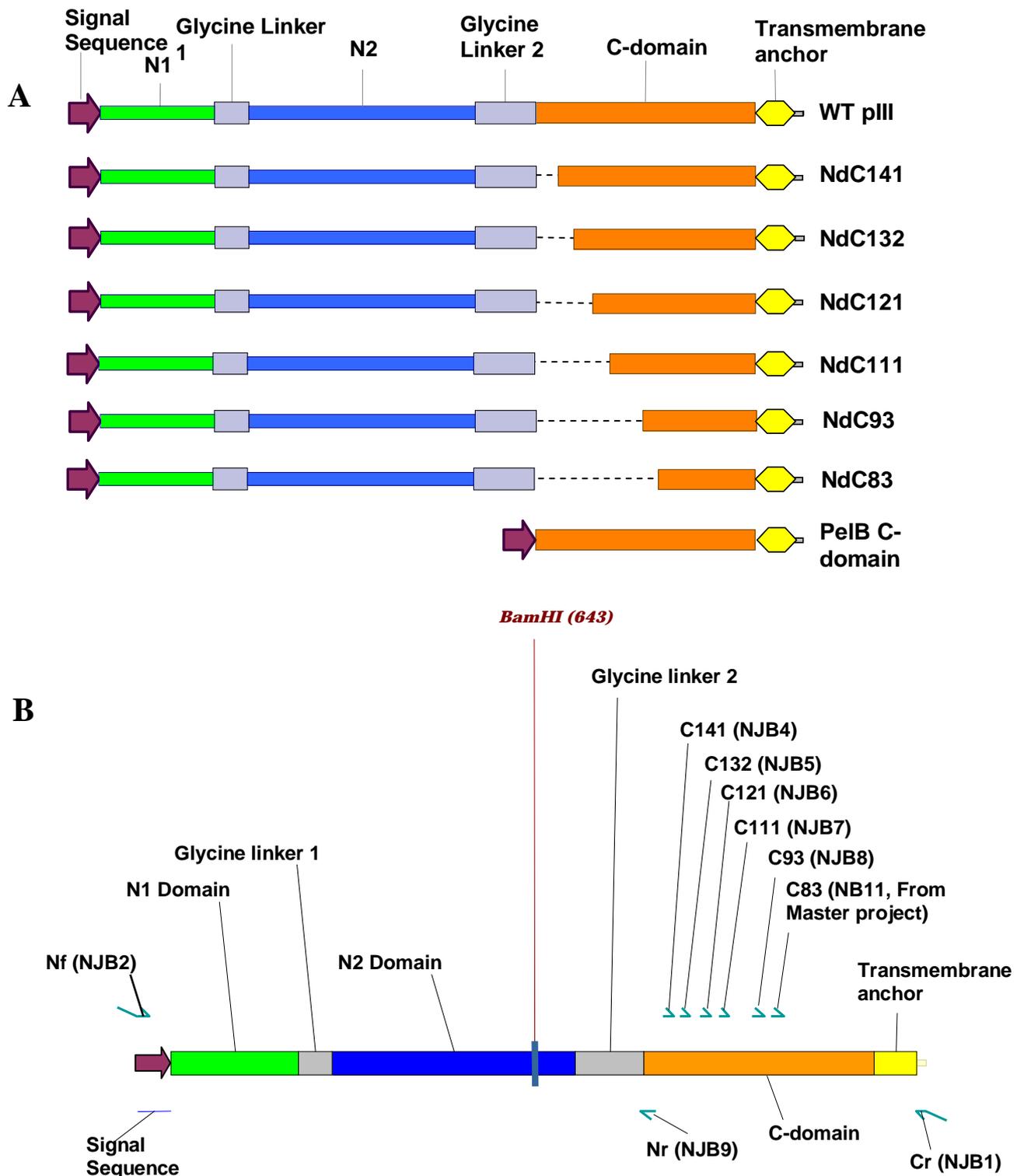
<b>Name</b>	<b>Backbone</b>	<b>Description</b>	<b>Resistance</b>	<b>Origin</b>	<b>Reference</b>
pNJB118	pJARA200	pIII D307A under the control of the <i>lacUV5</i> promoter	Amp	<i>ColE1</i> (pBR322)	This study
pNJB7	pCR4 Blunt	Contains repaired microphage origin of replication from pLS7	Kan	<i>ColE1</i> (pUC)	(Bennett, 2005)
pCR Blunt II	pCR Blunt II		Kan	<i>ColE1</i> (pUC)	Invitrogen
pCR4 Blunt	pCR4 Blunt		Kan, Amp	<i>ColE1</i> (pUC)	Invitrogen

**Table 4. Phage strains**

<b>Name</b>	<b>Genotype</b>	<b>Reference</b>
f1	Wild-type	(Loeb, 1960)
f1d3	f1, $\Delta gIII$	(Rakonjac, Jovanovic & Model, 1997)
R408	f1, DPS, IX, T30A, IR1, gtrxA2	(Russel, Kidd & Kelley, 1986)
R408d3	f1, DPS, IX, T30A, IR1, gtrxA2 $\Delta gIII$	(Rakonjac et al., 1997)
VCSM13	M13, PA15 <i>ori</i> , <i>Kan<sup>R</sup></i> , IR	Stratagene
VCSM13d3	$\Delta gIII$	(Rakonjac et al., 1997)
R676	f1, <i>gVIII</i> E25am $\Delta IV$	(Linderoth, Model & Russel, 1996)
R777	f1, DPS, IX, T30A, IR1, gtrxA2 <i>gVIII</i> E25am	This Thesis

**Table 5. Bacterial strains list**

<b>Strain</b>	<b>Genotype</b>	<b>Source</b>
K1030	<i>S26 RIE, fadL701 phoM510 mcrB rrnB ton A22, gar B10, ompF, relA1, pit 10, spoT1, T2R supD zed508::Tn10</i>	Rockefeller strain collection
K561	<i>HfrC, λ+, relA1, spoT1 T2R, OmpF627, fadL701, supD, lacI<sup>q</sup></i>	Rockefeller strain collection
TG1	<i>Δ(lac-pro), supE44, thi, hsdR 5, F' traD36, proA+B+, lacI<sup>q</sup>, lacZDM14</i>	(Carter, Bedouelle & Winter, 1985)
Top10	<i>F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araΔ139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG</i>	Invitrogen
K1976	TG1//pJARA112	(Bennett & Rakonjac, 2006)
K1981	TG1//pYW01	(Jankovic, 2008)
K1994	TG1//pNJB50	This Study
K2177	TG1//pNJB51	This Study



**Figure 14. Diagram of the mutant pIII proteins used in NdC experiments and positions of primers** **A.** Schematic of proteins produced by NdC deletion series constructs and C-domain showing relative sizes of constructs compared to the wild-type pIII. **B.** Schematic showing positions of primers used in construction of NdC deletion series mutants.

Instead of pBAD vector, an existing plasmid containing wild-type *ori* under the tightly controlled *lacUV5* promoter was used as a destination vector. The *lacUV5* promoter is a mutated *lac* promoter that has a very low residual of expression in the absence of induction. The mutations in the *lacUV5* have been shown to make the promoter less sensitive to intracellular cAMP levels (Wanner, Kodaira & Neidhardt, 1977). Because cAMP participates in the de-repression of the *lac* promoter, the *lacUV5* mutations result in a lower basal transcription level through the promoter. Concurring with this, the WT pIII controlled by the *lacUV5* promoter on pJARA200 has been shown to have a very low background in the absence of IPTG (Davis, 1985; Rakonjac & Model, 1998).

The *ori* encoded for on pJARA200 contains a naturally occurring *Bam*HI site within the Nd domain. This was used to transfer the series of internal deletion mutants into pJARA200. The fragment containing NdC internal deletion mutant was amplified by PCR using primer NJB20 located upstream of the *Bam*HI site and reverse primer NJB19 which inserted a *Hind*III site downstream of the C-terminus of pIII allowing insertion into *Bam*HI-*Hind*III digested pJARA200 to replace the corresponding portion of the WT *ori* with the deletion mutant. Plasmids pNJB30, pNJB31, pNJB32, pNJB33 and pNJB34 were constructed by PCR amplification of the fragments encoding NdC141, NdC132, NdC121, NdC111, NdC93 using flanking primers NJB20 and NJB19 and the corresponding pCR Blunt II clones as the template. Plasmid pNJB35 was constructed by PCR amplification of NdC83-encoding fragment using the primers NJB20 and NJB19 primer and the araC-NdC83 plasmid pNJB4 as a template (Bennett & Rakonjac, 2006). The fragments were digested with *Bam*HI and *Hind*III, gel-purified and cloned into *Bam*HI-*Hind*III digested pJARA200. All resulting plasmids were isolated and sequenced using primers NJB19 and NJB20 to ensure accuracy. All constructs rendered the host cells resistant to filamentous phage infection when pIII production was induced by the addition of IPTG to the culture. This confirmed that pIII was being produced by these constructs.

Even though the *lacUV5* promoter has a very low expression of pIII in the absence of IPTG, host cells expressing NdC mutants from pJARA200-based plasmids could only be infected at rate between 40%-90%, depending on the construct.

The complete C-domain of pIII was provided from either a compatible plasmid or phagemid depending on whether the particles required for analysis were to be phage or phagemids, respectively. Plasmid pNJB50 is a derivative of plasmid pJARA24 which expresses the C-domain of pIII under the control of the *tac* promoter and is derived from the *ColD ori* vector pGZ119EH (Lessl et al., 1992). To replace the *tac* promoter with the *psp* promoter, pJARA24 was combined with pJARA220 (Rakonjac, unpublished research), a plasmid with *psp* promoter in the same (pGZ119EH) vector backbone. pJARA24 was cut with *Xba*I and *Hind*III, and the small fragment containing the *ori* coding sequence was gel purified and ligated into *Xba*I-*Hind*III cut pJARA220. The phage shock protein promoter or *pspP*, is a tightly regulated promoter, which is induced by phage infection or periplasmic stress (Brissette et al., 1990; Brissette et al., 1991; Model, Jovanovic & Dworkin, 1997; Weiner, 1993; Weiner, Brissette & Model, 1991). The construct was sequenced to ensure accuracy. The construct was also tested for its ability to complement the assembly in the  $\Delta$ *gIII* phage f1d3, confirming that the C-domain is expressed.

Phagemid pNJB51 is derived from pJARA220 (Rakonjac, unpublished research). pNJB51 contains the *psp* promoter followed by a multiple cloning site (MCS), f1 *ori*, *cat* gene ( $\text{Cm}^{\text{R}}$ ) and *colD* plasmid origin. To construct pNJB51, the f1 origin of replication and 5' portion of the *cat* gene from a  $\text{Cm}^{\text{R}}$  phagemid pC3S (Rakonjac, unpublished research) was inserted between *Hind*III site, at the distal end of the MCS, and *Nco*I site of the *cat* gene. Plasmid pC3S was cut with *Hind*III and *Nco*I. The cut fragment was gel-purified and ligated into *Hind*III-*Nco*I cut pJARA220. Colonies showing insertions were tested for their ability to produce phagemid derived-transducing particles by infecting them with helper phage R408.

The alanine scan series of plasmids pNJB93 (G332A), pNJB97 (A329G), pNJB106 (G319A), pNJB107 (D318A), pNJB108 (I317A), pNJB109 (A316G), pNJB110 (A315G), pNJB111 (G314A), pNJB115 (A311G), pNJB116 (V310A), pNJB117 (S309A), pNJB118 (D308A) are all derivatives of pJARA200 and contain the indicated mutations in the C-domain of WTgIII. The plasmid number indicates the position of mutated residue counted from the C-terminus of pIII. The number and letters in the brackets indicate the amino acid mutation and position from the N terminus of mature pIII. Plasmids were constructed in a two-step process using ligation-mediated PCR.

Plasmid pJARA200 was used as a template in all amplifications. Position of primers is indicated in Figure 15.

In the first round of PCR for construction of plasmids pNJB093 and pNJB097, the common N-terminal fragment, N93-97, was constructed using NJB20 flanking forward primer and NJB3000 reverse primer. The C-terminal fragments were amplified using forward mutagenic primers NJB3093 (C93) or NJB3097 (C97) and NJB19 as the reverse flanking primer.

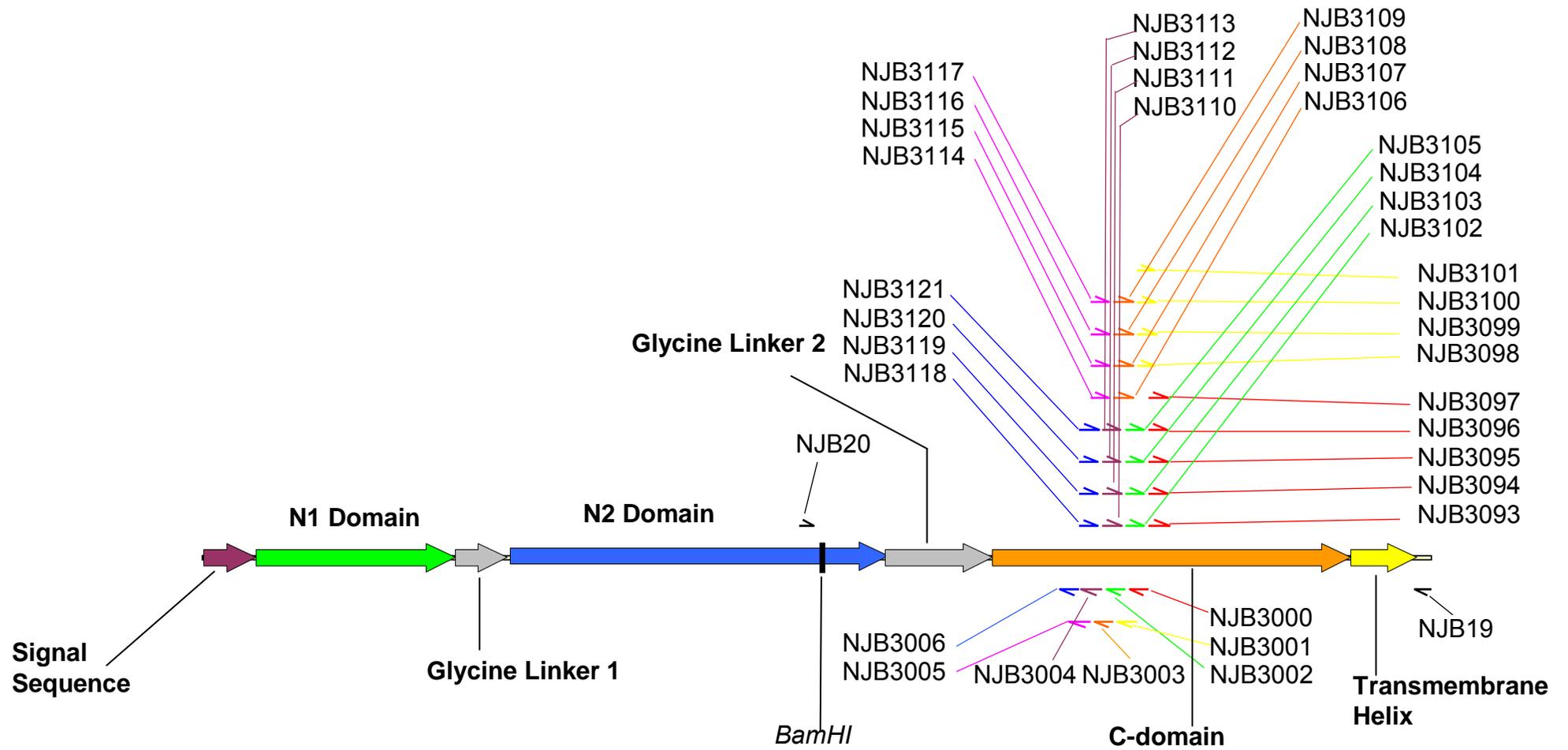
In the first round of PCR for construction of plasmids pNJB106, pNJB107, pNJB108 and pNJB109, the common N-terminal fragment, N106-109 was constructed using forward flanking primer NJB20 and the reverse primer NJB3003. The C-terminal fragments were amplified using forward mutagenic primers NJB3106 (C106), NJB3107 (C107), NJB3108 (C108), or NJB3109 (C109) to amplify the respective mutants; NJB19 was the reverse flanking primer.

In the first round of PCR for construction of plasmids pNJB110 and pNJB111, the common N-terminal fragment N111-114 was amplified using the forward flanking primer NJB20 and reverse primer NJB3004. The C-terminal fragments were amplified using forward mutagenic primers NJB3110 (C110), or NJB3111 (C111) to amplify the respective mutants with NJB19; was the reverse flanking primer.

For plasmids pNJB115, pNJB116 and pNJB117, the common N-terminal fragment N114-117 was amplified using the forward flanking primers NJB20 and reverse primer NJB3005. The C-terminal fragments were amplified using forward mutagenic primers NJB3115 (C115), NJB3116 (C116), and NJB3117 (C117) to amplify the respective mutants; NJB19 was the reverse flanking primer.

For plasmid pNJB118, the N-terminal fragment N118-121 was amplified using forward primer NJB20 and reverse primer NJB3006. The C-terminal fragments was amplified using the forward mutagenic primer NJB3118 (C118) and NJB19 was the reverse flanking primer.

First round PCR products of the N-terminal and C-terminal fragments were gel-purified and phosphorylated. The N-terminal fragments were ligated to the corresponding C-terminal fragments. The appropriate ligation products were amplified then by PCR using the flanking primer pair NJB20 and NJB19. The resulting products were gel-purified and ligated into pCR Blunt II vector (Invitrogen) using the corresponding cloning kit. Inserts were sequenced using M13 forward and M13 reverse primers. Correct clones were then excised from pCR Blunt II recombinants using BamHI and *Hind*III and inserted into *Bam*HI-*Hind*III digested pNJB35 encoding the NdC83 deletion mutant, to replace the fragment containing a 213 nucleotide internal deletion within *gIII*. This strategy of using a shorter-insert plasmid as a starting vector allowed easy identification by size of full-length inserts containing alanine scan point mutations. Recombinant plasmids containing the inserts of expected size were then sequenced again to ensure accuracy.



**Figure 15. Relative positions of the Alanine scan primers.** (For accurate position, sequences and lengths of all *gIII* primers see Table 2 and Appendix 2). Flanking primers for all amplifications were NJB19 and NJB20. Internal primers were designed in sets, indicated by colour. Each set consists of one N-terminal reverse primer for amplifying the N-terminal fragments. The set of a few mutagenic primers with identical 5' ends were ligated to a single N-terminal fragment, and then the ligation products were amplified in the second round of PCR.

Helper phage R777 (R408 *gVIII<sup>am</sup>*) was constructed from helper phage R408 and phage R676 ( $\Delta gIV$ , *gVIII<sup>am</sup>*) to combine the origin of replication of R408 and *gVIII<sup>am</sup>* of R676. R408 and R676 were cut with *Bam*HI and *Bsp*GI. The purified large *Bam*HI/*Bsp*GI fragment of R408 containing *gVI*, *gI*, *gXI*, *gIV*, *gII*, an intergenic sequence containing an interference resistant origin of replication and most of *gV* was ligated to the purified small *Bam*HI/*Bsp*GI fragment of R676 containing *gVII*, *gIX*, *gVIII* and of the 5' moiety of *gIII*. The ligation was transformed into electrocompetent cells of *supD* strain K1030 and immediately plated on a lawn of K1030 to identify plaques of individual recombinant phage. To identify R408 *gVIII<sup>am</sup>* phage, plaques were passaged on the suppressor strain K1030 and a non-suppressor strain K561. Recombinant phage clones that formed plaques on K1030 but not on K561 were R408 *gIII<sup>am</sup>* mutants named R777.

## **2.4 Phage protocols**

### **2.4.1 Phage stocks**

All working stocks of phage were prepared by the following methods. An exponentially growing culture (around  $10^8$  cells/mL) of an appropriate bacterial strain was infected with phage at an m.o.i = 50. The phage culture was then further incubated for four hours (37°C/200 rpm) to allow for phage growth. Cells were then removed by centrifugation (6000 g/15 min). To remove any remaining viable bacterial cells from the supernatant it was either heated, to kill any remaining cells. (65°C/20 min) and centrifuged again (6000 g/15 min) to remove the cellular debris or filtered through a 0.2  $\mu$ m filter. Phage stocks were titrated as described in Section 2.4.6 and stored at 4°C.

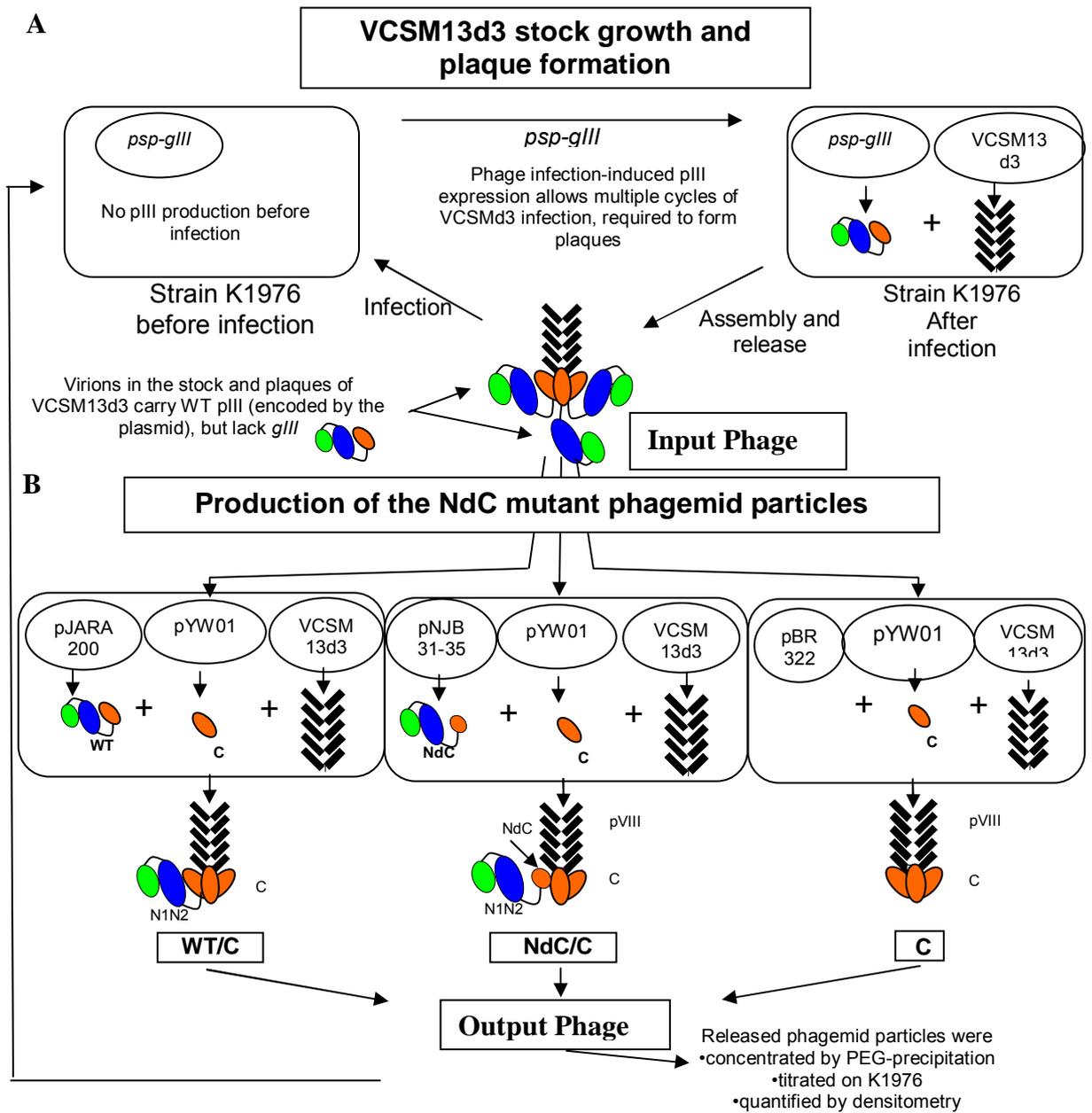
## 2.4.2 Growth experiments for producing phage or phagemid particles (PPs) containing combinations of pIII mutants

The host *E.coli* strain used for producing mutant virions is TG1. This strain was chosen because it is *lacI<sup>q</sup>*, and *F<sup>+</sup>*. This means that TG1 allows the tight control of the *lac* promoter used to control the expression of various constructs and can be infected with the helper phage for the production of virions. Strain K1981 is a derivative of TG1 containing the phagemid pYW01 which expresses the C-domain of pIII under the control of the *psp* promoter (Jankovic, 2008). K2177 is a derivative of TG1 and contains phagemid pNJB51. K1981 and K2177 were used for production of phagemid particles. K1994 is a derivative of TG1 and contains plasmid pNJB50 which expresses the C-domain of pIII under the control of the *psp* promoter. K1994 was used in production experiments to produce phage particles only.

For the NdC deletion mutant growth experiments, all NdC deletion mutant plasmids were transformed into strain K1981 (phagemid particles) or K1994 (phage particles) just prior to the growth experiment. For the alanine scan mutant series growth experiments, all alanine scan mutant plasmids were transformed into K2177 just prior to the growth experiment. These strains were not stored long-term because of concerns that pYW01/pNJB50/pNJB51 and NdC deletion/alanine scan plasmids will undergo recombination causing reconstitution of WT *gIII*. Corresponding bacterial strains are referred to by the particular NdC internal deletion or alanine scan mutation plasmid that they contain. An outline of the growth experiment used to produce phagemid samples in the NdC and alanine scan mutant experiment is described in Figures 16 and 17, respectively.

Strains containing plasmids expressing NdC or alanine scan mutants and compatible phagemids/plasmid (pYW01, pNJB50 or pNJB51) were propagated in 2xYT media supplemented with Amp (100 µg/mL) and Cm (25 µg/mL). Exponentially growing cultures (OD<sub>600</sub>= 0.2) were infected with VCSM13d3 phage (for phagemid experiments) or f1d3 (for phage experiments) at a multiplicity of infection of 100 phage per cell, for 1 hour (without shaking). For the NdC deletion series, infected cells were then separated from unabsorbed phage by centrifugation (5000g /10 min/37 °C) and

resuspended in 2xYT supplemented with Amp (100 $\mu$ g/mL) Cm (25 $\mu$ g/mL) and IPTG (0.1mM). The cultures were then incubated for a further 4 hours to allow phage or phagemid particle production. For the alanine scan series, the centrifugation step was omitted and instead IPTG (0.1mM) was supplemented into the culture after infection to induce expression of pIII mutant proteins. The culture was then incubated for a further period of 4 hours. Before the four-hour incubation began, a small sample was taken from the culture and plated to determine the efficiency of infection of the producer cells. Following an incubation period of 4 hours, the culture was centrifuged (8000 g/20 min/30°C) to remove the producer cells. The supernatant containing the released phage or phagemid particles was collected for further analysis.



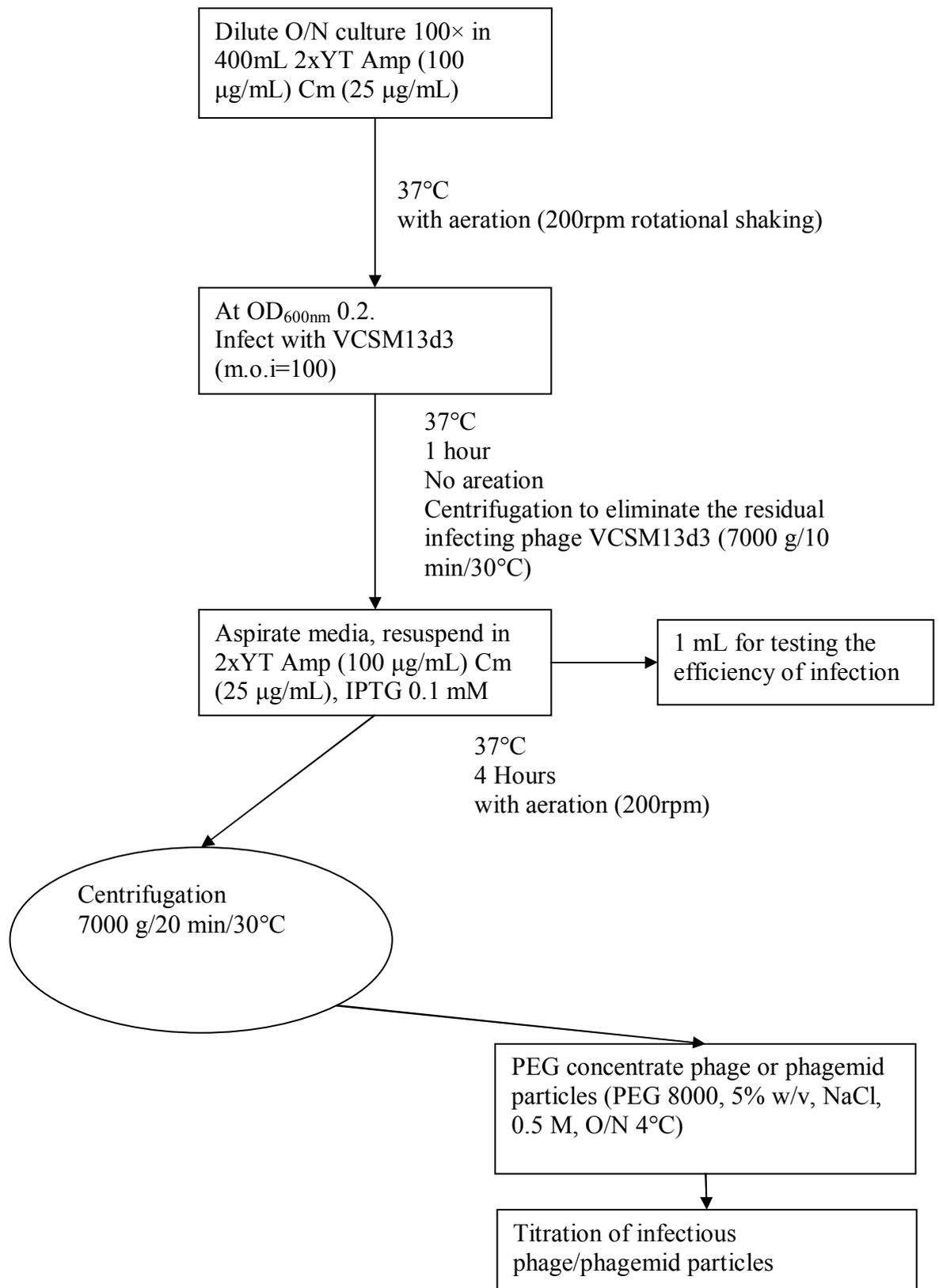
**Figure 16. Method for producing mutant phagemid particles.** **A.** Growth of the VCSM13d3 helper phage stock containing wild-type pIII on the surface but carrying genome with complete deletion of gene III ( $\Delta gIII$ ). **B.** Production of phagemid particles containing the NdC mutants. pJARA200, WT *gIII* under the control of a *lacUV5* promoter; pYW01, phagemid expressing the complete C-domain under the control of the *psp* promoter. pNJB31-35, NdC internal deletion series mutants expressed from a *lacUV5* promoter. pBR322, cloning vector from which pJARA200 and pNJB131-35 were derived. Phagemid particles are virion particles that contain an ssDNA genome derived from a phagemid. A phagemid is a plasmid which carries both a plasmid and a filamentous phage origin of replication. Thus, upon infection with a filamentous helper phage it replicates via the phage origin of replication to form ssDNA genome, which is then packaged into virions to produce phagemid particles (PP's).

### 2.4.3 Producer cell infection test

All NdC deletion and alanine scan series producer strains contain a multi-copy plasmid that expresses *gIII* constructs containing the N1N2 domains. Because of this, these strains were expected to produce a small amount of pIII in the absence of IPTG induction. Expression of the N1 and/or N2 domains of pIII renders the cells resistant to Ff infection. If N1 and/or N2 expression occurs before infection, the level of infection of the producer cell culture is low. This results in a drastic reduction in the yield of produced output phage. Therefore, for virion production experiments, it is important to check the percentage of producer cells that were infected with the helper phage.

To test the producer cell infection level with helper phage VCSM13 or VCSM13d3 (which carry a  $Km^R$  marker), a small sample of phage-infected culture was taken one hour after infection and titered on 2xYT plates and 2xYT supplemented with Kan (30 $\mu$ g/mL) plates. Phage infected cells become  $Km^R$ , whereas non-infected cells remain  $Km^S$ . Thus, the fraction of phage infected cells in the producer culture was determined as the ratio of the titres of  $Km^R$  to  $Km^S$  colony-forming units and expressed as a percentage of infected cells, relative to the total cell count.

In all phage growth experiments between 40-90% of cells were infected. This was deemed an acceptable level of infection to produce a high enough yield of phage/phagemid particles for analysis. In situations where less than 40% of the producer cells were infected, the phage/phagemid yield was not sufficient. In this thesis, I cloned *gIII* into several vectors until finding that only pJARA200-derived constructs consistently allowed infection of more than 40% of the host cells.



**Figure 17. Flow diagram of the production of the virions carrying combination of pIII constructs.**

#### **2.4.4 Concentration of virions by PEG precipitation**

Cultured phage supernatants were concentrated by overnight precipitation in PEG8000 (5 % w/v) and NaCl (0.5 M) at 4°C. Precipitate was collected by centrifugation (16 500 g/45 min/4 °C). The pellet was then resuspended in 5mL of DNase I buffer (10 mM Tris, 2.5 mM MgCl<sub>2</sub>, 0.1 mM CaCl<sub>2</sub>, pH 7.5) and centrifuged again (16 500g/10min/4°C) to remove any precipitated protein. Concentrated phage stocks were then treated with DNase (1 µg/mL) and RNase A (1 µg/mL) for 1 h at room temperature on a rotator. Phage were precipitated again in PEG8000 (5% w/v) and NaCl (0.5 M) at 4°C for 2 h. Precipitate was collected by centrifugation (16500 g/45 min/4 °C). The pellet was then resuspended in an appropriate amount of either TE buffer (10mM Tris, 2 mM EDTA, pH 7.6), or if they were going to be further purified using the Sarkosyl and Triton X-100 as outlined in the next section (2.4.5), the pellet was resuspended in TBS (50 mM Tris, 150 mM NaCl, pH7.5) and centrifuged to remove any precipitated debris (16 9500 g/10 min)

#### **2.4.5 Purification of concentrated phage using Sarkosyl and Triton X-100**

If concentrated phage were to be analysed by western blots or transmission electron microscopy then the sample was further purified to remove as much cellular debris as possible. This helped to reduce the background of non-specific cell-derived bands in western blots. Only phage that are stable in detergents (Sarkosyl and Triton X100) can be subjected to this procedure (Smith, 2005; Wickner, 1975).

Concentrated phage from the previous procedure were resuspended in 10 mL of TBS at room temperature and centrifuged to remove any precipitated debris (16 500g/10 min). The supernatant was transferred into a fresh tube and Triton X-100 solution (10% v/v) was added to 1%. The solution was then gently shaken at room temperature for 1 hour. Phage were then precipitated in PEG 8000 (5% w/v) and NaCl (0.5 M) at room temperature for 2 h. The precipitate was collected by centrifugation (16 500 g/45 min) and the phage pellet was resuspended in 10 mL of TBS (50 mM Tris, 150 mM NaCl, pH7.5). Phage suspension was centrifuged (16 500 g/10 min) to remove any debris. Supernatant was then transferred into a fresh tube and to this 10 mL of Sarkosyl

solution (2 % w/v). The solution was then gently shaken at room temperature for 1 h. Phage were precipitated in PEG8000 (5 % w/v) and NaCl (0.5 M) at room temperature for 2 hours. Precipitate was collected by centrifugation (16 500 g/45 min) and the phage pellet was then resuspended in an appropriate volume of TBS. The concentrated phage suspension was centrifuged to remove any precipitated debris and the supernatant was transferred to a fresh tube.

#### **2.4.6 Titration of infectious phage and phagemid particles**

Phage fl (wild-type), R408 and VCSM13 were titrated on TG1 or K561. Phage fl d3, R408d3 and VCSM13d3 were titrated on K1976 (TG1 containing pJARA112). The *gVIII<sup>am</sup>* phage R777 was titrated on strain K1030 (a *supD* strain, see table 5).

Phage were first titrated by a quick method. 10µl drops of phage dilutions were placed on the surface of a lawn of the appropriate *E. coli* strain in a soft 2xYT agar on 2xYT agar plates. After O/N incubation at 37°C, plaques in the area of the absorbed drops were counted to determine the titre. Once a rough titre was determined using this quick method, the titre was determined more accurately by plating in triplicate at a dilution that gave ~300 plaques per plate. An average (mean) from three plates was determined as the infectious titre for a sample.

The infectious phagemid-encapsulating particles (PPs) were titrated through their ability to transfer or transduce an antibiotic resistance marker present on the phagemid to a recipient strain by infection. Phagemid titration was carried out by using a protocol similar to the one outlined in Russel et al. (1988). For titring of phagemid pYW01 (Cm<sup>R</sup>), agar plates containing 21 mL of 2xYT supplemented with Chloramphenicol (25 µg/mL) were overlaid prior to use with 9 mL 2xYT agar (1.2 %) without antibiotic. This antibiotic-free layer served to delay the contact of recipient bacteria with chloramphenicol, allowing the in situ expression of the chloramphenicol acetyltransferase to occur after transduction of the antibiotic resistance into the host cell. Only the cells that were infected with PP (phagemid particles) produced a colony on the antibiotic selective plates after overnight incubation. The quick “drop” method was

used first followed by the precise titration that involved counting of ~300 colonies per plate and 3 plates per sample as described above for the phage titration.

#### **2.4.7 Agarose gel electrophoresis of the phage and quantification**

The f1 wild-type virions normally contain one genome. However, under the circumstances when elongation of the virion filament is favoured over termination, a range of longer virions can be released from the infected cells. All virion samples in this thesis contained a similar range of virion sizes (e.g. Figure 21); hence, the ratio of the number of virions in all samples to the number of virion encapsulated genomes should be equivalent. Given no major differences in the sizes of virions among the samples, genome equivalent was used as a measure of the number of phage or phagemid particles in a sample.

A genome equivalent is a measure of particle mass, and is defined as a particle (or its portion) containing one encapsulated genome (Figure 18). Thus a virion particle containing ten genomes represents ten genome equivalents, as do ten virion particles containing one genome each (Rakonjac & Model, 1998). In all experiments, the number of genome equivalents was determined from agarose gel electrophoresis of phage or phagemid ssDNA, released from SDS disassembled virions (Nelson et al., 1981).

Prior to electrophoresis, virions were disassembled by incubation in SDS-containing buffer (1 % SDS, 1× TAE Buffer (40 mM Tris Acetate pH 8.3, 1 mM EDTA), 5 % Glycerol, 0.25 % BPB) at 70 °C for 20 min. After electrophoresis, phage ssDNA was stained with either Ethidium bromide or SYBR gold (Invitrogen) and quantified densitometrically. Since the amount of ssDNA in a band is not linearly proportional to the intensity of the fluorescence, every gel contained a set of twofold dilutions of a standard used for calibration. For Ethidium bromide-stained gels the standard range was 1280 ng to 40 ng per lane, whereas for SYBR gold the standard range was typically 1000 ng to 0.976 ng per lane. The ssDNA standard used was purified phage ssDNA from f1d3. The concentration of the ssDNA standard was determined by UV spectrometry. The gels for analysis by densitometry were photographed using a CCD camera (BIORAD), under conditions below pixel saturation. Quantitative analysis was

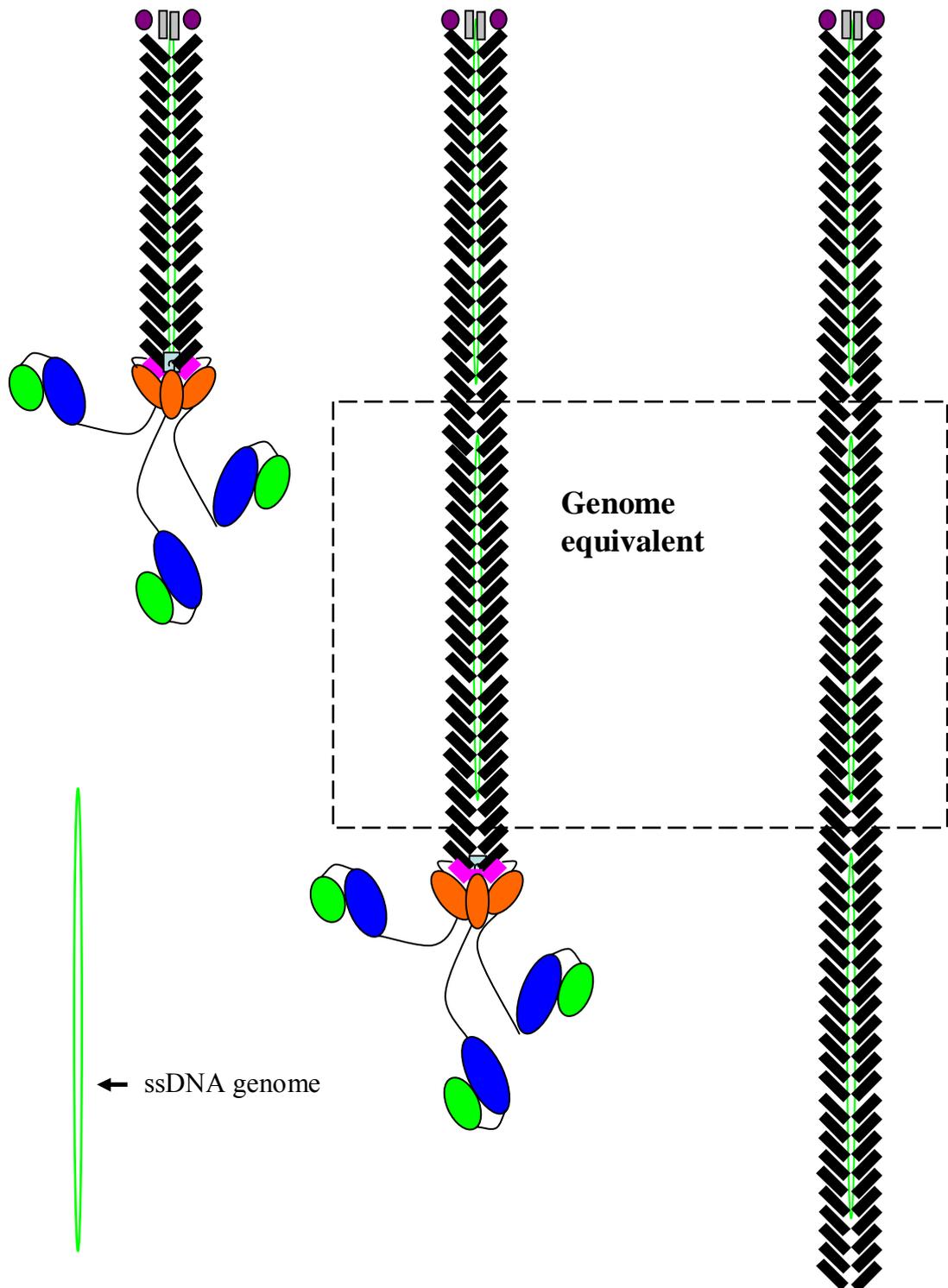
performed using software packages Image Gauge (Fuji Film), Quantity One (Biorad), and Excel (Microsoft). Each gel photograph was analysed by densitometry three times to ensure accurate results and a standard curve was fitted over the standard data range. This standard curve was then used to calculate the mass of ssDNA in each band. Conversions of the calculated amount of ssDNA (ng) in the sample into the amount of genome equivalents was carried out based on the molecular mass of a particular ssDNA genome, which was again calculated from the base composition and length (Appendix 1)

To minimize the error in densitometric analysis of the gel bands, the following steps were carried out: All band measurements in an analysis were taken using the same surface area. The background density was removed from each band by measuring the density of an equal area directly below the band. This should limit the error caused by uneven background fluorescence and band spreading. All samples and standards were prepared and loaded in the same volume. Gels were photographed and processed at the maximal possible resolution, to ensure that the highest amount of pixels possible was available for analysis.

Native virion agarose gel electrophoresis was used to separate virions of various lengths and to detect free ssDNA when the stability or size distribution of virions was analysed (Nelson et al., 1981). Samples were loaded onto low density agarose gels (0.6 %) in DNA loading buffer (1×TAE, 5 % Glycerol, 0.25 % BPB). When stability of the phage was analysed, Sarkosyl (N-lauryl sarcosine, 0.1 %) was added to the loading buffer and the sample was incubated at room temperature for ten minutes prior to loading. The gel was electrophoresed at 3 V/cm for sixteen hours. After electrophoresis, free phage ssDNA was detected by staining the gel in Ethidium Bromide. To detect the position of the virions in the gel, they were then disassembled by soaking the gel in an alkaline solution (0.2 M NaOH) for 1 h, followed by neutralising the gel in 0.45 M Tris pH 7.1, and then by staining the gel again with Ethidium Bromide. For the Southern blots analyses, DNA from SDS-disassembled or native virion electrophoresis was transferred to nitrocellulose membrane and probed using appropriate probes. Southern blot detection was carried out using the ECL direct Nucleic Acid Labelling and Detection System (GE healthcare Bioscience) according to manufacture's protocols.

## Monophage

## Polyphage



**Figure 18. A diagram describing the concept of genome equivalent.** A monophage is a virion that contains only one ssDNA genome. A polyphage is a virion particle that carries multiple copies of ssDNA genome. A genome equivalent is a portion of a phage or phagemid particle that contains one ssDNA genome (Rakonjac & Model, 1998).

#### **2.4.8 Protein electrophoresis and western blots**

Proteins from the phage samples were separated by SDS-PAGE, using either glycine (12%, pIII) (Laemmli, 1970) or tricine (16%, pVI) (Schägger & von Jagow, 1987) gel systems. Proteins were then transferred to nitrocellulose filters and detected using appropriate antibodies. Affinity purified polyclonal antibody R164 (Rakonjac et al., 1997) was raised in rabbits against the C-terminal decapeptide of pIII (FANILRNKES). Therefore it is able to detect all C-terminal fragments of pIII (Rakonjac et al., 1999). Antibody 19-38 was raised in rabbits against the C-terminal peptide of pVI (Endemann & Model, 1995). Antibodies were detected using an anti-rabbit antiserum conjugated to Alkaline phosphatase and detected using substrates Nitro blue tetrazolium (NBT) and 5-Bromo-4-chloro-3-indolyl phosphate (BCIP) in alkaline buffer (Blake et al., 1984). The base buffer used in western blots was TBS (30 mM Tris, 150 mM NaCl, pH 8.0, 0.05 % Tween 20). Blocking and antibody binding buffers also contained either 5% non-fat milk powder or 0.5 % I-Block (Applied Biosystems).

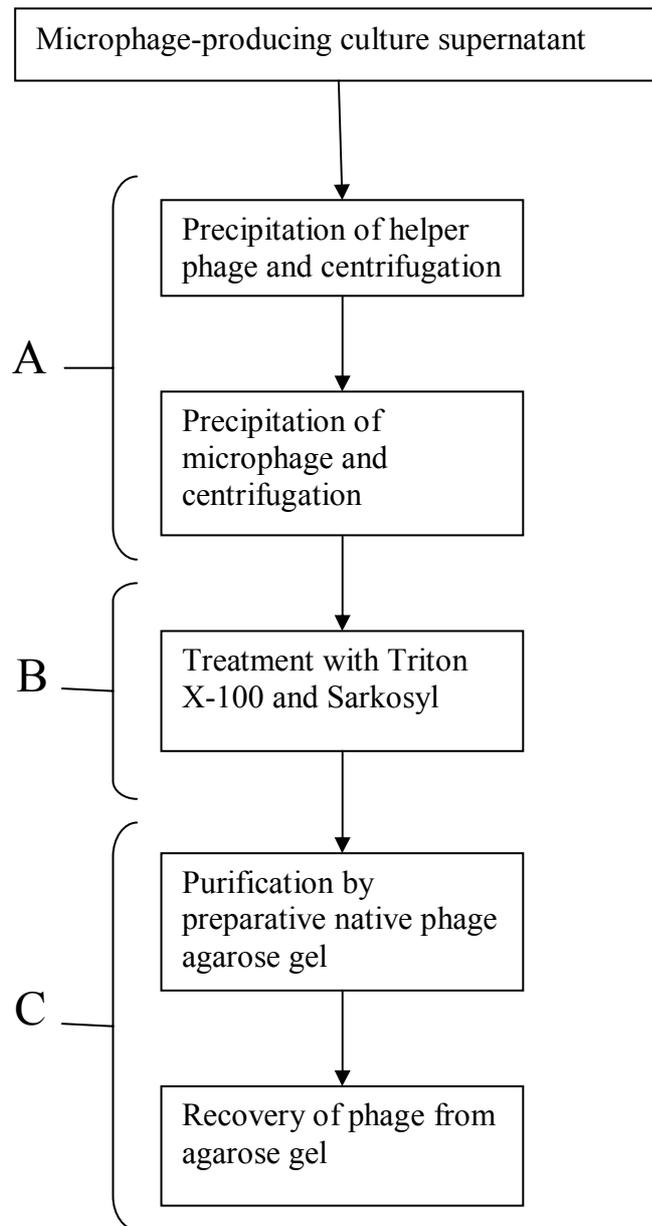
#### **2.4.9 Growth, concentration and purification of microphage**

Microphage are very short Ff filamentous phage, (50 nm in length, as opposed to the WT fl that is 860 nm). They are produced by infecting an *E. coli* strain carrying the microphage-producing plasmid pNJB7 with helper phage R777. Microphage producing plasmid pNJB7 is a better plasmid for producing microphage than the previous microphage plasmid pLS7. Firstly, pNJB7 contains a corrected microphage producing origin sequence; the microphage producing origin in pLS7 contained a mutation that severely reduces microphage genomes production. Secondly, pNJB7 is a high copy number plasmid (ColE1, pUC ori) whereas pLS7 is a low copy plasmid (ColE1, pBR322 ori). The increased copy number increases the number of pNJB7 plasmids available for pII-mediated rolling circle replication upon infection with a helper phage. In turn, this increases the yield of microphage genomes available for packaging, as well as the yield of microphage virions produced. For microphage growth experiments, pNJB7 was transformed into *E. coli* strain K1030 prior to growth experiments. Microphage produced were then separated from the R777 helper phage by a protocol developed in this thesis. The protocol combined a two-step PEG precipitation,

Sarkosyl/Triton X-100 purification and purification using preparative native phage agarose gel electrophoresis.

In detail, an exponentially growing culture of K1030 containing microphage-producing plasmid pNJB7 in 2xYT supplemented with Kanamycin (30 $\mu$ g/mL) was infected with helper phage R777 at an m.o.i of 100 phage per cell. Culture was then incubated for 1 hour (without shaking) at 37°C to allow infection with helper phage to occur. After infection, the culture was incubated for a further four hours at 37°C with aeration (200rpm). Producer cells were then separated from the released phage particles by centrifugation (7000 g/20 min/4°C). The microphage-containing supernatant was then subjected to further purification.

Long helper phage were precipitated from the supernatant overnight in “low PEG” (2.5 % w/v PEG8000) and 0.5 M NaCl at 4°C. The helper phage precipitate was collected by centrifugation (16 500 g/ 45 min/4°C) and resuspended in an appropriate volume of TBS (50 mM Tris, 150 mM NaCl, pH 7.5) for further analysis. PEG was added to the supernatant to a final concentration of 15% w/v and further incubated overnight at 4°C. The microphage-rich precipitate was collected by centrifugation (16 500g/ 45 min/ 4°C). The pellet was resuspended in 10 mL TBS and subjected to Sarkosyl/Triton X-100 purification as outlined in section 2.4.5. This step removed any lipids or cell envelope-embedded proteins that may have been precipitated with the microphage. This allowed the concentration of the microphage to a higher degree than would have been possible without this treatment. The flow-chart diagram of microphage purification is shown in Figure 19.



**Figure 19. Optimised microphage purification protocol.** **A.** Two step PEG-precipitation; **B.** Sarkosyl and Triton X-100 treatment; **C.** Preparative native phage gel electrophoresis, followed by microphage recovery from the gel slice.

Differential precipitation of the microphage does not eliminate all helper phage virions. Helper phage virions represent the major virion fraction in the supernatant of a microphage-producing culture (Chapter 5, Section 5.4, Figure 31). To eliminate the remaining helper phage virions after Triton X-100/ Sarkosyl purification, the microphage were further purified from any contaminating helper phage by preparative native phage agarose gel electrophoresis followed by recovery from agarose.

A Southern blot, using the microphage origin of replication as a probe, conducted on a native phage agarose trial gel, was used to locate the microphage band relative to a 1kb<sup>+</sup> dsDNA ladder (In this case the 1kb<sup>+</sup> ladder was used as a sign-post rather than as a size standard). On a 1.5% TAE agarose gel after a three-hour electrophoresis at 3 V/cm the native microphage were detected in the region of the gel corresponding to the space between the 1.3 kb and 1.6 kb bands of the 1kb<sup>+</sup> dsDNA ladder (Invitrogen).

To purify microphage, a preparative native phage gel was run loaded with microphage and a 1kb<sup>+</sup> dsDNA ladder. After electrophoresis, the preparative gel was stained with Ethidium Bromide for 1 h. This stained the 1Kb<sup>+</sup> ladder which was then visualised using UV transillumination. The area within the microphage-loaded lanes corresponding to the portion of the gel at the level between 1.3 kb to 2.0 kb bands was excised from the gel. Phage were isolated from the gel by centrifugation (4500 g) of the agarose slice through a small filter made of siliconized glass wool until no more fluid could be recovered from the gel. The remains of the gel fragment were then soaked in TBS and subjected to centrifugation again to remove all microphage from the gel slice. All fluid collected from the gel slice was centrifuged (7000 g/10 min) to remove any further remains of agar. The microphage were then precipitated from the fluid extracted from the gel slice by overnight precipitation in “high PEG” (PEG8000 15% w/v, 0.5 M NaCl). Precipitated microphage were then collected by centrifugation (16 500 g/45 min/4 °C) and resuspended in an appropriate volume of TE. The concentration of microphage was determined by SDS-disassociated phage agarose electrophoresis followed by densitometric analysis as described in section 2.4.7.

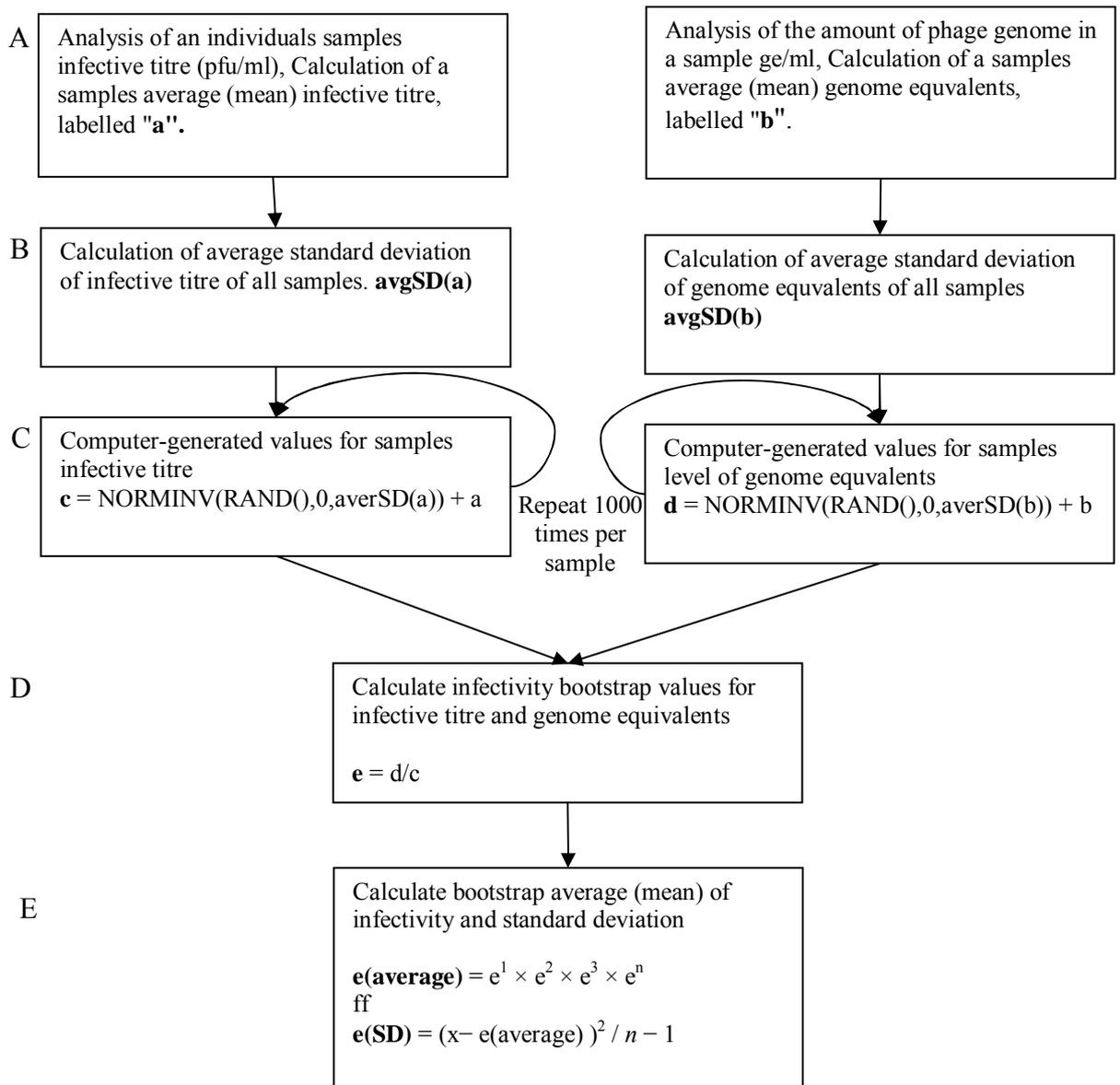
The purity of the concentrated microphage was assessed by native phage agarose electrophoresis and transmission electron microscopy (TEM) of the negatively stained samples.

## 2.5 Bootstrap Analysis

To statistically predict the error of the infectivity of the alanine scanning mutants bootstrapping analysis was used (Efron, 1979; Efron, 1981). Bootstrapping is a computer-based data resampling method that can be used to determine standard deviation of a ratio. For bootstrap analysis the following assumptions were made: Since all alanine mutants and WTPIII/Cd control had infective titres (tdp/mL) within one order of magnitude, and the titres were measured using the same procedures, it was assumed that the error between samples should be due to a common error derived from the experimental procedure used in analysis. Thus, the average standard deviation of infective titre could be used in the bootstrap calculations. Likewise, the same assumption was made of the calculation of error for genome equivalents. Because all measurements of genome equivalents are within one order of magnitude and the same procedure was used to measure all samples the average standard deviation of genome equivalents was used in calculations.

Bootstrap analysis was carried out using the following procedure: bootstrap values were generated for both infective titre and genome equivalents using the inverse normal distribution function (NORMINV) within Excel (Microsoft). The probability values were seeded by a random number generated using the random number function (RAND) within Excel. The average standard deviation of either the infective titre or genome equivalents was used as the standard deviation, depending on which value was being generated. The mean value used in all calculations was zero. The values generated by the inverse normal function were then added to either the average infective titre or genome equivalent of each sample, depending on value being calculated.

The process of creating a bootstrap value for both infectivity and genome equivalents was reiterated one thousand times for each sample. Bootstrap-generated numbers were then combined to calculate an array of bootstrap values for the infectivity for each sample. Bootstrap-generated infectivities that were less than zero or higher than one were removed from standard deviation calculations. Computer generated infectivities were then used to calculate the standard deviation of each alanine scan sample.



**Figure 20. Flow diagram of the bootstrap analysis used to predict the standard deviation of infectivity.** **A.** Calculation of the average (mean) infective titre (a) or genome equivalents (b) in a sample. **B.** Calculation of the average standard deviation of both infective titre and genome equivalents measurements. Because both measurements gave values within one order of magnitude and had a normal distribution the average standard deviation of all measurements of a value were used to calculate the average standard deviation of either infectivity (avgSD(a)) or genome equivalents (avgSD(b)). **C.** Computer-aided generation of bootstrap values for infective titre (c) and genome equivalents (d). Values were generated by using the inverse normal distribution function (NORMINV) within the Microsoft Excel program. The probability factor was provided by a computer-generated random value (RAND). The mean used for the distribution was zero and the standard deviation was used was either the average standard deviation of the infective titre or genome equivalent. The mean of the infective titre measurement or genome equivalent measurement for that sample was then added to the generated value, producing a computer-generated value for infective titre of the number of genome equivalents. This process was repeated one thousand times per sample for each value. **D.** Calculation of a bootstrap value of infectivity (e). Computer-generated bootstrap values for infective titre (c) and genome equivalents (d) were used. All computer generated values for infectivity less than zero were removed from further infectivity calculations. Values produced were tested to ensure that they had a normal distribution **E.** Calculation of average (mean) bootstrap values for infectivity (e(average)) and standard deviation (e(SD)). For bootstrap analysis a statistically significant difference was defined as an infectivity that varied over two standard deviations from the mean of the WTpIII positive control.

## Chapter 3

### Mapping of the pIII C-domain for the minimal fragment required for phage infection

#### 3.1 Introduction

Bennett and Rakonjac (2006) showed that a termination incompetent C-terminal fragment of the C-domain (83 residues) fused to the receptor binding domains N1 and N2 cannot mediate phage entry. It has also been shown that a longer C-terminal fragment of 93 residues is sufficient for termination of phage assembly the release from the membrane (Rakonjac et al., 1999). However, the minimal C-domain requirement for infection has not been determined.

The aim of this study is to investigate whether the C-domain requirement for phage entry is the same as that for phage release. By comparing the minimal fragment of the C-domain required for phage entry and for phage release, it will be possible to determine if the two processes have similar or different mechanisms.

#### 3.2 Experimental system for the production of composite virions

To determine what was the minimum requirement for the C-domain of pIII for infection, a series of C-domain nested deletions were fused to intact receptor-binding N-terminal domains (N1N2) and incorporated into phagemid-containing virions. This was achieved by using a plasmid-based expression system to produce the pIII mutant proteins in conjunction with a helper phage that carries a complete deletion of *gIII* and a phagemid expressing the intact C-domain, to assemble pIII mutant-containing virions. The ability of these virions to infect *E. coli* was then examined. To ensure that pIII mutant virions were assembled correctly, the physical properties of derived virions and presence of mutant pIII was analyzed.

The pIII mutants were named to reflect the presence of the N-terminal domains (Nd) and the size of the remaining C-terminal fragment: NdC141, NdC132, NdC121, NdC111, NdC93 and NdC83 (Figure 14). These particular sizes of fragments were chosen to match the deletion series used to investigate the minimal C-domain fragment required for phage assembly (Rakonjac et al., 1999). The mutants were constructed using ligation-mediated PCR as described in Material and Methods (Chapter 2) and were placed under the control of the *lacUV5* promoter of plasmid pJARA200 (Rakonjac & Model, 1998). The resulting plasmids were named pNJB30, pNJB31, pNJB32, pNJB33, pNJB34 and pNJB35, expressing NdC141, NdC132, NdC121, NdC111, NdC93 and NdC83 respectively.

Given that the NdC83 mutant is not sufficient for the termination of phage assembly (Bennett & Rakonjac, 2006) and that mutants NdC93-121 assemble defective virions (Rakonjac et al., 1999), a complete C-domain of pIII mutant was expressed *in trans* in the same host cell from pYW01, a compatible phagemid derived from vector pGZ119EH (Lessl et al., 1992). The expression of the C-domain in pYW01 was controlled by the *psp* (phage shock protein) promoter (Jankovic, 2008). This mutant (Cd) does not carry the receptor-binding domains N1 and N2, and thus cannot mediate phage infection. Thus, expression of complete C-domain ensures the release of stable virions, yet it does not contribute to ability of those virions to infect host cell (Bennett & Rakonjac, 2006).

Plasmid pYW01 that expresses the C-domain (Cd) is also a phagemid. Phagemids are plasmids that carry both a plasmid and a filamentous phage origin of replication. The *fl* origin of replication enables the phagemid, if the host cell is infected with a filamentous helper phage, to be packaged into virions protein coats, thus forming phagemid particles (PPs). In this experiment, the phagemid was used to distinguish the output mutant phage from the helper phage stock used to infect the producer culture (input helper phage), given that an antibiotic marker ( $Cm^R$ ) is carried by the phagemid particles, but not by the input helper phage virions.

The mosaic phagemid particles (PPs) carrying a combination of an NdC series mutant and Cd will be referred to as NdC/Cd through this thesis. These PPs were generated by infection of cells carrying an NdC-expressing plasmid and a Cd-expressing phagemid

with a helper phage that contains a deletion of entire *gIII* (VCSM13d3). Hence, the assembled phagemid particles do not contain wild type pIII and their infectivity solely relies on NdC mutants expressed in the infected cell. This allows testing of the ability of the NdC mutants to mediate infection. The schematic representation of the virion particles produced in this thesis is shown in Figure 16 and amino acid sequence and schematics of the NdC deletion mutants are shown in Appendices 2 and 3.

The positive control for this experiment are phagemid particles containing wild-type pIII and the Cd (labelled WTpIII/Cd) produced from cells expressing full-length wild type pIII under the control of the *lacUV5* promoter (pJARA200) and Cd under the control of the *psp* promoter (pYW01). WT pIII/Cd phagemid particles are infectious and provide a reference for comparison to NdC/Cd-carrying virions.

The negative control sample for this experiment is composed of phagemid particles carrying only the complete C-domain of pIII (labelled Cd through this thesis). These mutant phagemid particles were produced from host cells expressing the intact C-domain without N1N2 from pYW01. Phagemid particles produced by this system terminate assembly efficiently, but are non-infectious due to the lack of the N1N2 domains (Figure 16).

### **3.3 The complete C-domain complements the assembly deficiency of the NdC83 mutation and stabilizes the NdC93-121 virions.**

The length of Ff virions is the most significant indicator of pIII-mediated release from the membrane. If pIII is not functional, the virion filaments continue to elongate, and extremely long particles, carrying more than 20 genomes are formed (Crissman & Smith, 1984; Rakonjac & Model, 1998). These filaments are initially attached to the host cell; those found in the supernatant represent filaments broken off by mechanical shearing. When subjected to native virion electrophoresis, the long filaments appear as a slow-migrating smear (Nelson et al., 1981). Phage virions whose release from the cell is mediated by pIII typically run as a ladder of bands on a native phage gel, representing particles that are carrying between one and ten phage genomes. Phage carrying more than one genome are referred to as “polyphage” (Figure 18). In a wild-type fl culture

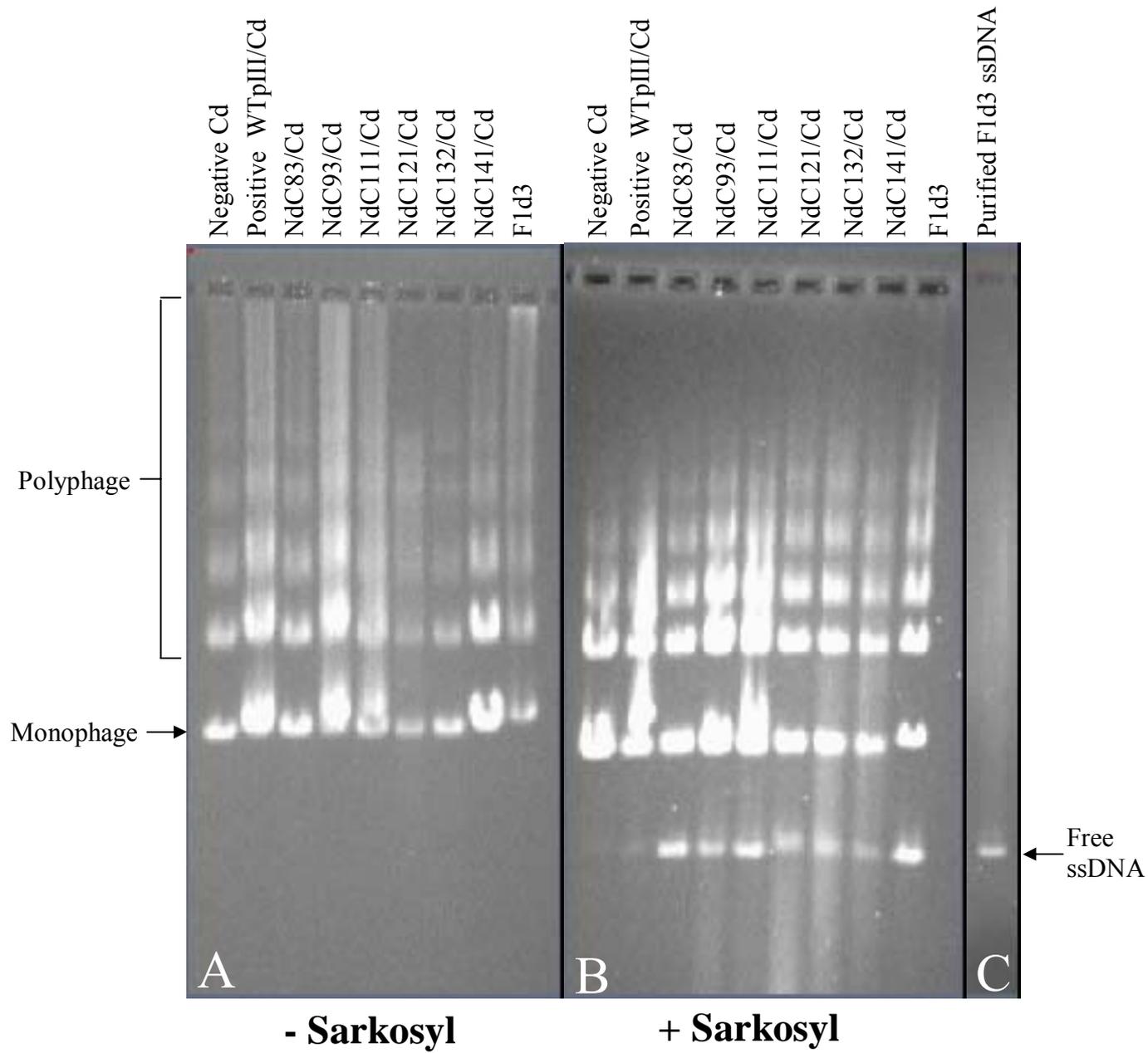
>90% of virion particles are produced as monophage. However, in  $\Delta gIII$  phage complemented by pIII expressed from a plasmid, the amount of polyphage is elevated and the fraction of monophage is less than 50%.

The native virion gel electrophoresis was used to separate virions according to their length (Figure 21). The NdC/Cd virions analysed in the presented native virion electrophoresis gel photograph were obtained using a phage system rather than a phagemid system. In this phage system, Cd was expressed from a plasmid (pNJB50) rather than a phagemid and thus only the virions containing phage genomes were produced. The phage system was used for this analysis because it produced a much higher titre of the virion particles compared to the phagemid system. This makes the native virion gels much easier to interpret.

The native virion electrophoresis showed that all NdC/Cd phage contained similar distributions of monophage and short polyphage relative to the WTpIII/Cd control (Figure 21). Phagemid particle samples produced from NdC and WTpIII-expressing cells using phagemid pYW01 to supply Cd and VCSM13d3 as the helper phage were also subjected to this analysis and were all found to consist of monophage and short polyphage containing various combinations of helper phage and phagemid genomes.

In Rakonjac et al. (1999) it was shown that the C-domain fragment 93, 111 and 121 residues in length were sufficient to terminate phage assembly, but the resulting virions were unstable in the ionic detergent Sarkosyl (N-lauryl sarcosine). The virions constructed in this thesis contain the complete Cd in addition to the NdC mutants with truncated C-domains. Truncated domains within the NdC mutant could also possibly destabilise the virions even though these virions contained the complete Cd. To ascertain that this was not the case, the mutant carrying virions were tested for stability in the detergent Sarkosyl. If the phage termination complex was not correctly formed then virion particles would have been degraded in Sarkosyl. Degradation would eliminate a ladder of bands formed by phage virion particles but instead would give a single band in the same position as a purified ssDNA genome (Bennett & Rakonjac, 2006). All NdC/Cd containing mutant virions were stable in Sarkosyl; hence, a combination of destabilising and stabilising mutants resulted in stable virions. It is however noted that all NdC/Cd containing mutant virions, WTpIII/Cd and Cd did not

have a band of ssDNA after the Sarkosyl treatment. Therefore, these samples contained very little, if any broken off virions without the pIII/pVI cap.



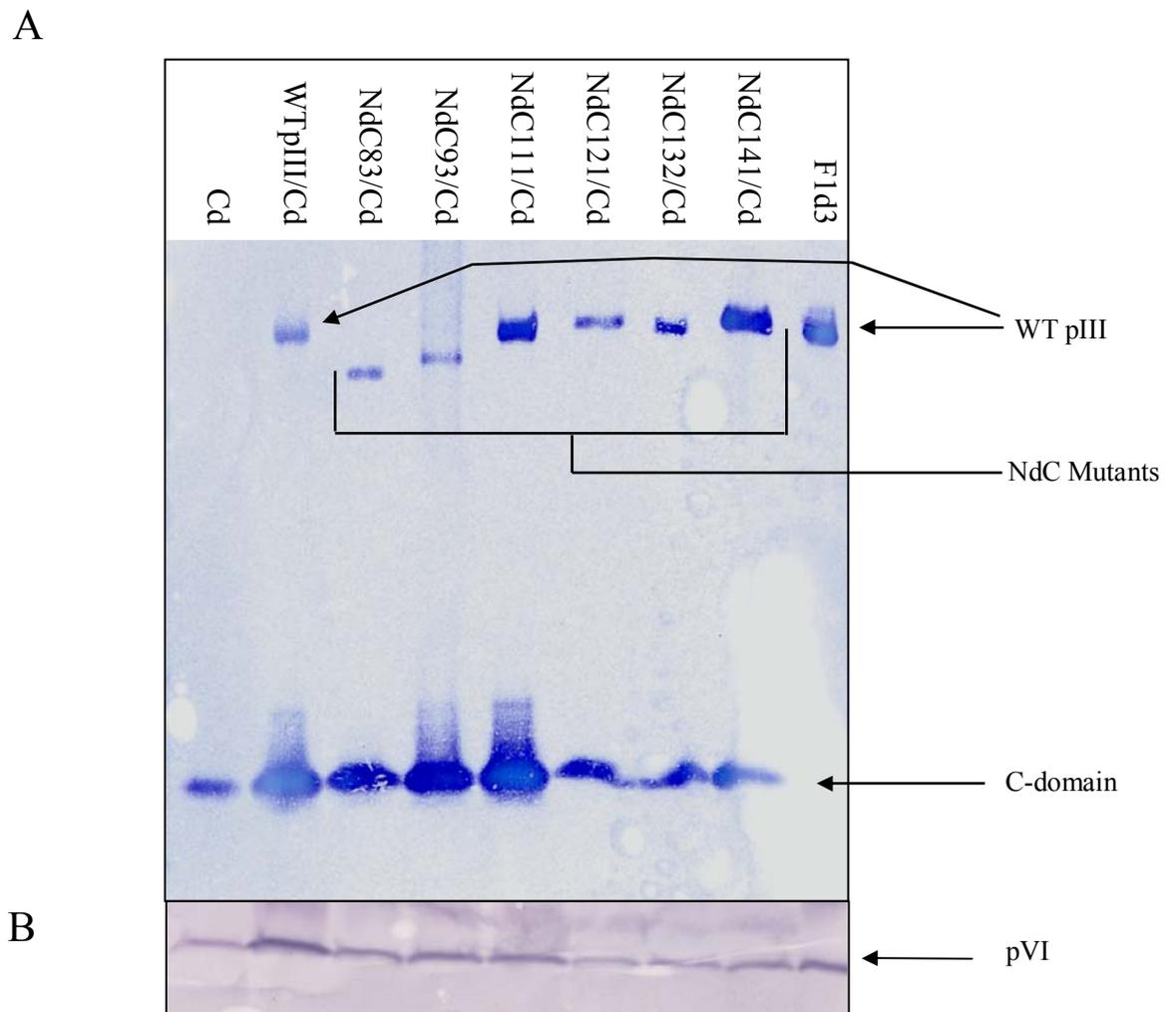
**Figure 21. Native phage gel electrophoresis.** **A.** Native electrophoresis of phage samples in the absence of the ionic detergent Sarkosyl. Virions of different lengths form a ladder of bands, indicating that termination of phage assembly is occurring. Polyphage are virions that contain two or more genomes. Monophage virions contain a single genome (See Figure 18). If phage assembly termination were not occurring properly there would be only a smear just below the wells. **B.** Native phage electrophoresis of virions treated with the ionic detergent Sarkosyl. **C.** Purified f1d3 ssDNA. If the phage termination complex was not correctly formed then all virion particles would have been degraded in Sarkosyl and there would be no ladder of bands formed by phage virion particles (Bennett & Rakonjac, 2006). A small amount of ssDNA is released during incubation with Sarkosyl from a fraction of mechanically sheared incomplete polyphage filaments in the sample.

### **3.4 The NdC mutants are incorporated into the virions**

For the investigation of the role of the C-domain in infectivity, it was important to demonstrate that the NdC deletion mutant protein fragments are incorporated into the virions efficiently. This would indicate if the low infectivity of a deletion mutant was due to degradation of the mutant protein within the host cell periplasm before assembly into phage virions, or due to poor incorporation of mutants into the virions, rather than due to the effect of the internal deletion of the C-domain of pIII in infection.

The incorporation of NdC mutant pIII proteins into the virion particles was monitored by western blotting using an anti-pIII antibody (R164) (Rakonjac et al., 1997). This antibody is raised against the decapeptide FANILRNKES that corresponds to 10 C-terminal residues of pIII. Therefore, it recognises all pIII mutants used in the experiments (Figure 14A). All NdC mutant proteins were detected in the virions confirming that they were incorporated into the virions (Figure 22).

Low infectivity could also be caused by the poor incorporation of pVI into the virions. pVI together with pIII forms the phage cap complex. To examine this possibility, the presence of pVI was determined by western blotting using anti-pVI antibodies (Endemann & Model, 1995). pVI was detected in all NdC mutant series virions, and therefore it was incorporated efficiently into the phagemid particles carrying the NdC mutant pIII proteins. Therefore, the phagemid particles contained correctly assembled pIII-pVI termination cap complex (Figure 22).



**Figure 22. Western blots of virions A.** Detection of pIII using anti-pIII antibodies raised against C-terminal decapeptide of pIII. Bands corresponding to WTpIII, NdC mutant series and Cd are indicated. **B.** Detection of pVI using anti pVI antiserum.

### 3.5 Analysis of the NdC mutant virion infectivity

To measure the ability of the NdC mutants to infect the host cells, the capacity of NdC/Cd phagemid particles to mediate phage infection was measured (Table 5). The infectivity is defined as the number of infectious phagemid particles per total amount of phagemid particles in a sample. Therefore, infectivity is calculated as a ratio of the infectious phagemid titre to the number of phagemid particles in a sample (expressed as genome equivalents, as defined in Material and Methods, Section 2.4.7).

The infectious phagemid particles were titrated by transduction of  $Cm^R$  marker of phagemid pYW01 into the indicator strain TG1. Phagemid particle titration showed that the infective titres of the NdC mutants NdC83/Cd and NdC93/Cd were approximately four orders of magnitude lower than the titre of the positive control (WTpIII/Cd) phagemid particles and were very similar to that of the negative control (Cd). NdC111 particles had a titre two orders of magnitude higher than the titre of the negative control. NdC121, NdC132 and NdC141 had titres that were very close to or within the same order of magnitude as that of the positive control (Table 5).

In the phage system analyzed by native gel electrophoresis (Figure 21), a *gIII* deletion mutant of the f1 phage (f1d3) was used, and a compatible plasmid (not a phagemid) was a source of Cd. This system gave about hundredfold higher titre relative to the system that produced phagemid particles, but also had a relatively high background of infectious virions carried over from the f1d3 stock used to infect the producer cells. The infectious titre of the f1d3 phage (or the number of plaque-forming units per ml) was determined by titration of samples on the corresponding complementing strain K1976 (Table 6).

A low background of *gIII*<sup>+</sup> phage arises by low-frequency recombination with the complementing plasmid used to generate the stocks (Rakonjac et al., 1997). To monitor the level of *gIII*<sup>+</sup> phage of VCSM13d3 in the phagemid system and f1d3 in the phage system, the phage were titrated on strain TG1 that does not provide pIII (data not shown). The *gIII*<sup>+</sup> titres were, as expected, about six orders of magnitude lower than those of the  $\Delta gIII$  phage.

The total amount of phagemid particles was determined by quantification of encapsulated phagemid ssDNA. Phagemid ssDNA was released from virions by disassociation of the virion capsids in an SDS-containing buffer, followed by separation by agarose gel electrophoresis. The amount of phagemid ssDNA in bands was determined by densitometry and converted to number of phagemid ssDNA genomes or genome equivalents as described in the Material and Methods (Section 2.4.7). Quantification of the f1d3 particles in the phage system was carried out using the same method.

The results showed that all NdC mutant samples contained similar amounts of phagemid particles (Table 5). Therefore, phagemid assembly was of similar efficiency in all samples, as ssDNA derived from the disassociated virions was present at a similar concentration within all samples.

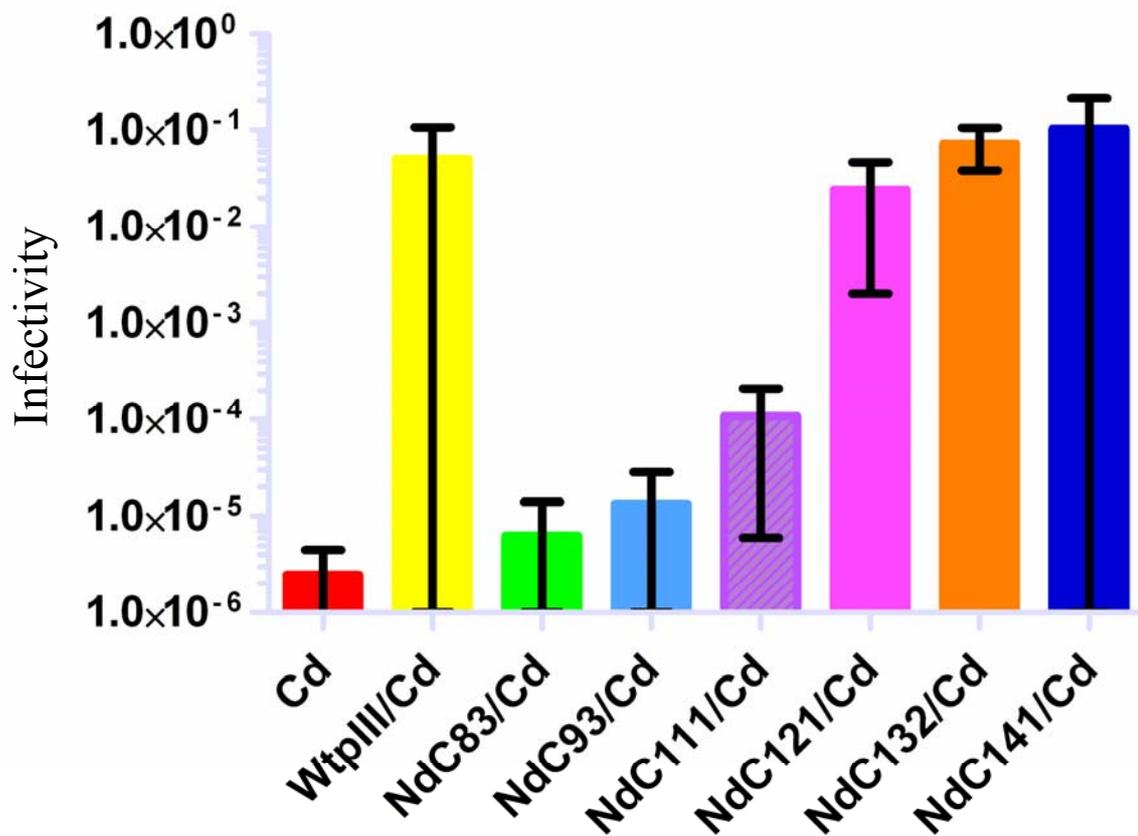
**Table 6. Infectivity of the NdC deletion mutant phagemid particles**

Phagemid	Infectious titre (tdp/mL) <sup>a</sup>	Quantity of virions (ge/mL) <sup>b</sup>	Infectivity (tdp/ge) <sup>c</sup>
Cd	1.17±0.66×10 <sup>6</sup>	4.69±1.16×10 <sup>11</sup>	2.48±2.01×10 <sup>-6</sup>
WTpIII/Cd	1.73±1.54×10 <sup>10</sup>	3.42±0.75×10 <sup>11</sup>	5.07±5.61×10 <sup>-2</sup>
NdC83/Cd	4.87±2.80×10 <sup>6</sup>	7.75±5.17×10 <sup>11</sup>	6.28±7.79×10 <sup>-6</sup>
NdC93/Cd	9.33±6.66×10 <sup>6</sup>	6.96±2.84×10 <sup>11</sup>	1.34±1.50×10 <sup>-5</sup>
NdC111/Cd	1.32±0.47×10 <sup>8</sup>	1.22±0.71×10 <sup>12</sup>	1.08±1.02×10 <sup>-4</sup>
NdC121/Cd	1.51±0.85×10 <sup>10</sup>	6.20±2.19×10 <sup>11</sup>	2.44±2.24×10 <sup>-2</sup>
NdC132/Cd	4.15±0.15×10 <sup>9</sup>	5.76±2.44×10 <sup>11</sup>	7.20±3.31×10 <sup>-2</sup>
NdC141/Cd	2.75±1.97×10 <sup>10</sup>	2.66±0.91×10 <sup>11</sup>	1.03±1.09×10 <sup>-1</sup>

<sup>a</sup> Titre; the number of infectious phagemid particles that can transduce Cm<sup>R</sup> into the indicator strain (tdp) per mL

<sup>b</sup> Quantity of virions is the number of genome equivalents (ge) per mL. Genome equivalents are defined in Material and Methods (Section 2.4.7; Figure 18).

<sup>c</sup> The error of infectivity was calculated by the following equation  $e[y] = (e[a] + y \times e[b])/b$ , where a = infectious titre, b = quantity of virions, and y = infectivity.



**Figure 23. Infectivity of phagemid particles containing the NdC mutants.**

Infectivity is a ratio of the infectious phagemid particles titre (transducing particles per mL) to the amount of phagemid particles (in genome equivalents per mL), expressed as a percentage value. The error of infectivity was calculated by the following equation  $e[y] = (e[a] + y \times e[b])/b$ , where  $a$  = infectious titre,  $b$  = quantity of virions, and  $y$  = infectivity. Error bars represent one standard deviation from the mean.

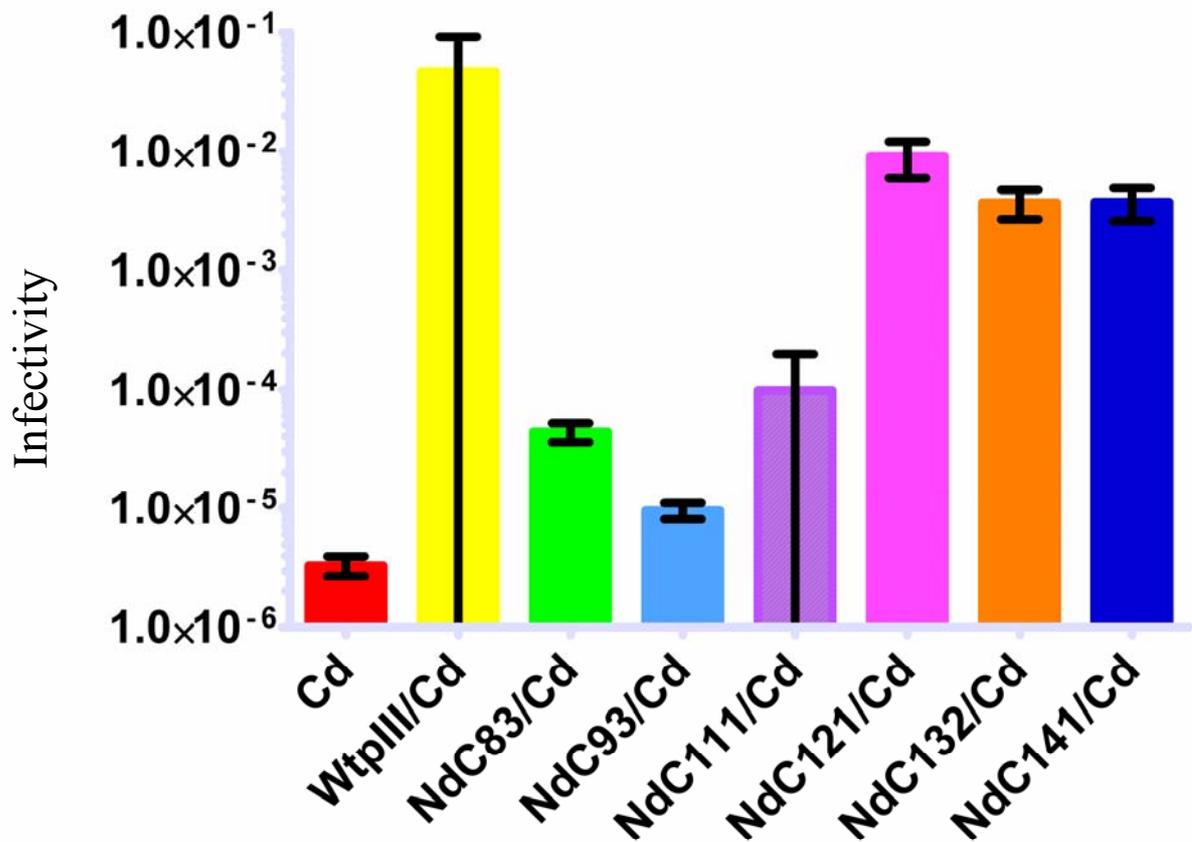
**Table 7. Infectivity phage particles containing the NdC mutants**

<b>Phage</b>	<b>Infectious Titre (pfu/mL)<sup>a</sup></b>	<b>Quantity of virions (ge/mL)<sup>b</sup></b>	<b>Infectivity (pfu/ge)<sup>c</sup></b>
Cd	$2.90 \pm 0.54 \times 10^7$	$8.69 \times 10^{13} \pm 3.74 \times 10^{10}$	$3.34 \pm 0.63 \times 10^{-6}$
WTpIII/Cd	$2.40 \pm 1.04 \times 10^{12}$	$5.21 \times 10^{13} \pm 5.68 \times 10^{10}$	$4.61 \pm 2.01 \times 10^{-2}$
NdC83/Cd	$2.52 \pm 0.45 \times 10^9$	$5.77 \times 10^{13} \pm 1.47 \times 10^{11}$	$4.36 \pm 0.80 \times 10^{-5}$
NdC93/Cd	$5.16 \pm 0.79 \times 10^8$	$5.44 \times 10^{13} \pm 1.06 \times 10^{10}$	$9.49 \pm 1.46 \times 10^{-6}$
NdC111/Cd	$6.47 \pm 2.48 \times 10^9$	$6.55 \times 10^{13} \pm 2.86 \times 10^{10}$	$9.87 \pm 3.79 \times 10^{-5}$
NdC121/Cd	$1.05 \pm 0.35 \times 10^{11}$	$1.16 \times 10^{13} \pm 8.30 \times 10^{10}$	$9.05 \pm 3.11 \times 10^{-3}$
NdC132/Cd	$9.33 \pm 2.52 \times 10^{10}$	$2.51 \times 10^{13} \pm 8.83 \times 10^{10}$	$3.72 \pm 1.02 \times 10^{-3}$
NdC141/Cd	$1.40 \pm 0.42 \times 10^{11}$	$3.72 \times 10^{13} \pm 6.82 \times 10^{10}$	$3.76 \pm 1.15 \times 10^{-3}$

<sup>a</sup>Titre of the number of infectious phage per mL, determined as the number of plaque-forming units.

<sup>b</sup>Quantity of virions, the number of genome equivalents (ge) per mL. Genome equivalents are defined in Material and Methods (Section 2.4.7; Figure 18)

<sup>c</sup>The error of infectivity was calculated by the following equation  $e[y] = (e[a] + y \times e[b])/b$ , where a = infectious titre, b = quantity of virions, y = infectivity.



**Figure 24. Infectivity of phage particles containing the NdC mutants.** Infectivity is a ratio of the infectious phage particles titre (plaque-forming units per mL) to the amount of phage particles (in genome equivalents per mL) expressed as a percentage value. The error of infectivity was calculated by the following equation  $e[y] = (e[a] + y \times e[b])/b$ , where  $a$  = infectious titre,  $b$  = quantity of virions,  $y$  = infectivity. Error bars represent one standard deviation from the mean.

### 3.6 A specific region of the C-domain is required for N1N2-mediated f1 infection

Using the data above the infectivity of each sample was calculated (Table 5, Figure 23). Infectivity is expressed as a percentage ratio of the number of infectious phagemid particles or transducing particles per genome equivalent (tdp/ge). Calculation showed that the infectivity of NdC83/Cd and NdC93/Cd particles was very low, at 0.0007% and 0.0014% respectively. This means in the case of NdC83/Cd only 0.0007% of the phagemid particles within the samples were able to infect a host cell. For NdC83/Cd and NdC93/Cd, this value is close or within the same order of magnitude as the Cd, the negative control phagemid particles devoid of the receptor binding domains N1N2 (0.0002%).

Virions containing NdC111 had an infectivity of 0.01%, which is two orders of magnitude higher than that of the negative control. This increase in infectivity is unlikely to be solely accounted for by an increased level of infection from reversion mutant *gIII*<sup>+</sup> input helper phage given the two order of magnitude increase over the negative control. Therefore, the NdC111 mutant is likely to mediate infection at a low efficiency.

NdC121, NdC132 and NdC141 had an infectivity of 2.4%, 7.2% and 10.3% respectively. These are all much higher than the other virions containing other mutants and the negative control and are within the same order of magnitude as the WTpIII/Cd that contains full-length pIII (positive control 5.1%). The difference between the infectivity of NdC132 and NdC141 containing particles and WTpIII containing particles is probably not due to more efficient infection by these deletion mutants, but rather due to intrinsic error in titration and phagemid particle quantification methods.

Thus, NdC121/Cd, NdC132/Cd and NdC141/Cd are all as infective as is the positive control. Therefore, the smallest tested C-terminal fragment that supports phage infection is a pIII C-terminal fragment of 121 residues, covalently linked to the functional N1N2 domain. The fragment of C-domain between residue C121 and C153 (32 residues in total) is redundant for infection.

### **3.7 Evidence for correct folding of the N1N2 domains of pIII**

The functional manifestation that the N1N2 domains of pIII are folding correctly is the cell is the acquiring resistance to fl infection (Boeke et al., 1982). All NdC deletion mutant constructs rendered the cells immune to fl infection when expression was induced. When NdC construct expression was induced, over 99% of cells became resistant to phage infection. In contrast, in the absence of induction, 60% to 90% of the cells were infected, showing that the resistance to infection depends on NdC expression.

An alternative explanation for the five orders of magnitude reduction in the infectivity in the NdC83/Cd and NdC93/Cd mutant phagemid particles relative to the positive control could be the failure of the N1N2 domains to fold correctly, and not by the C-terminal domain playing an important role in phage infection. If the N1N2 domains were misfolded then they would be unable to bind to the host cell receptors (F pilus and TolA). This would render the phage virions non-infectious, as they would no longer be able to bind to the host cell. In these experiments, it is assumed that the N1N2 domains have folded in the correct conformation for phage infection to occur. This assumption is based on the ample evidence that the folding of the N1N2 domains of pIII is independent of the folding status of the C-terminal domain.

Multiple NMR and X-ray crystallography (Holliger et al., 1999; Lubkowski et al., 1998; Lubkowski et al., 1999) structural studies of the N1N2 domain of pIII have all been carried out using N1 and N2 domains expressed separately from the C-domain. The best example that the N1N2 domain folds independently of the C-domain is described by Holliger et al., (1999). In this paper, the N1N2 domains were expressed without the C-domain of pIII in the periplasm where they folded correctly despite the absence of the C-domain. In addition to this, Lubkowski et al., (1998; 1999) used N1N2 domain that had been isolated from purified cytoplasmic inclusion bodies and refolded using a redox shuffle before crystallization. In these studies and that of Reichmann and Holliger (1997); isolated N1N2 domains were also shown to be able to still bind to the primary and secondary receptors, therefore they retained their biological function.

More recently, evidence that the N domain can and does fold independently of the C-domain was presented by Martin and Schmid (2003a; 2003b). These authors used a variety of techniques and mutants to analyse the folding kinetics of the N domain. They demonstrated that the N domain could be used as a model system for the folding kinetics and domain docking of a two-domain protein, again in the absence of the C-domain.

In all constructs used here, the N1N2 domains were separated from the C-domain by the natural long glycine-rich linker (GGGS GGGT GGGS GGGT), ensuring the independent N1N2 domain folding.

In addition, the infection of F-negative cells was affected in NdC83 mutants to the same extent, as it was that of the F<sup>+</sup> cells, showing that the infection was blocked after the step of binding to the F pilus (Bennett & Rakonjac, 2006). Therefore, the reduction in infectivity seen in mutants (at least in the NdC83 mutant) occurs post F pilus receptor binding.

In summary, infection resistance shows that the N-terminal domains in the NdC mutants fold correctly. This is supported by ample evidence from the literature that N1N2 fold into functional domains in the absence of the C-domain

### **3.8 Comparison of the infectivity of WTpIII/Cd positive control to complemented f1d3**

The positive control virions WTpIII/Cd that carried a mixture of full length wild type pIII and C-domain were calculated to have a relatively low infectivity (ratio of infective to total amount of virions (expressed in genome equivalents) at 5.1%. This value seems quite low, but is actually typical to the complemented  $\Delta gIII$  phage virions.

Low infectivity in complemented  $\Delta gIII$  mutants is probably due to the increase in ratio of polyphage to monophage virions. As defined in the Material and Methods (Section 2.4.7), a polyphage virion is a filamentous phage particle which carries more than one ssDNA genome (Figure 18). Thus if infective, a polyphage virion is counted as one

infection event even though it contains more than one genome equivalent per particle. Because polyphage are a prevalent type of virions in the  $\Delta gIII$  phage stocks relative to the WTpIII phage stocks, the infectivity of  $\Delta gIII$  phage is expected to be significantly lower than that of a normal fl phage stock (which is composed of 95% of monophage and 5% of polyphage virions). The somewhat lower infectivity of composite WTpIII/Cd virions that contain WTpIII in combination with C-domain relative to virions containing only WTpIII could be due to the reduced number of copies of the N1N2 domains in the WTpIII/Cd particles. Given the lower number of WTpIII per virion and probability that a fraction of virions will contain only the C-domain, these particles have a lower chance of interacting with the host receptors.

### **3.9 Effect of N1N2 stoichiometry on opening of phage cap complex in NdC/Cd carrying phage**

As stated in the introduction (Section 1.3.3) there are five subunits of pIII within the phage cap complex. This was shown in the Lee et al.(2002) when ZnS nanocrystals were bound to the pIII tip of the phage through a ZnS-specific peptide displayed on pIII. Under high magnification TEM, five crystals were observed bound to the pIII/pVI terminus.

NdC/Cd phage produced in this work carry a mixture of different pIII molecules and it is impossible to control ratio of NdC to Cd mutants within a single virion. Because of this, there is a multitude of different outcomes in the ratios of NdC to Cd mutants within an individual virion, and thus there can be different outcomes in the numbers of N1N2 receptor binding domains present in individual virions. Nevertheless, because the WTpIII and the Cd construct are expressed at similar levels, the different pIII mutants should have an equal chance of being assembled into a virion particle. Assuming that the two pIII constructs have a normal distribution along the host cell membrane the majority of phage particles should carry two or three of each type of the pIII protein supplied in there pIII/pVI complex (out of five subunits total). However, some virions will be terminated entirely with the NdC mutant or the Cd mutant. Virions terminated with the C domain only will be non-infectious and would contribute to the non-infectious virions within a sample. It is assumed that the level of Cd-only virions within

the different NdC/Cd samples is similar, and thus do not bias the infectivity calculations.

In summary, it is likely that the ratio of NdC mutants to Cd mutants in the virions does not have any bearing on these experiments. As shown in Bennett and Rakonjac (2006) and in this thesis, filamentous phage carrying a mixture of WTpIII and Cd are infectious. This means for a virion to be infective there is no requirement that the phage cap complex to be entirely composed of full-length pIII. It is probably that the minimum number of full-length pIII within the phage cap complex for successful infection is one. At this point there is no method available for analysing the stoichiometric ratio of NdC to Cd mutants within the phage cap complex and testing this hypothesis.

### **3.10 Discussion**

Previous studies of Ff phage infection have elucidated in detail the binding to its cellular receptors, the F pilus and TolA (Click & Webster, 1997; Click & Webster, 1998; Deng & Perham, 2002; Lubkowski et al., 1999; Reichmann & Holliger, 1997). However, until recently there has been little research into the mechanism of the later steps that “unlocks” the Ff virion during infection, resulting in the reinsertion of the major coat protein into the host cell membrane and the transport of the phage ssDNA genome across the cellular membrane into the cytoplasm of a new host cell (Click & Webster, 1998).

In Bennett and Rakonjac (2006) it was shown that the C-domain of pIII plays a role in the post-receptor binding stages of Ff filamentous phage infection. In that work, it was shown that the termination-incompetent C-terminal domain (C83) was also unable to mediate infection, indicating that the C-domain of pIII must play an important role in the process of phage infection. This means that the C-domain of pIII plays an important role in the membrane transactions involved in both phage infection and phage termination. This leads to the hypothesis that the processes of termination and infection may share a common mechanism. If so, it would be expected that the minimal portion of C-domain required for infection and release would be the same.

In my analysis of a series of nested C-domain deletion fused to N1N2 domain, full infectivity was reached with a C-domain fragment length of 121 residues from the C terminus (NdC121). Thus, the minimal requirement of the C-domain during phage infection is a portion containing the 121 C-terminal residues of the pIII C-domain. The next shorter mutant containing 111 C-terminal residues has marginal infectivity, whereas mutants containing shorter C-terminal fragments could not mediate infection. This is a much longer fragment of the C-domain in comparison to that required for virion release (93 C-terminal residues). Therefore, the minimum test fragments of the C-domain required for phage entry versus termination/release are different. For further discussion, it is worthwhile to emphasise that the intact (complete) C-domains that are mixed with the infection-incompetent NdC mutants cannot mediate infection even though they form a mixed multimer in the virion with NdC mutants which contain full-length N1N2 domains. This implies that N1N2 receptor-binding domain has to be covalently linked to the functional C-domain in order to mediate infection.

NdC111/Cd mutant virions showed infectivity an order of magnitude lower than the infectivity of NdC121/Cd, but two orders of magnitude higher than that of the negative control Cd. This indicates that the NdC111 carrying mutant is partially functional in infection.

Overall, data presented in this chapter defines an infection-competence sequence for phage infection between residues C121 and residue C93. This infection-competence sequence (ICS) is: C121-GKLDV VATDY GAAID GFIGD VSGLA NGNG-C93. It is likely to carry the important residues involved in the post-receptor binding steps of filamentous phage infection, but is redundant for termination of assembly and release from the host membrane.

In summary, the process of phage infection has a different minimal requirement for the C domain than phage assembly termination and release. It appears that the mechanism of phage entry relies on different residues of the C-domain relative to those involved in phage release. An infection-competence sequence of the C-domain required for infection has been determined, allowing further investigation of the mechanism of post-receptor stages of filamentous phage infection.

## Chapter 4

### Search for the pIII C-terminal domain residues involved in infection using alanine scanning mutagenesis

#### 4.1 Introduction

Results of the deletion analysis (Chapter 3) indicated that the minimum tested C-domain requirement for full phage infection was a C-terminal portion of 121 residues in length covalently linked to N1N2 domain. The next shorter fragment NdC111, containing 111 C-terminal residues, was partially infectious (1% the infectivity of the "121" fragment), whereas the further shorter mutants, NdC93 and NdC83 containing 93 and 83 C-terminal residues had infectivities that were close to that of the negative control. These findings suggest the region between residue C121 and C93 contains residues that are important for the process of phage infection, but are not required for phage release. This region will be referred to in this thesis as the infection-competence sequence (ICS). To determine the residues that are essential for infection, the ICS was systematically mutated by an alanine scanning mutagenesis approach.

Alanine scanning is the sequential replacement of amino-acid residues located within a region of interest with the amino acid alanine. Alanine is chosen for this purpose because removes the side chain atoms beyond the  $\beta$ -carbon that may be mediating inter- or intra-protein interactions, but does not greatly increase the flexibility of the main chain. Without a  $\beta$ -carbon, main chain flexibility increases, potentially resulting in misfolding due to electrostatic or steric effects (Cunningham & Wells, 1989).

If the process of infection is dependent on interactions mediated by the side chain of a particular amino acid within the ICS, then mutation of that particular residue to remove the side chain should cause a reduction in infectivity of virions containing the mutant.

## **4.2 Construction of the mutants**

pIII constructs containing alanine scan mutations of the ICS region were designed and constructed using ligation-mediated PCR as described in the Material and Methods (Section 2.3). Mutagenic oligonucleotides used in PCR were designed to replace individual wild-type codons of the ICS with the alanine codon GCG. The exception to this rule was where the positions to be mutated were already occupied by the amino acid alanine. In these cases, the residue was mutated into glycine (codon GGG).

The pIII mutants were then expressed and incorporated into the phagemid particles using a phage production system similar to the one used Chapter 3 (Section 3.2). The phagemid particles containing the alanine replacement mutants of pIII were then tested to measure their infectivity.

When constructing the alanine scan mutants, difficulties were encountered in isolating certain mutants. This may be possibly due to the toxic effects certain mutants can have on an *E.coli* host. Toxicity of certain pIII mutants has been previously observed in this thesis and in experiments performed by Rakonjac et al. (1999). Only 12 of the total 28 mutation have been successfully constructed and analysed. In this section of the thesis, mutants will be referred to by their position relative to the C terminus of the pIII protein. Thus, mutant C93 contains a mutation of the ninety-third amino acid counting from the C terminus of pIII.

## **4.3 Phagemid particle expression system**

For these experiments, infectivity of the phagemid particles carrying constructed pIII alanine replacement mutants of pIII was calculated. To produce particles carrying the pIII alanine replacement mutants, a few modifications of the phagemid production system were applied.

In the system for testing of the internal deletion mutants in Chapter 3, the phagemid that was used to produce the phagemid particles encoded a complete C-domain. This C-

domain was combined with the NdC deletion mutants to allow release from the host cell and the formation of the pIII-containing virion cap. This was required because some NdC deletion mutants were expected to either not terminate phage assembly correctly or to assemble a defective virion cap. In contrast to the internal deletion mutants, all alanine mutant constructs contain full-length pIII and no changes to the termination-competent portion of the C-domain were made. Hence, all point mutants were expected to terminate phage assembly properly so therefore there was no need to supply pIII C-domain to allow phage termination to occur correctly. Thus, a compatible phagemid (pNJB51) that did not express the pIII C-domain was used instead of the Cd-expressing phagemid pYW01. The exception to this was in the production of the negative control where pIII C-domain-expressing phagemid pYW01 was used to allow virion assembly.

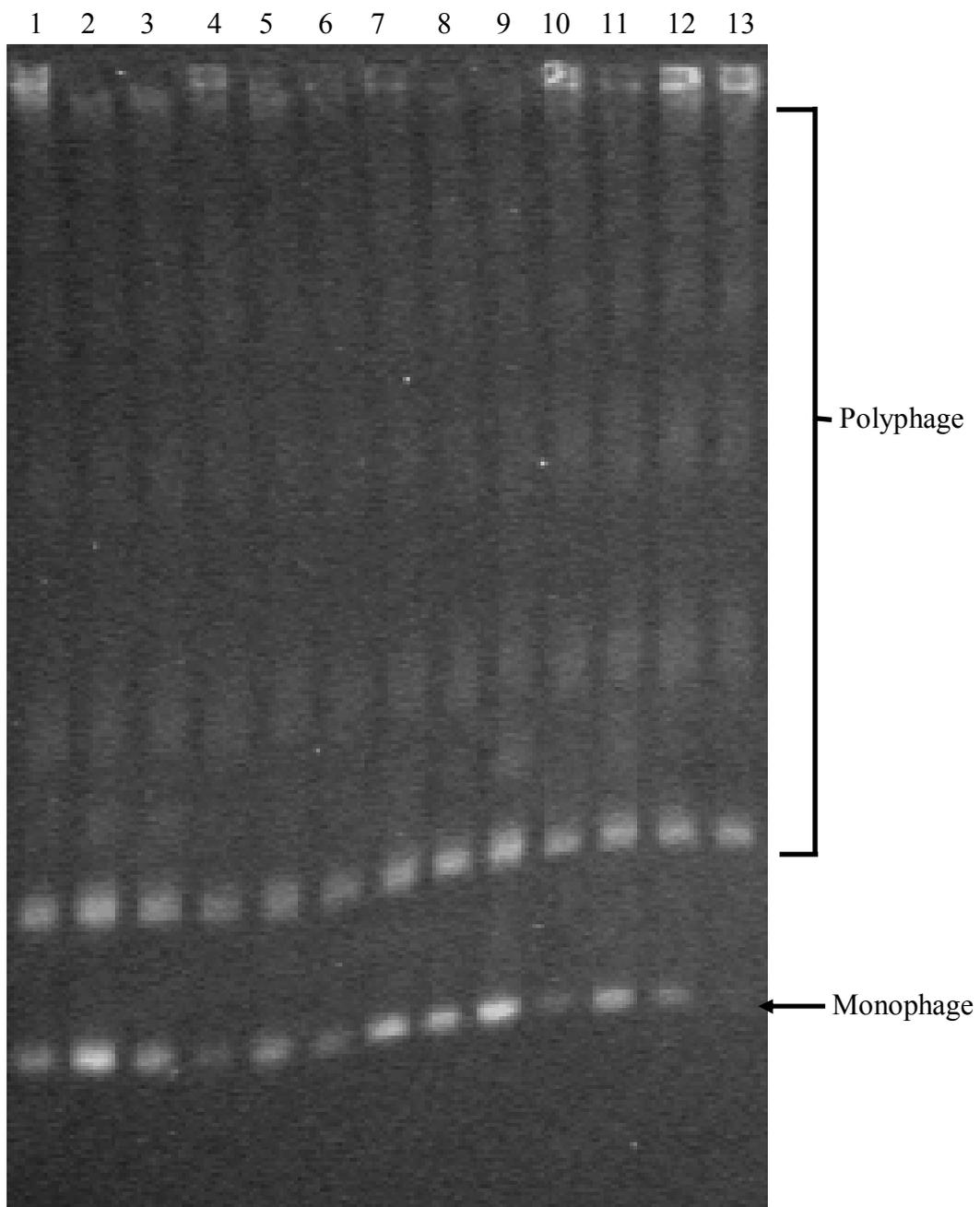
For the alanine scan series, the positive control phagemid particles contained only wild type pIII expressed from the pJARA200 plasmid. These phagemid particles were infectious and were used as the reference sample to which the infectivity of the phagemid particles carrying the alanine scan mutants was compared.

In the experimental setup, the acquired mutant is the only source of pIII used. Thus, the virion release is mediated solely by the supplied mutant pIII, and therefore the mutant pIII must be incorporated into the particles and function correctly in order to complete termination of virion assembly; hence, if monophage particles were being formed, assembly termination was occurring. Because monophage were produced, as shown by native virion agarose electrophoresis (Section 4.4), termination was occurring and the supplied mutant was being incorporated into the virions. Thus, it was not necessary to analyze the virions for the presence of pIII by western blotting. Given that the western blotting is the only experiment in which the excess helper phage from the input stock can affect the result, there was no need to remove unabsorbed input helper phage after infection. Therefore, the step of washing host cells of the excess input helper phage was omitted.

#### **4.4 Alanine scan mutants terminate phage assembly correctly**

As stated in Chapter 3, virion length is an important indication of pIII-mediated phage release. To ensure all alanine mutants terminated phage assembly correctly, the length of the phage virions was analysed by native phage electrophoresis (Figure 25). This showed that all of the alanine pIII mutants mediated phage assembly correctly, reflected by the similar monophage-polyphage ratio to that of the positive control (Figure 25). Unlike the native phage gel presented in chapter 3, (Figure 21) samples containing both phage and phagemid particles were present.

The alanine scan mutations were not tested for their stability in the detergent Sarkosyl, because the point mutations were not expected to alter stability of the virion.



**Figure 25. Native phage agarose electrophoresis of virions containing pIII point mutants.** Lane 1. Positive Control, virions containing WTpIII, Lanes 2-13. Virions containing mutations of individual ICS residues: C93, C97, C106, C107, C108, C109, C110, C111, C115, C116, C117, and C118. Monophage - pNJB51 phagemid particles containing a single genome. Polyphage - pNJB51 phagemid particles containing two or more sequentially packaged genomes. Helper phage titre in the sample was only 1/100 that of the phagemid (transducing) particles, therefore the amounts of virions containing the helper phage genomes or combination of helper phage and phagemid particle genome were too low to be detected. Gel was stained using SYBR gold.

## 4.5 Analysis of infectivity of mutant phagemid particles

To observe the effect of the alanine mutations on phage infection, the infectivity of the phagemid particles containing constructed pIII point mutants was determined.

Infectivity was calculated as a ratio of the number of infectious virion particles to total amount of virions in the sample (As defined in Section 2.4.7). As in the Chapter 3, infectivity calculations required determination of two values: the infectious titre (titre of transducing particles) and the phagemid particle concentration (number of genome equivalents per mL) in the samples.

The titre of the transducing particle was calculated using plate titration. This showed that the infective titre of the transducing particles containing pIII point mutants ranged between  $5 \times 10^{10}$  and  $5 \times 10^{11}$  cfu/mL. This is very similar to the titre of the positive control phagemid particles containing WTpIII ( $1.01 \times 10^{11}$  cfu/mL) and is four orders of magnitude greater than the titre of the negative control ( $1.72 \times 10^7$  cfu/mL) (Table 6).

Phagemid particle concentration in the pIII point mutant samples was assessed by densitometry. It was determined that the samples displaying the pIII point mutants contained between  $9 \times 10^{10}$  and  $1 \times 10^{12}$  phagemid genome equivalents per mL (Table 6). The positive control contained  $1.66 \times 10^{12}$  phagemid genome equivalents per mL and the negative control contained  $3.55 \times 10^{12}$  phagemid genome equivalents per mL

Using this data the infectivity of the alanine scan mutant phagemid particles was analysed (Table 6, Figure 26). Calculations show that the phagemid particles containing the alanine scan pIII mutants had infectivities ranging from 5.4% (C118) to 58% (C110). The WTpIII positive control had an infectivity of 6% and the negative control had an infectivity of 0.0005%. This means that all phagemids containing the pIII alanine scan mutants were infective at levels close to or higher than the WTpIII positive control.

Because some of the alanine scan mutants analysed produced phagemid particles that had infectivities tenfold higher relative to that of the positive control, a bootstrap analysis was used (Efron, 1979; Efron, 1981) to determine the standard deviation of infectivity and thus determine whether infectivities of any of these samples were

statistically significantly higher than the positive control. If so, the analysis thus would indicate that these mutations enhance infectivity. Bootstrap analysis was used because infectivity is measured as a ratio of infective virions to total virions, thus standard deviation cannot be calculated using simple statistical methods (Figure 26).

Bootstrap analysis determined that no sample had infectivity that was statistically significantly higher than that of the positive control. A statistically significant difference was defined by the predicted bootstrap data falling outside 95% of the positive control data (two standard deviations; Figure 26). This means that the large gain in infectivity shown by some mutants is due to experimental error and is not a real gain in infection.

The error for the negative control was not calculated because it varied too significantly from the other data for it to be subjected to bootstrap analysis.

**Table 8. Infectivity and Relative Infectivity of Alanine Point Mutations**

<b>Sample</b>	<b>Modified Residue<sup>c</sup></b>	<b>Infective Titre (tdp/mL)</b>	<b>Phagemid Genome Equivalents<sup>d</sup> (ge/mL)</b>	<b>Infectivity<sup>e</sup> (tdp/ge)</b>
WTpIII (+) <sup>a</sup>	N/A	$1.01 \pm 0.58 \times 10^{11}$	$1.66 \pm 0.73 \times 10^{12}$	$6.07 \pm 0.19 \times 10^{-2}$
Cd (-) <sup>b</sup>	N/A	$1.72 \pm 1.50 \times 10^7$	$3.55 \pm 1.34 \times 10^{12}$	$4.83 \times 10^{-6}$
93	Gly	$3.72 \pm 1.62 \times 10^{11}$	$7.98 \pm 3.45 \times 10^{11}$	$4.66 \pm 4.28 \times 10^{-1}$
97	Ala	$1.11 \pm 0.83 \times 10^{11}$	$1.79 \pm 1.05 \times 10^{12}$	$6.20 \pm 0.15 \times 10^{-2}$
106	Gly	$5.67 \pm 2.60 \times 10^{10}$	$2.39 \pm 1.43 \times 10^{11}$	$2.38 \pm 4.29 \times 10^{-1}$
107	Asp	$2.09 \pm 1.34 \times 10^{11}$	$1.36 \pm 0.55 \times 10^{12}$	$1.54 \pm 2.77 \times 10^{-1}$
108	Ile	$5.86 \pm 5.38 \times 10^{10}$	$9.05 \pm 3.46 \times 10^{11}$	$6.48 \pm 0.32 \times 10^{-2}$
109	Ala	$3.60 \pm 3.93 \times 10^{11}$	$1.91 \pm 0.85 \times 10^{12}$	$1.89 \pm 2.29 \times 10^{-1}$
110	Ala	$2.60 \pm 1.37 \times 10^{11}$	$4.48 \pm 1.36 \times 10^{11}$	$5.80 \pm 4.60 \times 10^{-1}$
111	Gly	$3.57 \pm 2.03 \times 10^{11}$	$1.80 \pm 1.05 \times 10^{12}$	$1.98 \pm 2.44 \times 10^{-1}$
115	Ala	$4.68 \pm 2.89 \times 10^{10}$	$3.75 \pm 2.51 \times 10^{11}$	$1.25 \pm 4.16 \times 10^{-1}$
116	Val	$3.45 \pm 2.36 \times 10^{11}$	$1.66 \pm 0.62 \times 10^{12}$	$2.08 \pm 2.58 \times 10^{-1}$
117	Ser	$7.18 \pm 3.25 \times 10^{10}$	$2.05 \pm 0.96 \times 10^{11}$	$3.50 \pm 4.48 \times 10^{-1}$
118	Asp	$1.26 \pm 0.85 \times 10^9$	$2.29 \pm 1.39 \times 10^{11}$	$5.49 \pm 0.44 \times 10^{-2}$

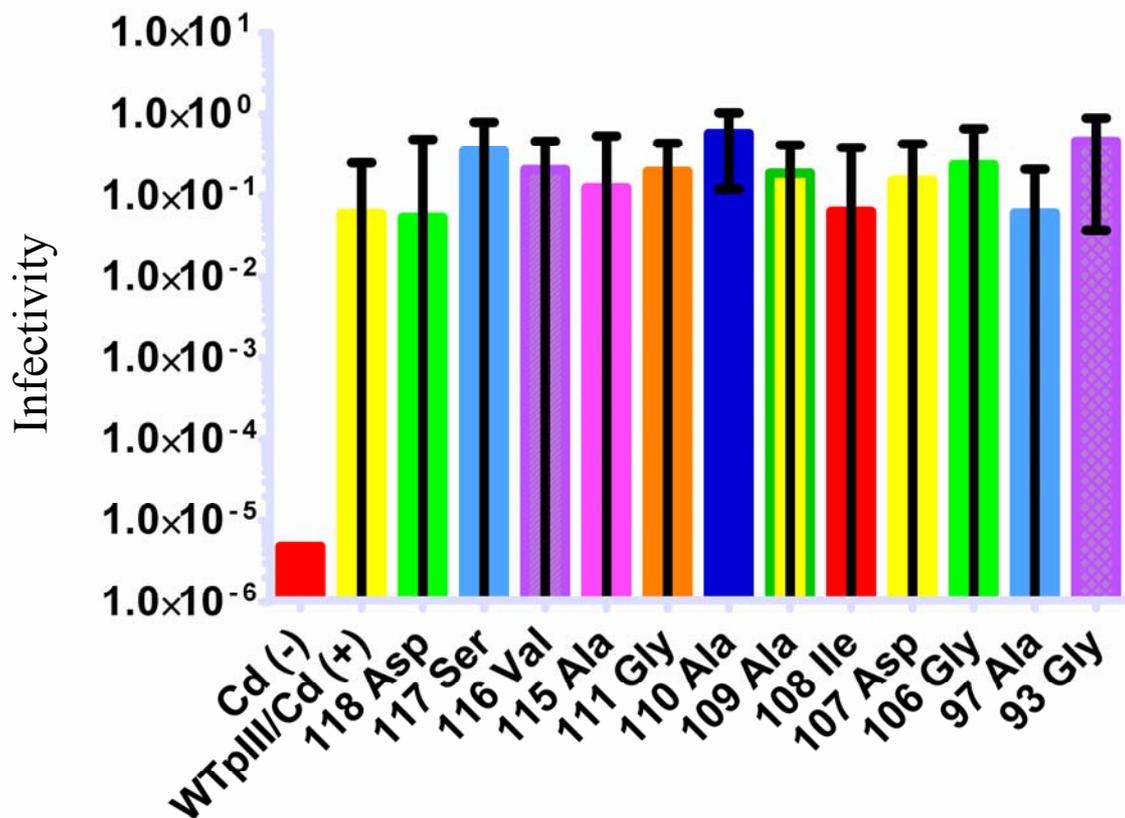
<sup>a</sup> Positive control, phagemid particles containing WTpIII.

<sup>b</sup> Negative control, phagemid particles containing only the C-domain of pIII.

<sup>c</sup> Alanine residues were mutated to Glycine; all other residues were mutated to Alanine.

<sup>d</sup> Phagemid genome equivalents were determined by densitometry as described in the material and methods (Section 2.4.7), except that the DNA was stained with SYBR-gold instead of Ethidium Bromide.

<sup>e</sup> Error of infectivity was measured using Bootstrap analysis as described in the Material and Methods (Section 2.5), except for the Cd (-) control which could not be measured using this form of analysis.



**Figure 26. Infectivity of virions containing pIII point mutants.** Number below the graph denotes the position (relative to the C terminus of the mutated residue) and the amino acid in the WTPIII. Alanine 109, 110 and 115 were replaced with Glycine. The error obtained from bootstrap analysis. Error bars represent two standard deviations from the mean infectivity for each sample. Number below the graph denotes the position (relative to the C terminus of the mutated residue) and the amino acid in the WTPIII.

## 4.6 Summary of results

The partial alanine scan of the ICS sequence tested 12 out of 28 amino acid residues: C118, C117, C116, C115, C111, C110, C109, C108, C107, C106, C97, C93 for their role in fl infection. Mutations of none of the 12 mutated residues examined thus far have caused any statistically significant changes in the infectivity of virions relative to the positive control. Hence, the simple conclusion of the results of the 12 pIII point mutants analysed this far, is that the side chains of the mutated residues are not important in the process of infection. Some of the mutated residues were charged (C107 Asp, C118 Asp), yet the removal of their charge did not affect infectivity, suggesting no role in infection.

For a full assessment of the ICS region of the C-domain and its involvement in infection, the remaining positions in the ICS need to be mutated.

## 4.7 Secondary structure prediction

While the high-resolution three-dimensional structure of a protein can only be determined using experimental methods like X-ray crystallography and NMR spectroscopy, there are many statistics-based computational methods that can be used to predict the secondary structure elements of a protein based on the primary sequence.

The pIII sequence was submitted for analysis using the JPRED secondary structure prediction server which operates using multiple algorithms based on the primary sequence and the propensity of certain amino acids to be found in certain secondary structures (Cole, Barber & Barton, 2008). JPRED predicted that the ICS was statistically most likely to be a  $\alpha$  helix (Figure 27C). A helical wheel plot of the indicated helix predicts that the ICS sequence forms an amphipathic helix (Figure 27F)(O'Neil & Grisham, 1997).

The pIII sequence was also submitted the PONDR (Predictor of Naturally Disordered Regions) to predict disordered region in the C-domain (Garner et al., 1999; Li et al., 1999; Romero, Obradovic & Dunker, 1997; Romero et al., 2001). PONDR predicted

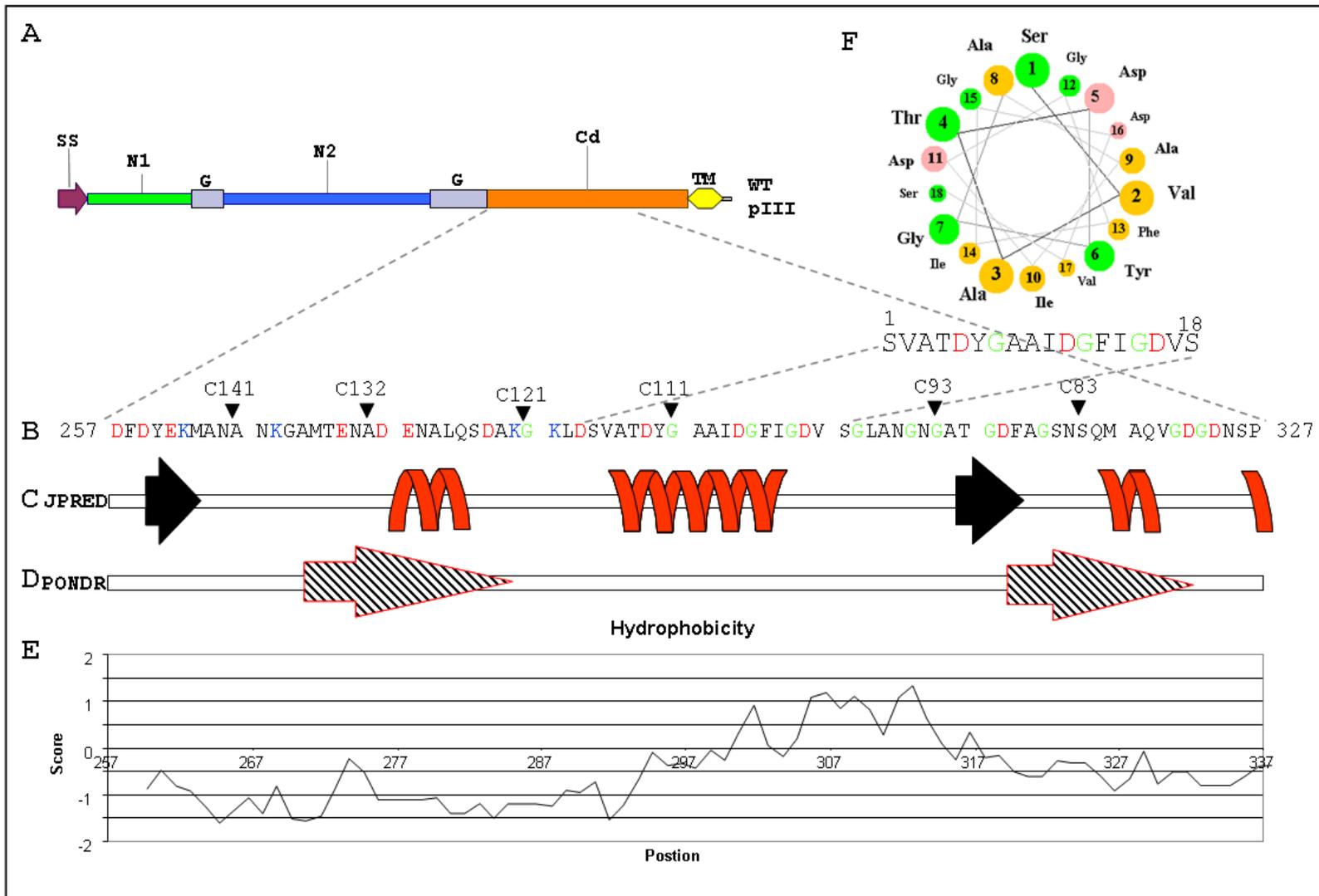
that the ICS region within the C-domain of pIII is not likely to be a naturally disordered region (Figure 27D) and hence is likely to form an ordered structure. This reinforces the prediction made by JPRED that the ICS region of pIII is likely to be a  $\alpha$  helix.

A hydrophobicity plot of the C domain of pIII also shows that the ICS region is significantly more hydrophobic than the rest of the C domain of pIII (Figure 27E)(Kyte & Doolittle, 1982).

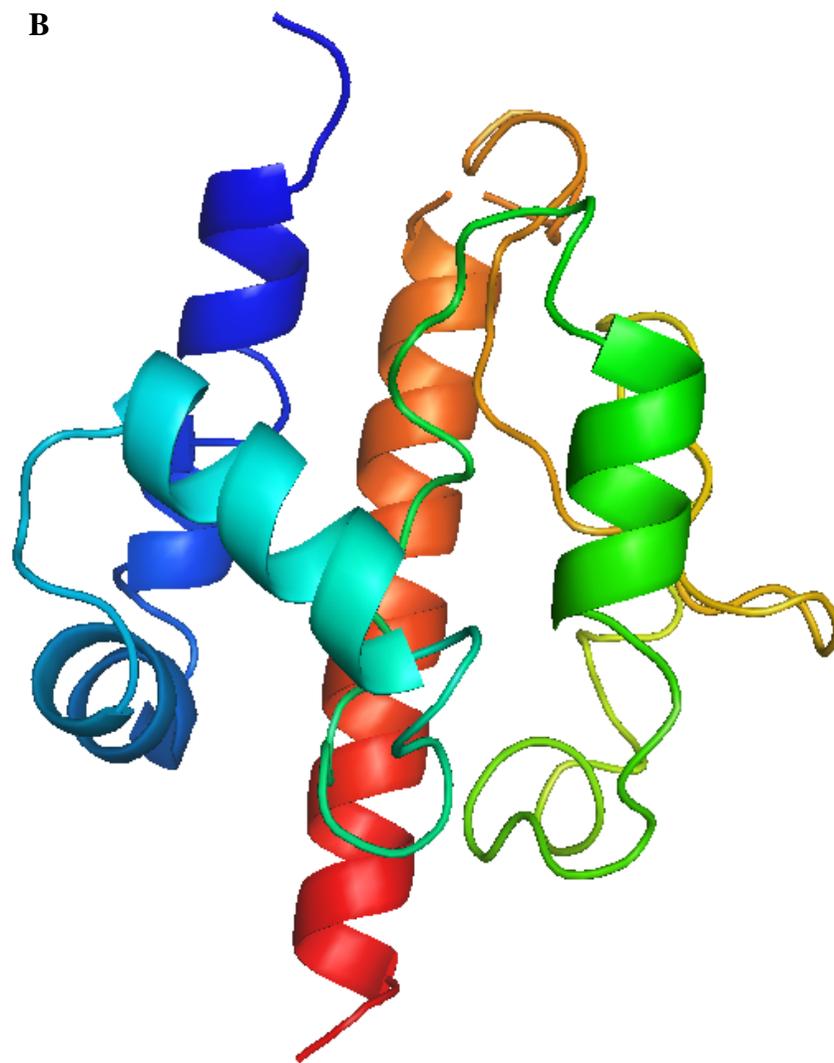
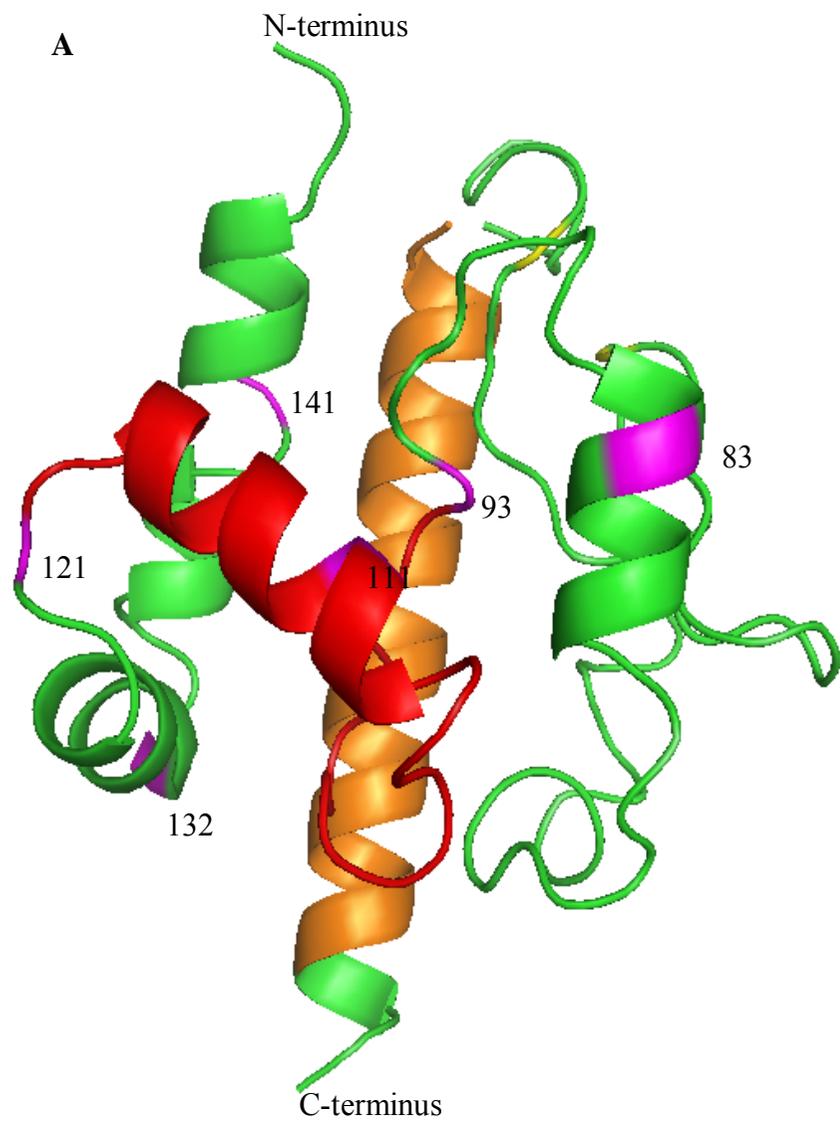
With no available high-resolution structure of pIII C-domain or any of the filamentous phage homologues, no significant hits were obtained by submitting the C-domain to SWISS-MODEL for analysis (Arnold et al., 2006; Guex & Peitsch, 1997; Kopp & Schwede, 2004; Peitsch, 1995; Schwede et al., 2003). SWISS-MODEL is a web-based homology modelling server which scans the query sequence against the ExNRL-3D database using a BLASTP2 algorithm. The ExNRL-3D contains the sequence of all proteins with known structures. SWISS-MODEL failed to find any structures currently within the ExNRL-3D database that had sufficient sequence homology (<25%) to the pIII C-domain for modelling to be feasible.

Because SWISS-MODEL failed to find a homologues structure based on sequence alignments to model the C-domain of pIII, the C-domain sequence, including the transmembrane anchor helix, was submitted to the I-TASSER server (Zhang, 2007; Zhang, 2008). I-TASSER is also a homology modelling server like SWISS-MODEL. However, I-TASSER uses the LOMETS meta-search algorithm to search the PDB database. This analysis based on secondary structure prediction homology rather than sequence homology. I-TASSER then uses various reconstruction methods to produce a series of predicted 3D models of the query protein based on homology modelling against structures found using the LOMETS search. The I-TASSER model best fitting all known structural data is presented in Figure 28. Full I-TASSER results are shown in Appendix 4. This model was determined to be the best fitting model, because firstly it contained the di-sulphide bond known to within the C-domain required for pIII function (Kremser & Rasched, 1994). This model (shown in Figure 28) also fits the available molecular biology data. Because a fragment of containing the final 93 residues of the C-domain is capable of terminating phage assembly (Rakonjac et al., 1999) it implies that the folding and structure of the final 93 residues of the C-domains is independent of the

folding the 60 N-terminal residues of the C-domain. The model chosen also fits this data because it does not contain any intermixed sheets or helices, hence the folding of the C-terminal 93 residue is not affected by the N-terminal residues of the C-domain.



**Figure 27. Bioinformatics analysis of the C domain of pIII.** **A.** Domain organisation of full-length pIII. SS, signal sequence (purple); N1, N1 domain (green); G, Glycine linker (grey); N2, N2 domain (blue); Cd, C domain (orange); TM, transmembrane anchor (yellow). **B.** Primary sequence of pIII C domain from residue 257 to residue 337, covering the region of nested deletions in NdC mutants. Acidic residues are highlighted in red; basic residues in blue; Glycine residue are in green. **C.** Schematic representation of the secondary structure regions predicted by JPRED algorithm (Cole et al., 2008). **D.** Schematic representation of predicted regions of naturally disordered sequence using the PONDR VLXT algorithm (Garner et al., 1999; Li et al., 1999; Romero et al., 1997; Romero et al., 2001). **E.** Hydrophobicity plot obtained using hydrophobicity values from (Kyte & Doolittle, 1982) and a nine residue window. **F.** Helical wheel diagram of region corresponding to predicted helix SVATDYGAAIDGFIGVS, obtained using a helical wheel algorithm (O'Neil & Grisham, 1997).



**Figure 28. Structural model of pIII C-domain A.** I-TASSER predicted structural model of pIII C-domain orientated with ICS region toward the viewer. Regions indicated: ICS (red), transmembrane anchor helix (orange), C-domain deletion mutant points (magenta), Numbers correspond to numbering from C terminus. **B.** I-TASSER predicted structural model of pIII C-domain in same orientation as model A. Model is coloured in a spectrum from Blue (N-terminal) to Red (C-terminal) to aid visualisation of the progression of the polypeptide chain. Images were obtained from pdb file supplied by I-Tasser and imaged using PyMol.

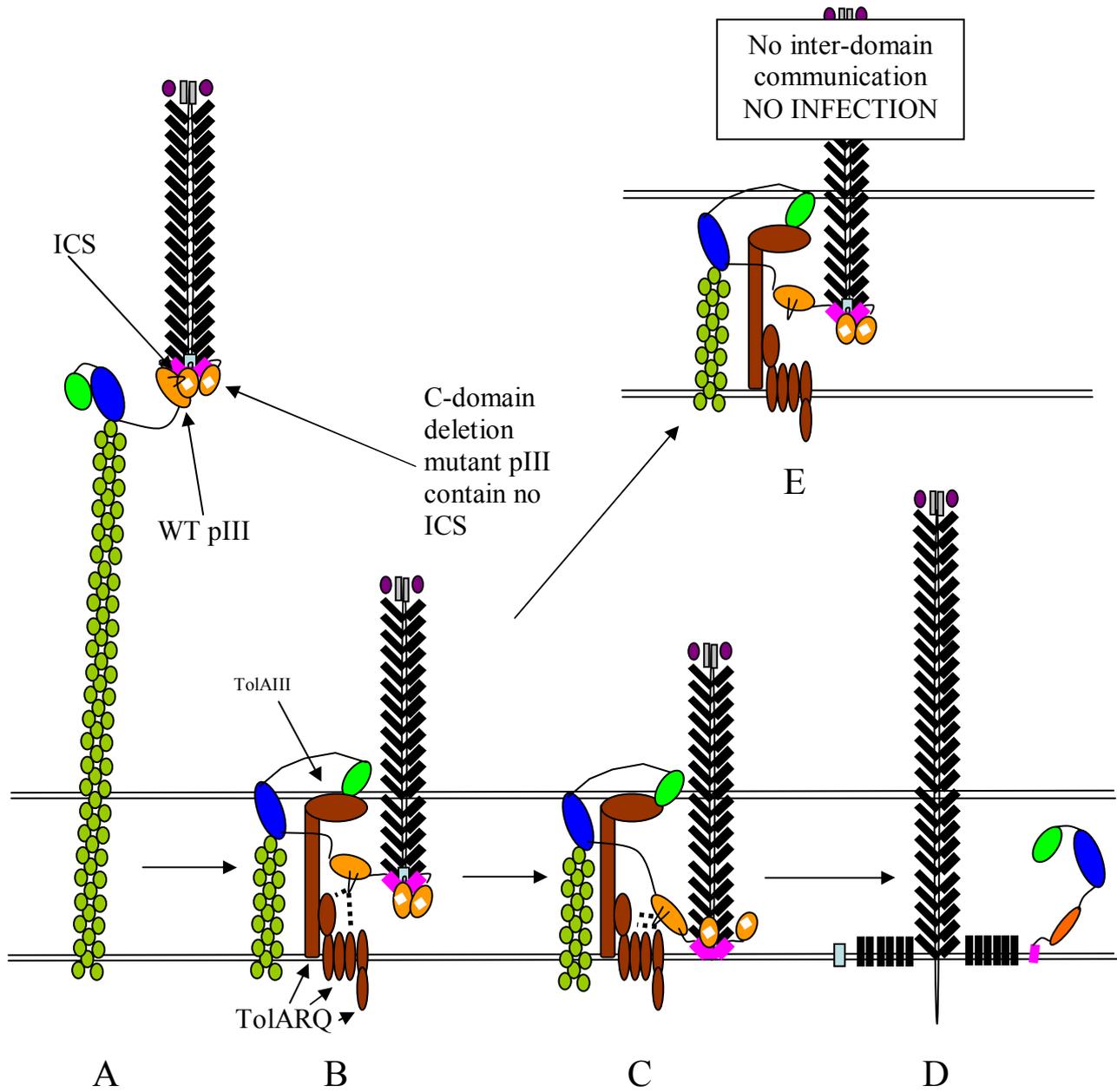
#### **4.8 The ICS region is not involved in the intra-cap complex signalling**

The C-domain of pIII and pVI form a complex, which terminates virion assembly and caps the virion. This complex contains five subunits of pIII (Lee et al., 2002). Because of the multi-subunit structure of the phage cap complex it is possible that the ICS may be involved in the transmission of the "unlock" signal to the other C-domains within the phage cap complex.

To test this possibility an experiment was designed and performed. If the ICS region was involved in transmitting the signal from the receptor-bound N1N2 domains of a pIII that had bound to the pIII C-domains of the other pIII subunits within the phage cap complex, then a C-domain mutant shorter than C111 should block infection. This would occur because the transmission of the intra-domain signal would be blocked at the post-receptor binding step, preventing infection.

Phagemid particles containing a mixture of full length WTpIII (pJARA140) and a series of nested deletions of the C-domains that lack the N1N2 domains were produced (Rakonjac, 1998). Titration analysis of these virions showed that none of the truncated C-domains affected the infectivity of the produced phagemid particles (Table 9). All WTpIII/C truncated phagemid samples had an infective titre similar to that of particles containing a mixture of WTpIII and a complete C-domain.

In conclusion, the ICS of the C-domain is mediating entry only *in cis* with the receptor-binding N1N2 domains. Therefore, the ICS within the subunit that mediates entry does not affect or require the ICS sequence within the other subunits in the same virion.



**Figure 29. Scenarios of pIII inter-subunit cooperation during entry.** **A.** Virions carrying WTpIII and a C-domain deletion mutant that do not contain ICS bind to F pilus via N2 domain of pIII. **B.** The F pilus retracts and N1 domain binds TolQRA via TolAIII domain. C-domain of pIII undergoes conformational change. If ICS is involved in intra phage cap complex signalling during infection, situation E occurs. If ICS is not involved in intra phage cap complex signalling then ICS forms further interactions with TolQRA (or another host cell protein). **C.** The C-domain opens and pIII transmembrane helix is inserted into host cell membrane. **D.** Phage ssDNA genome is transported into host cell cytoplasm. Phage virion proteins integrate into the host cell membrane. **E.** If ICS is involved in a receptor-triggered coordinated conformational change of pIII subunits in the virion cap to open the virion, the ICS-negative subunits are expected to block infection due to the lack of required cooperative conformational change. Because pIII C-domain deletion mutants do not contain ICS, "the open" signal is not perpetuated within phage cap complex. Phage complex remains shut. No infection occurs.

**Table 9: Infective titre of phagemid particles containing WTpIII and nested deletions of the C-domain**

<b>Sample</b>	<b>Phage particles (pfu/ml)</b>	<b>Phagemid particles (tdp/ml)</b>
WTpIII/Full length Cd	$5.37 \pm 0.97 \times 10^8$	$1.43 \pm 0.45 \times 10^9$
WTpIII/C148	$1.28 \pm 0.17 \times 10^8$	$3.22 \pm 0.62 \times 10^9$
WTpIII/141	$2.97 \pm 1.10 \times 10^9$	$1.73 \pm 0.15 \times 10^{10}$
WTpIII/C132	$2.53 \pm 0.18 \times 10^9$	$1.16 \pm 0.47 \times 10^{10}$
WTpIII/C121	$7.57 \pm 1.50 \times 10^{10}$	$3.01 \pm 0.81 \times 10^{10}$
WTpIII/C111	$1.67 \pm 1.15 \times 10^9$	$3.67 \pm 0.85 \times 10^{10}$
WTpIII/C93	$3.33 \pm 1.15 \times 10^9$	$5.72 \pm 1.15 \times 10^{10}$
WTpIII/C83	$1.03 \pm 0.42 \times 10^{10}$	$6.22 \pm 0.85 \times 10^{10}$

## Chapter 5

### Construction of an efficient microphage-producing system

#### 5.1 Introduction

Published work and work presented in this thesis have demonstrated that the C-domain is essential for both virion assembly (Rakonjac et al., 1999; Rakonjac & Model, 1998) and phage infection (Bennett and Rakonjac, 2006; Chapter 3). However, the mechanism by which the C-domain of pIII mediates these processes remains obscure, given the lack of information on the C-domain structure before assembly into the phage, after assembly into the phage and after phage infection. In the host cell, the C-domain is an integral membrane protein that does not interact with other coat proteins and has a short half life (Rakonjac, unpublished research; Rakonjac et al., 1999), whereas in the phage virion it forms multiple strong interactions with virion coat proteins pVI and pVIII within the virion cap (Gailus & Rasched, 1994). This makes purification of the pIII C-domain from both cells and virions and subsequent high-resolution structural analysis very challenging.

Once incorporated into the virion, the C-domain of pIII and another minor virion protein, pVI, form a multimeric "cap" complex at one of the two ends of the virion. This structure is highly stable (Gailus & Rasched, 1994) and is probably the best form of the pIII C-domain for structural analysis. Unfortunately, in the native virion, this cap structure is located at the very end of the long and flexible filament (860nm × 6nm) that makes up the virion coat, and represents only a very minor fraction of the virion. This makes native virions unsuitable for structural analysis of the cap structures. To increase the proportion of the virion occupied by the cap structures and to reduce virion flexibility, a system for the production of extremely short virions (microphage) was used. An existing low-efficiency system for producing microphage (Specthrie et al., 1992) was modified, to obtain a highly efficient system. In addition, the protocol for purification of the microphage was modified to obtain highly purified microphage.

## **5.2 Reconstruction of the microphage origin and cloning into the high copy number plasmid pCR4Blunt-TOPO.**

The original microphage-producing plasmid pLS7, constructed by Specthrie, Bullitt et al., (1992) used as a starting point for this work, was obtained from M. Russel. This system consisted of a microphage producing cassette, 200bp in length, which is composed of the fl packaging signal flanked on both sides by fl positive origins of replication (Chapter 2, Figure 11). This, in conjunction with a helper phage, produces small circular ssDNA fragments by initiation at the upstream positive origin of replication and termination at the downstream positive origin of replication. The small ssDNA circles corresponding to the microphage-producing cassette are then packaged, by the phage export machinery, into very short virion particles. In this thesis, these short virion particles will be referred to as microphage.

Sequence analysis of plasmid pLS7 showed that the upstream positive strand origin of replication contained a single nucleotide deletion. It has been shown that single nucleotide deletions within phage positive origins of replication are sufficient to prevent replication (Dotto et al., 1981). This was consistent with the testing of pLS7, which showed that no microphage ssDNA was produced upon infection with a helper phage (Bennett, 2005). To obtain the active origin the microphage origin cassette was amplified and repaired by PCR-mediated mutagenesis, using pLS7 as the template, appropriate flanking primers (NB26 and NB28) and a mutagenic primer (NB27, Table 2). The PCR product was then inserted into pCR4Blunt-TOPO vector (Invitrogen) and examined using DNA sequencing. This showed that the mutation in pLS7 had been successfully corrected. The new microphage-producing plasmid was named pNJB7 (Bennett, 2005).

The vector pCR4Blunt-TOPO was chosen to host the microphage producing origin as it is a high-copy number plasmid (200 copies per cell). The original microphage plasmid was derived from the low-copy number vector pBR322 (20 copies per cell). Since the microphage-producing cassette does not contain the fl negative strand origin of replication, the efficiency of the production of microphage ssDNA genomes is poor. Therefore, it was assumed that a high-copy number plasmid would provide larger numbers of templates for positive strand synthesis and would therefore result in a larger

number of microphage ssDNA genomes produced, in turn producing a larger number of microphage virions relative to a low-copy number plasmid.

Interestingly, when the repaired microphage-producing region was inserted into pCR4Blunt-TOPO, the recovered plasmid formed a ladder of bands when subjected to agarose gel electrophoresis, corresponding in size to multiple-length recombinant plasmid. Investigation of the new plasmid (pNJB7) by digestion with the restriction endonuclease *EcoRI* gave only two linear dsDNA products (Data not shown). These were of the size expected if the plasmid contained a single copy of the microphage origin cassette insert and a single copy of the original linearized vector. This confirmed that the plasmid had undergone simple duplexing upon microphage origin insertion. Because of this propensity of the microphage vector to induce duplex formation, in all microphage growth experiments plasmid pNJB7 was transformed into the host strain K1030 only just prior to the experiment.

### **5.3 A helper phage for efficient microphage production**

In the paper by Specthrie, Bullitt et al., (1992) it was found that classic helper phage R408 did not produce a high yield of microphage. Instead, a modified helper phage, R474, was found to produce the highest yield of microphage. Helper phage R474 contains a mutation within the *gVIII* promoter causing a decrease in the production of pVIII. It is thought that this reduction in pVIII available for phage assembly favours termination over elongation in phage assembly. This in turn favours the release of short microphage over the assembly of the much longer helper phage. However, this mutation also lowers the expression of minor coat proteins pIX, pIII and pVI, of which the latter two are required for phage release (Lopez & Webster, 1983; Rakonjac & Model, 1998). This results in stocks of R474 having much lower titres than those of the standard helper phage R408 ( $10^6$  vs  $10^{11}$ ).

Stocks of the standard helper phage R408 always have a 4 to 5 orders of magnitude higher infective titer than do R474 stocks and were therefore tested for microphage production, however very few microphage particles were produced (Data not shown).

Thus, neither R408 nor R474 were suitable as a helper phage for high-efficiency production of microphage.

To combine the high titer of R408 with the decreased pVIII production required for the preferential assembly of microphage over the helper phage, a low-efficiency suppression of *gVIII<sup>am</sup>* mutation at base pair 73 of *gVIII* (Figure 30), converting a GAG codon to a TAG stop codon (Feng, Model & Russel, 1999), was used. The resulting amber mutation truncates the protein product at residue 2 of the mature pVIII protein; hence, no pVIII protein is produced. When this mutation is suppressed in a *supD* strain of *E.coli* (K1030), full-length pVIII is translated at about 10% of the wild type levels, which is high enough to produce a relatively high titre stock of helper phage. The virions produced by suppression of the *gVIII<sup>am</sup>* mutant are exclusively monomeric particles (Rakonjac, 1998). This is in contrast to WT virions stocks that normally contain 5% longer virions. Thus, in the case of the *gVIII<sup>am</sup>* mutation, virion assembly termination is strongly favoured over virion elongation.

To construct the new helper phage, the *gVIII<sup>am</sup>* mutation was introduced into R408 helper phage. R408 contains the interference resistant origin of replication and truncated packaging signal (Russel et al., 1986). The interference resistant origin of replication prevents interference between the microphage origin and the R408 origin (See Chapter 1, Section 1.4.6). The truncated packaging signal in R408 decreases the efficiency at which this helper phage is packaged, thus favouring the packaging of microphage ssDNA, which contains an intact packaging signal. The R408 *gVIII<sup>am</sup>* phage R777, produced stocks whose titre was in the range of between  $10^{10}$ - $10^{11}$  pfu/mL. This helper phage was used to obtain microphage.

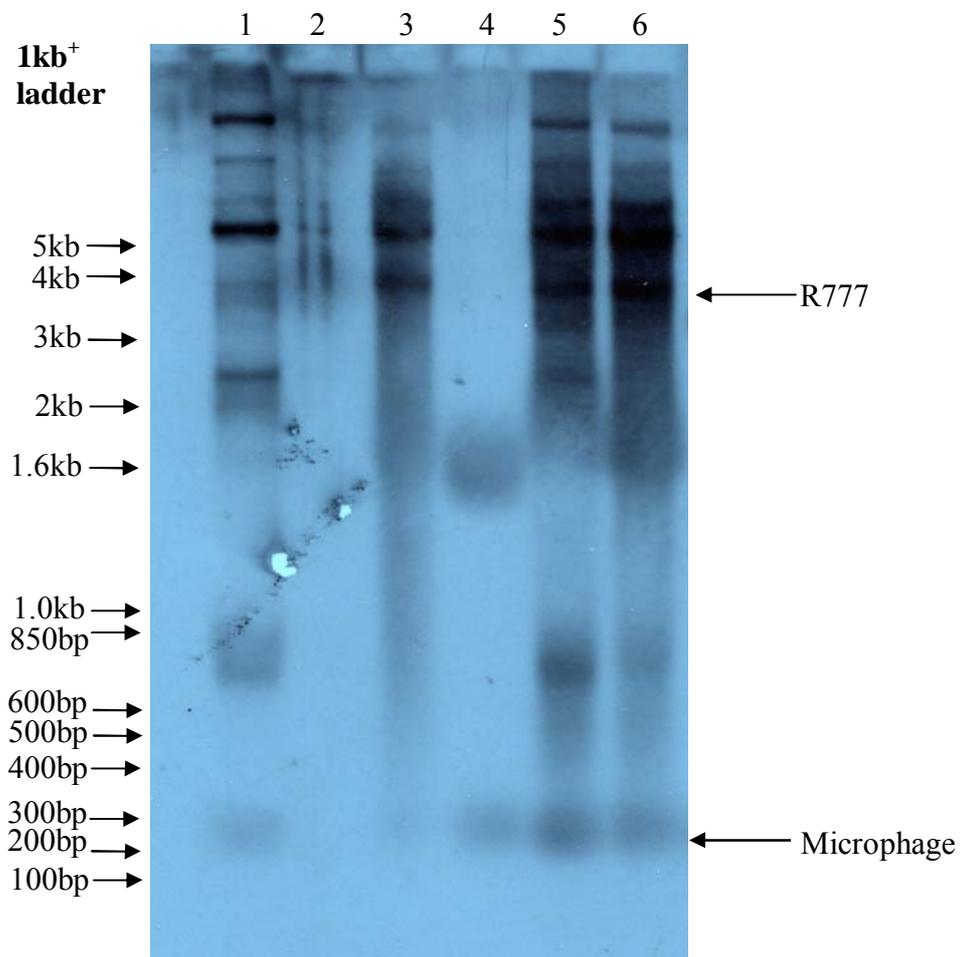
	1	50
	MetLysLysSerLeuValLeuLysAlaSerValAlaValAlaThrLeuVal	
<i>gVIII</i>	(1)	ATGAAAAAGTCTTTAGTCCTCAAAGCCTCTGTAGCCGTTGCTACCCTCGT
<i>gVIII<sup>am</sup></i>	(1)	ATGAAAAAGTCTTTAGTCCTCAAAGCCTCTGTAGCCGTTGCTACCCTCGT
	51	100
	ProMetLeuSerPheAlaAlaGluGlyAspAspProAlaLysAlaAlaPhe	
<i>gVIII</i>	(51)	TCCGATGCTGTCTTTTCGCTGCTGAGGGTGACGATCCCGCAAAGCGGCCT
<i>gVIII<sup>am</sup></i>	(51)	TCCGATGCTGTCTTTTCGCTGCTTAGGGTGACGATCCCGCAAAGCGGCCT
		<div style="border: 1px solid black; display: inline-block; padding: 2px;">TAG</div> 
	101	150
	AspSerLeuGlnAlaSerAlaThrGluTyrIleGlyTyrAlaTrpAla	
<i>gVIII</i>	(101)	TTGACTCCCTGCAAGCCTCAGCGACCGAATATATCGGTTATGCGTGGGCG
<i>gVIII<sup>am</sup></i>	(101)	TTGACTCCCTGCAAGCCTCAGCGACCGAATATATCGGTTATGCGTGGGCG
	151	200
	MetValValValIleValGlyAlaThrIleGlyIleLysLeuPheLysLys	
<i>gVIII</i>	(151)	ATGGTTGTTGTCATTGTTCGGCGCAACTATCGGTATCAAGCTGTTTAAGAA
<i>gVIII<sup>am</sup></i>	(151)	ATGGTTGTTGTCATTGTTCGGCGCAACTATCGGTATCAAGCTGTTTAAGAA
	201	222
	PheThrSerLysAlaSer***	
<i>gVIII</i>	(201)	ATTACCTCGAAAGCAAGCTGA
<i>gVIII<sup>am</sup></i>	(201)	ATTACCTCGAAAGCAAGCTGA

**Figure 30. Position of mutated codon in *gVIII<sup>am</sup>* mutant used in this thesis.**

## 5.4 Concentration and initial purification of microphage

Microphage were produced by infecting a *supD E.coli* strain (K1030) transformed with plasmid pNJB7. Microphage were concentrated and partially separated from the helper phage by differential PEG precipitation. This involved a two-step PEG precipitation as outlined in Specthrie et al (1992) and Material and Methods (Section 2.4.9). Firstly, a low PEG (2.5%) concentration was used to precipitate the majority of the helper phage virions. Then, after a centrifugation step, the supernatant was subjected to precipitation with a high PEG (15%) concentration to precipitate the microphage.

Upon resuspension of the "high" PEG precipitation three distinct layers as well as insoluble material were obtained, suggesting a high concentration of lipids in the precipitated material. Analysis of fractions by agarose gel electrophoresis of DNA released from SDS-disassociated virions followed by a Southern blot indicated that the "high PEG" fractions, which were expected to be enriched for the microphage, indeed contained microphage (Figure 31, lanes 4-6). All Microphage-positive "high PEG" fractions were pooled, diluted in DNase I buffer and then subjected to digestion with DNase I and RNase to remove any precipitated cellular DNA and RNA from the sample. Remnants of the cellular debris and lipid membranes were then further solubilized using detergents Triton X-100 and Sarkosyl (Material and Methods, Section 2.4.5). After these treatments, the PEG precipitation was repeated to further purify and concentrate the phage. This highly concentrated microphage suspension was then subjected to the final fractionation step, preparative native virion agarose electrophoresis.

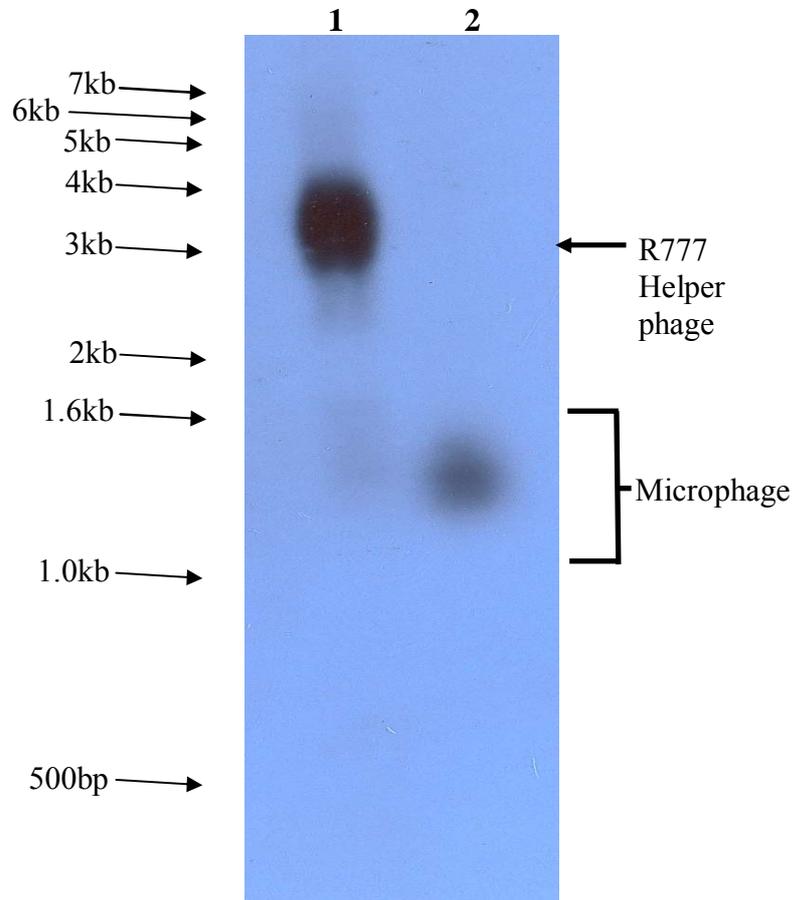


**Figure 31. Southern blot detection of microphage ssDNA from "high PEG" fractions.** Virions were disassembled by SDS, separated by electrophoresis on an agarose gel and probed with microphage origin-derived PCR-generated probe. Lane 1, soluble fraction of "Low PEG" precipitation pellet; 2, insoluble fraction of Low PEG precipitation pellet; 3, low layer of soluble fraction from high PEG precipitation pellet; 4, high layer of soluble fraction from high PEG precipitation pellet; 5, middle layer of soluble fraction from high PEG precipitation pellet; 6, insoluble fraction of High PEG pellet. Please note that the size standard is a 1kb+ dsDNA ladder (Invitrogen) and hence is not suitable for direct comparison of ssDNA sizes. It has only been used here as a "ruler" to monitor the progression of electrophoresis and the position of bands in the gel due to the lack of an appropriate ssDNA marker.

## **5.5 Purification of microphage using native phage gel electrophoresis**

Southern blot analysis of the SDS-disassociated virion gels before the purification with Triton X-100 and Sarkosyl indicated that the samples precipitated in high PEG contained a significant amount of microphage, but also a significant amount long helper phage virions. Specthrie, Bullitt et al., (1992), also noted this and in that publication the helper phage could not be completely eliminated from the microphage preparation. To purify the microphage away from the long helper phage virions, a protocol for preparative native phage agarose gel electrophoresis was developed.

The standard procedure for visualisation of the position of a virion band after native phage agarose gel electrophoresis involves the in situ disassembly the virions within the agarose gel using an alkaline buffer, and then staining the ssDNA with DNA intercalating dye Ethidium Bromide. Without this in situ disassembly of the virions, the ssDNA genome cannot be visualized using Ethidium Bromide. For this experiment it was not possible to use the disassembly procedure to locate the microphage as it a destructive technique. Instead to determine what the position of the microphage virions after native agarose electrophoresis was going to be, a replica of the preparative native gel was run (Figure 32). Using Southern blotting, microphage virions were detected migrating as a smeared band at approximately the same distance as the region between the 1.0 kb and 2.0 kb marker on the 1 kb plus DNA ladder (Invitrogen). The virion bands are always somewhat smeared on the native gels, because these are essentially native protein gels. The 1kb plus ladder was used simply for demarcation of migration distances, this was because there are no appropriate markers available for native virions.



**Figure 32. Detection of the microphage by native virion agarose gel electrophoresis.** Lane 1, phage precipitated by low PEG (2.5%); 2, phage precipitated by high PEG (15%). Arrows indicate the position of the dsDNA fragments of the 1kb<sup>+</sup> ladder (Invitrogen).

The preparative native virion agarose gel electrophoresis was performed under the same conditions, using the 1kb plus ladder for orientation. The appropriate region of the preparative gel was excised from the gel and microphage virions were extracted as described in the Material and Methods (Section 2.4.9).

## 5.6 Analysis of purified microphage

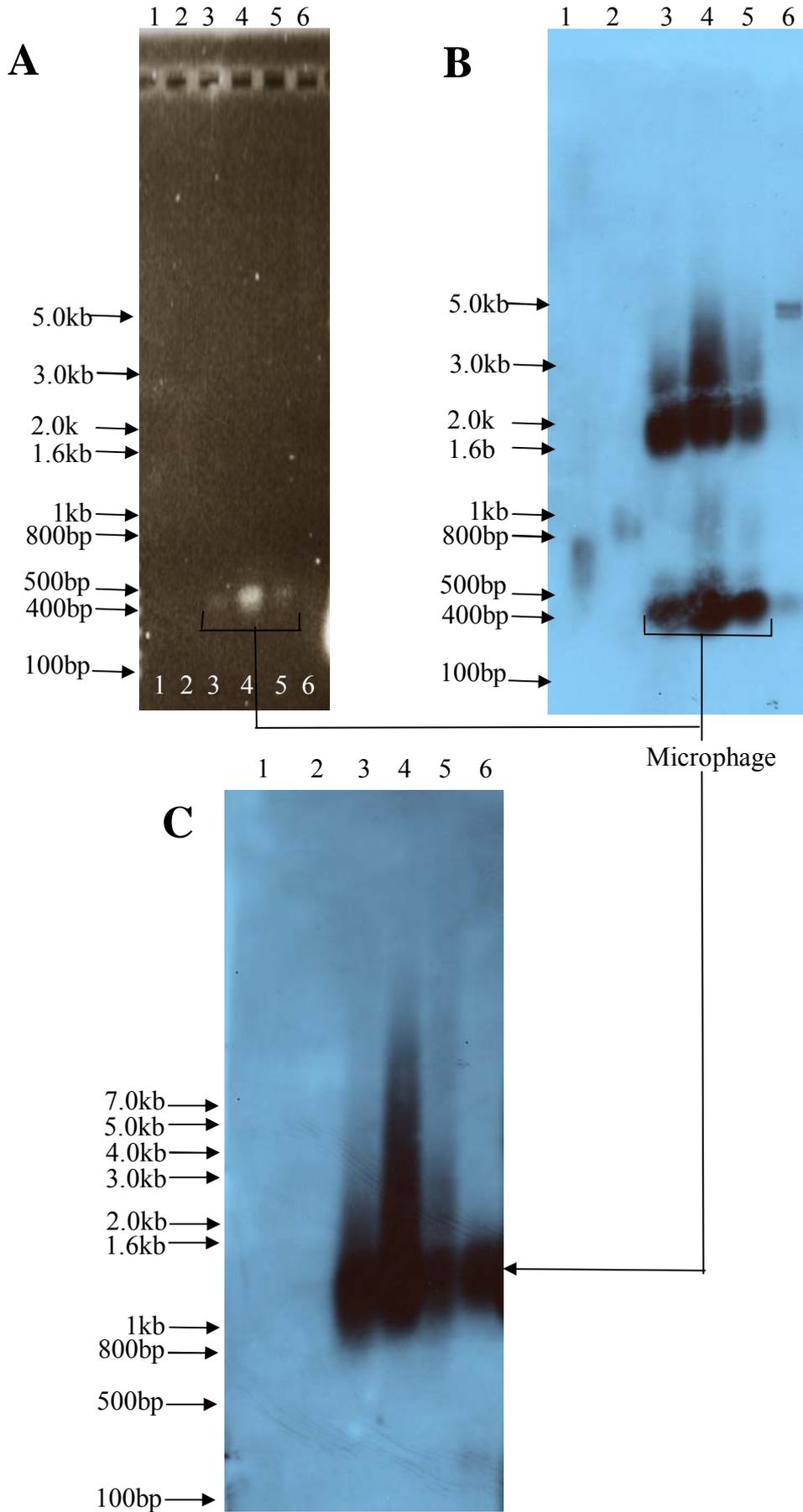
The gel-purified microphage samples were analysed by SDS-disassociated virion gel electrophoresis (Figure 33A and B) and native phage virion gel electrophoresis (Figure 33C). These methods analysed the size of the virion, the purity of the sample and concentration of the purified microphage.

The Southern blot analysis of the SDS-disassociated virion gel shows the final gel-purification step was highly efficient at removing the large helper phage from the sample. The majority of the microphage were recovered from the gel slices excised from the preparative native phage gel slices corresponding to regions between the following 1kb<sup>+</sup> ladder bands; 1.0 kb-1.3 kb, 1.3 kb-1.6 kb and 1.6 kb-2.0 kb (Figure 33A and B).

The increase in concentration and purity also allowed visualisation of the microphage genomes on the SDS-disassociated virion gel using Ethidium Bromide, allowing densitometric analysis of gel bands to determine concentration of microphage (Figure 33A). Densitometric enumeration of the three microphage-positive samples (1.0-1.3 kb, 1.3-1.6 kb and 1.6-2.0 kb) indicated that the concentration of microphage were  $1.2 \times 10^{13}$ ,  $9.57 \times 10^{13}$  and  $1.11 \times 10^{14}$  microphage per mL, respectively..

The native virion electrophoresis gel was used to assess the purity and size of the microphage sample (Figure 33C). Because the microphage genomes shown on the SDS-disassociated virion gel could be derived from poly-microphage (microphage virions containing more than one 200 nt microphage genome) and not from mono-microphage (microphage virions carry only one 200bp microphage genome), the length of the microphage virions was analysed. Southern blot analysis of the native virion gel showed that the major fraction of the microphage samples were of a similar size (Figure 33C).

In addition to this, there was no evidence of large poly-microphage or helper phage virions in any of the microphage samples purified from the gel slices.



**Figure 33. Purity and concentration of microphage** A. Ethidium-Bromide stained gel of SDS-disassociated microphage after preparative agarose gel electrophoresis and purification. B. Same as A, Southern blot detection of phage-derived bands using a probe corresponding to the microphage origin of replication C. Southern blot detection of microphage after the native phage agarose gel electrophoresis. All lanes are loaded with concentrated precipitates of microphage extracted from the preparative native virion agarose electrophoresis 1, 200-800 bp gel slice; 2, 800 bp-1.0 kb gel slice; 3, 1.0-1.3 kb gel slice; 4, 1.3-1.6 kb gel slice; 5, 1.6-2.0 kb gel slice; 6, 2.0-7.0 kb gel slice.

## 5.7 Visualisation of microphage by electron microscopy

Microphage virion samples 1.0-1.3kb, 1.3-1.6 kb, and 1.6kb-2.0kb were sent to James Conway at the University of Pittsburgh for the analysis using electron microscopy. Samples were prepared and visualized using negative staining with uranyl acetate (Figure 34A). Electron micrographs of these samples showed that they contained a high amount of microphage particles. The analysis also showed that the microphage samples were free of contaminating full-length virion particles; however, there was some contamination with ~20-40 nm vesicles.

Microphage were also analysed by electron microscopy using cryo-negative staining (Figure 34B). From this, it can be determined that the microphage virions are approximately 50 nm in length, which is the same length as reported in Specthrie et al., (1992).

The microphage particles also have unequal termini with one being pointy and the other being oval. It has been shown previously that the pIII/pVI cap forms the pointy end (Gray et al., 1979). It can be noted looking at the termini of the microphage that the virions are forming sheets that are composed of individual microphage aligned in alternating orientations.

In addition to this, six double-length poly-microphage particles out of a total of 172 can be viewed in the entire view field of the electron micrographs (Figure 34B, top left panel). These particles are approximately twice the length of the monophage microphage particles; this is consistent with these particles containing two microphage genomes.

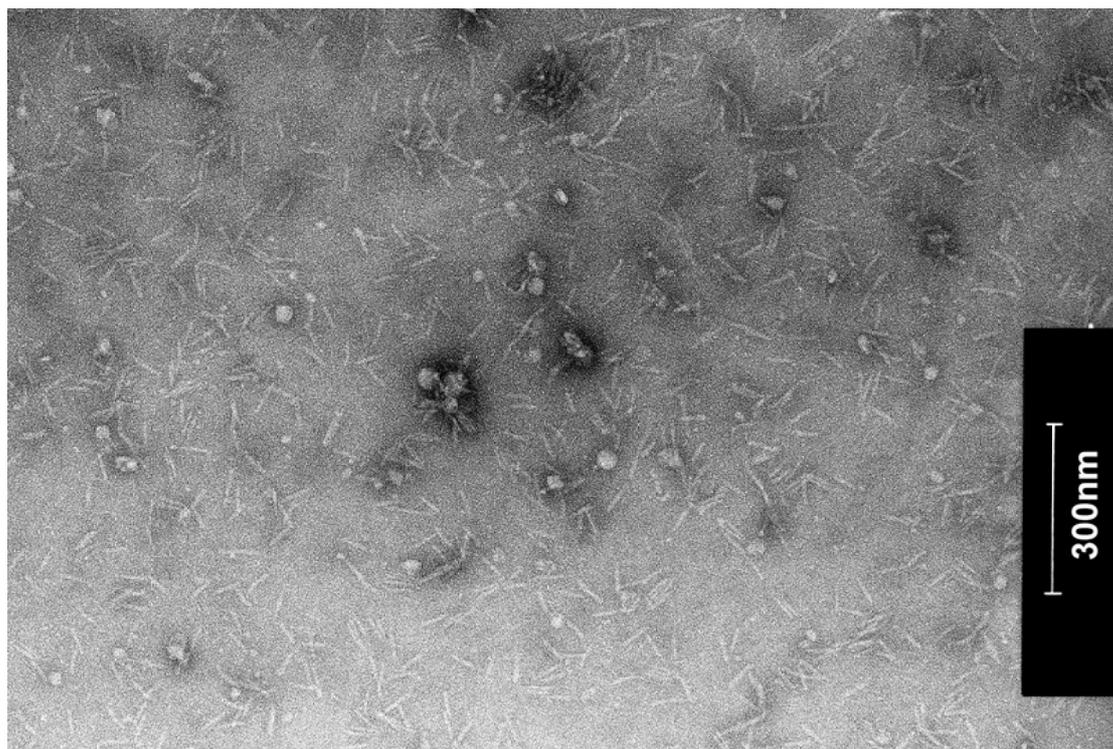
In the cryo-negative electron micrographs it can be observed that the microphage virions appear to be ringed with helical patches of alternating light and dark staining (Figure 35A). The suggested reason for these alternating regions of light and dark stain is due to grooves formed by the shingle-like helical array arrangement of pVIII. An image highlighting these grooves is shown in figure 35B. Based on the structural data available, it can be predicted that the a microphage particle should exhibit between 6-8

alternating regions of dark and light staining, depending on particle orientation, if the staining was due to this groove formation.

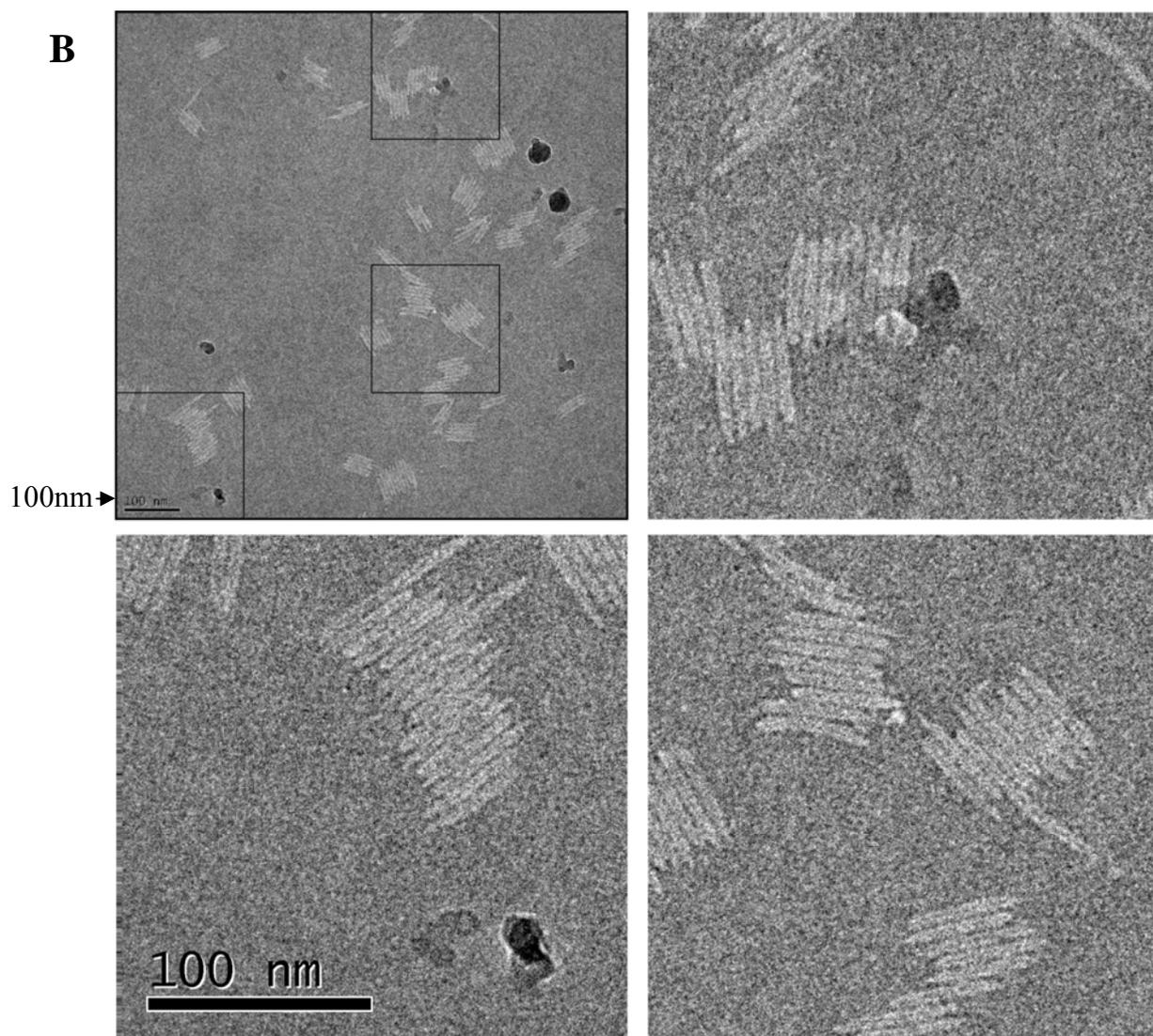
It is noted that no microphage visualized in these electron micrographs gird have any signs of the extra ball of density that could correspond to free-moving N1N2 domains as described in Gray, Brown et al., (1979). This is not unusual, and in fact, it is much more common that fl virions do not show the N1N2 "ball" than they do (Rakonjac, 2006).

The results of the analysis of the various electrophoresis gels and the electron micrographs show that an efficient microphage-producing system and purification protocol has now been established.

**A**

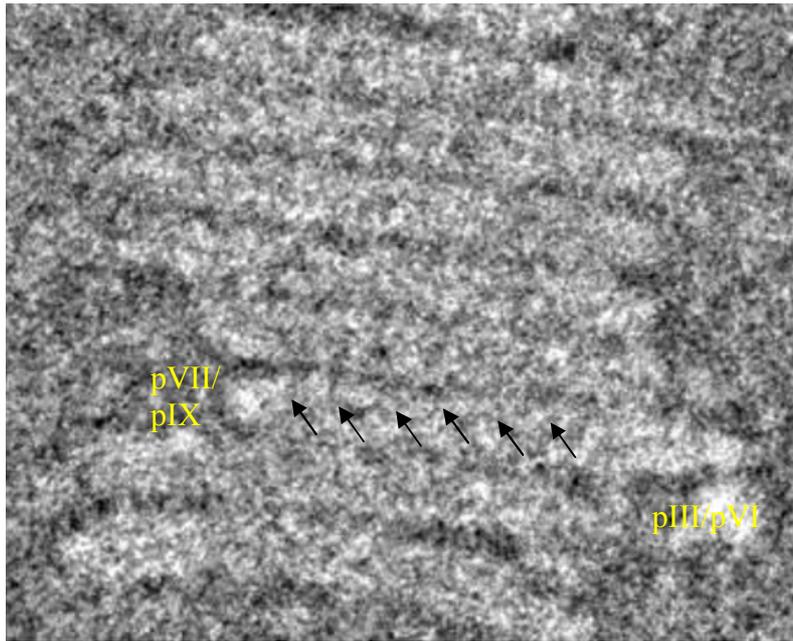


**B**

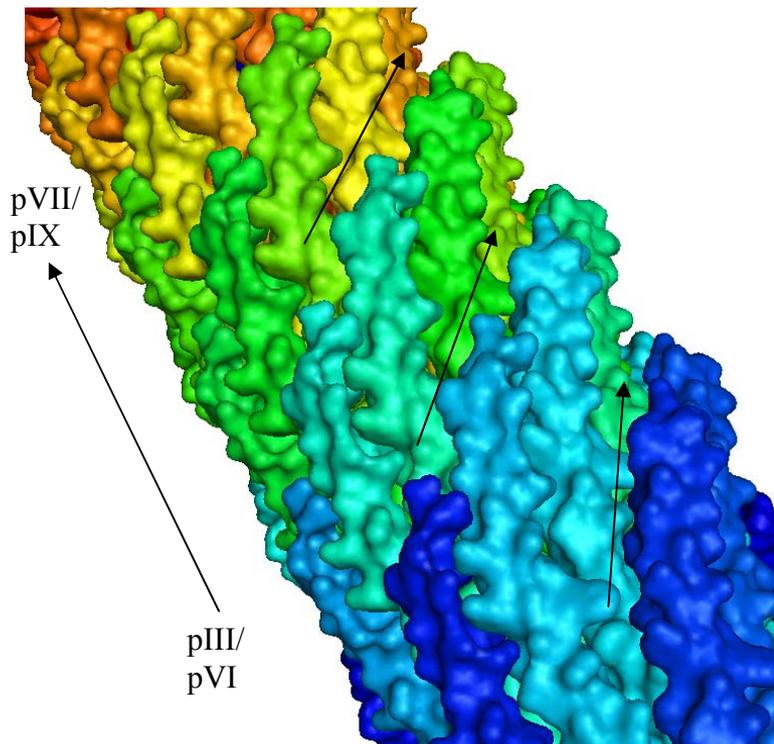


**Figure 34. Electron micrographs of gel-purified microphage samples** **A.** Negatively stained electron micrograph image of gel-purified microphage sample from the 1.0-1.3kb gel slice. The sample was negatively stained using Uranyl acetate. Bar corresponds to 300nm. **B.** A cryo-negative electron micrograph of purified microphage from 1.3-1.6 kb gel slice. Images courtesy of Dr James Conway, University of Pittsburgh.

A.



B.



**Figure 35. High-resolution TEM and 3D atomic projection of phage filament. A.** Ultramagnified Cryo-negative TEM image of microphage. Arrows show alternating regions of light and dark staining. **B.** 3D projection of phage filament. The phage filament is orientated so that the pVII/pIX terminus is pointing to the upper left and angled away from the viewer. Arrows highlight the grooves in filament caused by pVIII packing. This image was obtained using PyMOL from file 2c0w downloaded from the RCSB PDB database. This file is based on the work presented in Marvin D.A et al (2006).

## 5.8 Future uses of microphage production system for structural studies

Microphage will be used as an important tool for future structural studies of the filamentous phage termini. Currently the three-dimensional structure of both phage termini is unknown. Deciphering the structure of the phage termini would greatly contribute to the current understanding of the process of filamentous phage infection and phage assembly. Given that Ff filamentous phage are used in phage display, understanding the phage termini structure may increase the understanding of the limits of this technique.

The major limiting factor in gathering structural information about the filamentous phage termini from wild-type virions is the high length to width ratio, and the inherent flexibility of the virions. Microphage however are much more suitable for structural analysis, due to their much reduced length to width ratio, which in turn reduces the maximum flexibility of the filament.

The original plan for using microphage for structural analysis of the phage termini was to use Cryo-EM and single particle reconstruction to produce a low resolution structure of the entire microphage. This would yield the first low resolution structure of the filamentous phage termini. By using the series of C-domain deletion mutants outlined in Rakonjac et al (1999) assembled into the microphage virions it would be then possible to map the portions of the C-domain corresponding to specific regions of the termini structure.

Microphage may also be used, for X-ray crystallographic studies of the virion filament. In the original filamentous phage major coat protein papers using X-ray filament-diffraction studies of magnetically aligned and desiccated native virions, only the major repeating unit, which is the major coat protein pVIII, could be visualised. This is due to the phage termini being only a very minor component of the X-ray diffraction data. Because X-ray fibre diffraction methodologies produce an averaged structure, the structure of the phage termini could not be resolved from this form of data. This has also been the case with further studies of the filamentous phage structure conducted by NMR and Cryo-EM.

Microphage have a much reduced ratio of pVIII:pIII/pVI. In native virions the ratio of pVIII:pIII/pVI is 540:1, whereas in the microphage virions this ratio is reduced to 19:1. In addition, because microphage virions are shorter and less flexible than native virions it may be possible to crystallise microphage virions. If the microphage do crystallise and diffract, the analysis should provide superior structural data to previous studies and may provide the first high-resolution structure of the filamentous phage termini.

In addition to X-ray crystallography the microphage could also be used in other structural techniques. Using deuterium exchange mass spectroscopy (DXMS) it could be possible to map where the interaction between the N1N2 domains and the phage filament. In Gray et al (1979) the N1N2 domain of pIII were visualised using TEM under specific conditions. It can be noted from this paper that the release of the N1 and N2 domains can only be observed occasionally. This means that the N1N2 domains of pIII must spend the majority of the time in some form of complex with other proteins that form the phage cap. If this complex of N1N2 and the phage filament is a consistent complex then it should be possible to map its position using DXMS. By analysing the change in deuteration exchange rates between microphages with and without N1N2 domains this interaction could be mapped. For these experiments it should be possible to remove the N1N2 domain with subtilisin digestion of microphage virions. Subtilisin digestion of filamentous phage virions has been shown to remove the N1N2 domains of pIII only (Gray, Brown & Marvin, 1981).

Another option exists for the structural analysis of the phage termini. This method does not rely on the use of microphage. In Galius et al (1994) it was shown that pIII and pVI formed a structural complex strong enough to withstand removal from the phage using the detergent deoxycholate and the solvent chloroform. It is possible that this complex could be purified and analysed by Cryo-EM and single particle reconstruction.

## **5.9 Discussion of microphage production system**

In this chapter, a new microphage-producing plasmid and a new helper phage were constructed to aid the creation of an efficient system for the production of microphage.

In addition to this, a new microphage purification procedure was designed. This new purification protocol included a preparative gel purification step, which purifies the microphage virion away from contaminating helper phage. Together, these improvements resulted in a high yield system for producing pure microphage virions. The stocks of microphage produced using this system have been shown to have very low contamination from the helper phage and other contaminants.

This research also showed the significance of pVIII in the regulation of filamentous phage length. In normal filamentous phage assembly, the length of a phage virion is solely dependent on the length of the ssDNA genome available for packing. In contrast, at the extreme minimal length of the microphage genome it is not genome length which is the major determinant of virion particle length. Instead, it seems that particle length is more dependent on the local concentration of the phage coat proteins at the phage assembly site. In the case of the microphage, the concentration of the major coat protein pVIII seems to be most important. This is because when WT levels of pVIII were supplied microphage genome were found to be co-packaged with helper phage genomes, whereas when pVIII amber mutant helper phage R777 was used in conjunction with the microphage plasmid pNJB7, microphage genomes were packaged independently of the helper phage genome.

This research was significant as it took a previously inefficient microphage production and purification system (Speckthre et al., 1992) and modified it to produce a system that produced a much greater yield of highly purified microphage virions. This is a very important and critical step if microphage virions are to be used for structural studies. Given that this has been achieved, microphage virions can now be purified for a range of various structural analysis which hopefully will yield to the first structural analysis of the phage termini.

## Chapter 6

### Discussion

#### 6.1 Introduction

This thesis is composed of two different lines of investigation. The first project discussed will be the structure-function analysis of the role of the C-domain of pIII in phage infection. The second project discussed the optimisation of a microphage producing plasmid based system and the development of a method for purification of microphage virions away from the helper phage virions.

#### 6.2 A new model for the involvement of the C-domain of pIII in the process of phage infection

The work presented in Chapter 3 mapping of the minimal requirement of the C-domain for phage infection showed that a fragment containing the final C-terminal 121 residues is sufficient for entry of phage DNA in the host cell. A shorter C-terminal fragment containing 93 residues is sufficient for phage release but not for infection. The sequence between infection-competent and release-competent C-terminal fragments was designated the "infection-competence sequence" (ICS): (121-GKLDV VTDY GAAID GFIGD VSGLA NGNG-93). This finding shows that the process of phage infection is not symmetrical to the process of phage termination and release.

This functional asymmetry is consistent with the involvement of two different sets of proteins in the two processes. Phage assembly termination/release is mediated by the phage-encoded *trans* envelope complex of pI/pXI/pIV and host cell protein thioredoxin. In contrast, the process phage infection requires the F pilus and the TolQRA complex. Because the co-factors are different, it is likely the structural rearrangements and contacts involved in "locking" the phage structure during assembly termination and "unlocking" the phage structure during infection is different.

Bioinformatics analysis of the pIII C-domain in the region of the ICS using PONDR (Garner et al., 1999; Li et al., 1999; Romero et al., 1997; Romero et al., 2001) and

JPRED (Cole et al., 2008) showed that it is most likely that the ICS region of pIII has a defined secondary structure, a predicted  $\alpha$ -helix (Figure 27). Analysis also showed that the ICS region was flanked by predicted region of disordered segments. This arrangement, in which a folded segment is embedded in a disordered domain of a protein, is often an indication of the folded segment forming contacts with other proteins (Dyson & Wright, 2005).

Based on this new experimental evidence, it is possible to propose three different models of the role that the identified ICS region of pIII may play in the process of phage infection. Model one is that the ICS region is required to form contacts with a host cell-binding partner during infection, to mediate the reintegration of the phage virion coat into the host cellular membrane. Model two is that the ICS region is required for transmitting the "unlock" signal to the other subunits of the phage virion cap complex during the process of phage infection. Model three is that the ICS region is required for both the formation of contacts with a host cell protein during infection and to transmit the "unlock" signal within the phage virion cap. The two latter models account for the multi-subunit structure of the virion cap which contains five copies of pIII, presumably organised in five-fold radial symmetry arrangement.

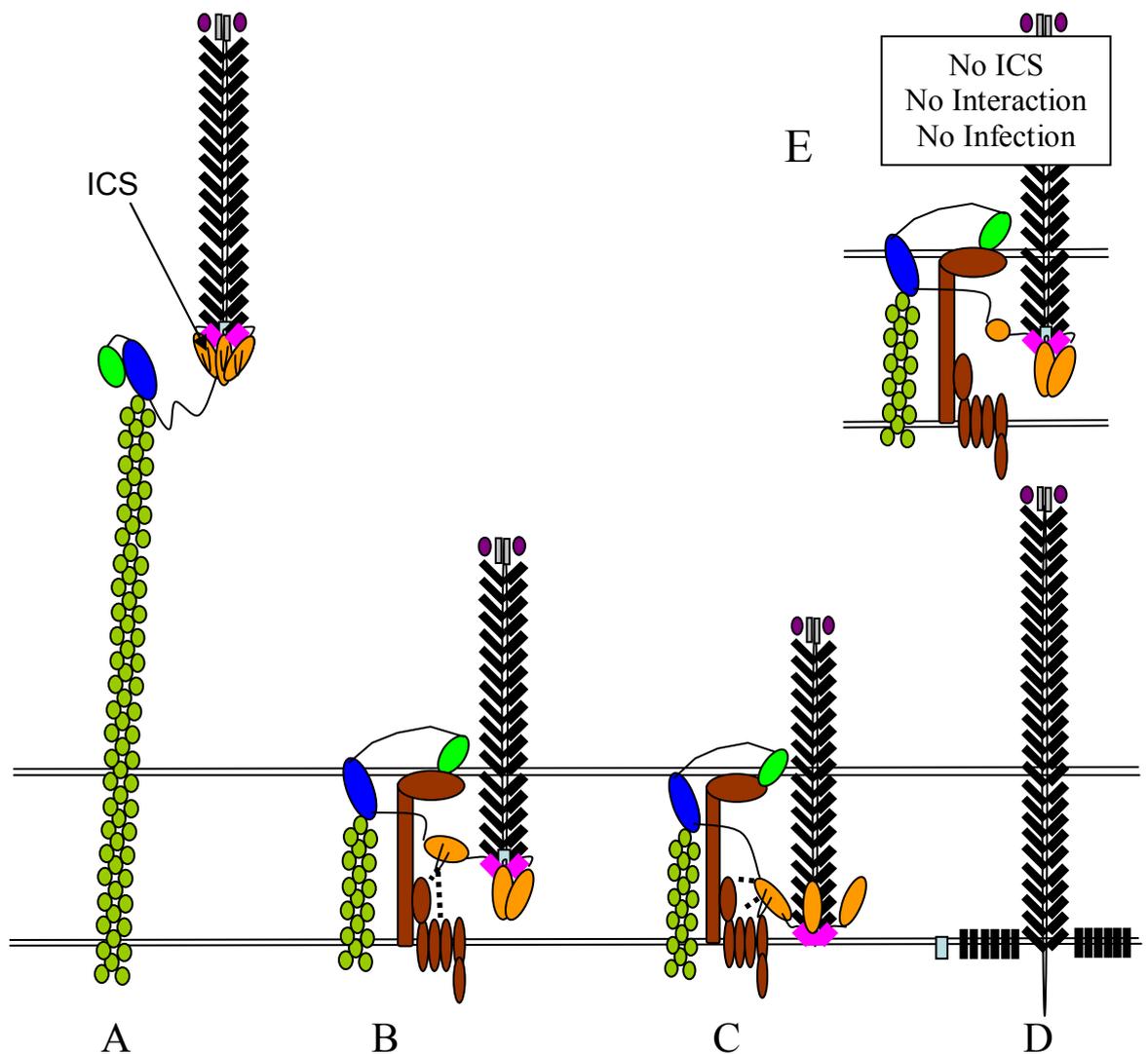
However, models two and three were immediately discounted based on the results of the experiment outlined in Chapter 4 that showed that the ICS region of the pIII C-domain was not involved in inter-subunit signalling within the cap complex. In this experiment a series of pIII C-domain fragments, that matched the deletion series used in this thesis, but were not attached covalently to the N domains, were assembled into phagemid particles together with a full length WTpIII (WTpIII/C83-WTpIII-C141). If the ICS was responsible for transmitting the signal to the other subunits of the phage cap complex, conformational changes within the C-domain that had been triggered by binding of N1N2 of the WTpIII subunit to the phage receptors, would not be able to propagate to other subunits and mutant virions that contain shortened C-domain fragments, lacking the ICS region, would be non-infectious (Figure 29 E).

Infectivity of these latter virions (WTpIII/C83-WTpIII/C111) was measured and was shown to be as high as that of the virions containing only the WTpIII, or those with the C-terminal fragments that contain the ICS (WTpIII/C121-WTpIII/C141). Therefore, it

appears the intra-complex communication, via the ICS, is not required during the process of phage infection.

Hence, it is hypothesised that the ICS is likely to form contacts with a host cell binding partner during the process of phage infection. The model for the involvement of the ICS region in infection proposed is model one (Figure 36A-D). The phage virion binds to the tip of the F pilus via the N2 domain of pIII, this causes the release of the N1 domain of pIII from N2 exposing the TolAIII domain binding site (Reichmann & Holliger, 1997). The F pilus then retracts bringing the N1 domain of pIII into contact with the TolQRA complex, and interactions are formed between the TolAIII domain and the N1 domain of pIII. Interaction of N1 with TolA domain exposes the ICS region of the pIII C-domain, allowing further contacts to form with the TolQRA complex or another unidentified host cell binding partner. This causes the "unlocking" of the phage virion allowing the insertion of the phage virion into the membrane and the transportation of the phage ssDNA into the host cell cytoplasm.

In the virions lacking an ICS region linked to the N1N2 domain (Ndc93-Ndc83) interactions with a host cell-binding partner during the process of phage infection are no longer formed, blocking infection (Figure 36E). In the mutant with intermediate infectivity (Ndc111), the infectivity was likely reduced because the shortened ICS region was structurally unstable, and could not efficiently form the structure required for the formation of interactions with the host cellular binding partner. The other possibility was that most of the residues required for forming contacts with the host cell-binding partner were missing in the shortened ICS region, thus weakening the interaction and reducing infectivity by a factor of 100.



**Figure 36. Proposed model one for f1 entry** **A.** Phage virion binds to F pilus via the pIII N2 domain causing F pilus to retract and N1 domain to be released **B.** N1 domain binds to TolQRA complex via TolAIII domain. This signals the pIII region to undergo a conformational change exposing the ICS region of the pIII C-domain. The ICS forms further interaction with TolQRA (or another unidentified binding partner). **C.** Phage virion "unlocks" and is inserted into the cytoplasmic membrane using the pIII transmembrane helix as a fusion peptide. **D.** Phage virion integrates into cytoplasmic membrane as ssDNA genome is transported into host cell cytoplasm. **E.** Situation when mutant NdC93-NdC83 was used. Total deletion of ICS region of pIII C-domain, No interactions with TolQRA are formed, and infection does not occur.

The best method for conclusively demonstrating this model of infection would be to obtain a high-resolution three-dimensional structure of the phage cap complex in its native closed conformation on the phage virion and in its open conformation within the periplasmic space after interaction with TolA, then attempt to model the events that occur during phage infection.

Without a three-dimensional protein structure, a genetic approach could be designed to examine whether the interacting partners of the ICS are proteins of the TolQRA complex.

### **6.3 Future experiments to evaluate the role of the pIII C-domain in phage infection**

Without a high-resolution structure of the pIII C-domain, a genetic approach could be used to attempt to further dissect the role of the pIII C-domain during infection. To achieve this, the following experiments are proposed. Firstly, scanning mutagenesis of the ICS must be completed to find an amino acid residue within the ICS region that plays an important role during infection. If this analysis fails to implicate a particular residue of the ICS in infection, then it is possible that infection may be mediated by an interaction with the main chain atoms of the ICS or by multiple interactions formed by a binding surface. If this is the case then mutagenesis of the ICS to bulkier amino acids or amino acids carrying opposite charge may be required to disrupt interactions.

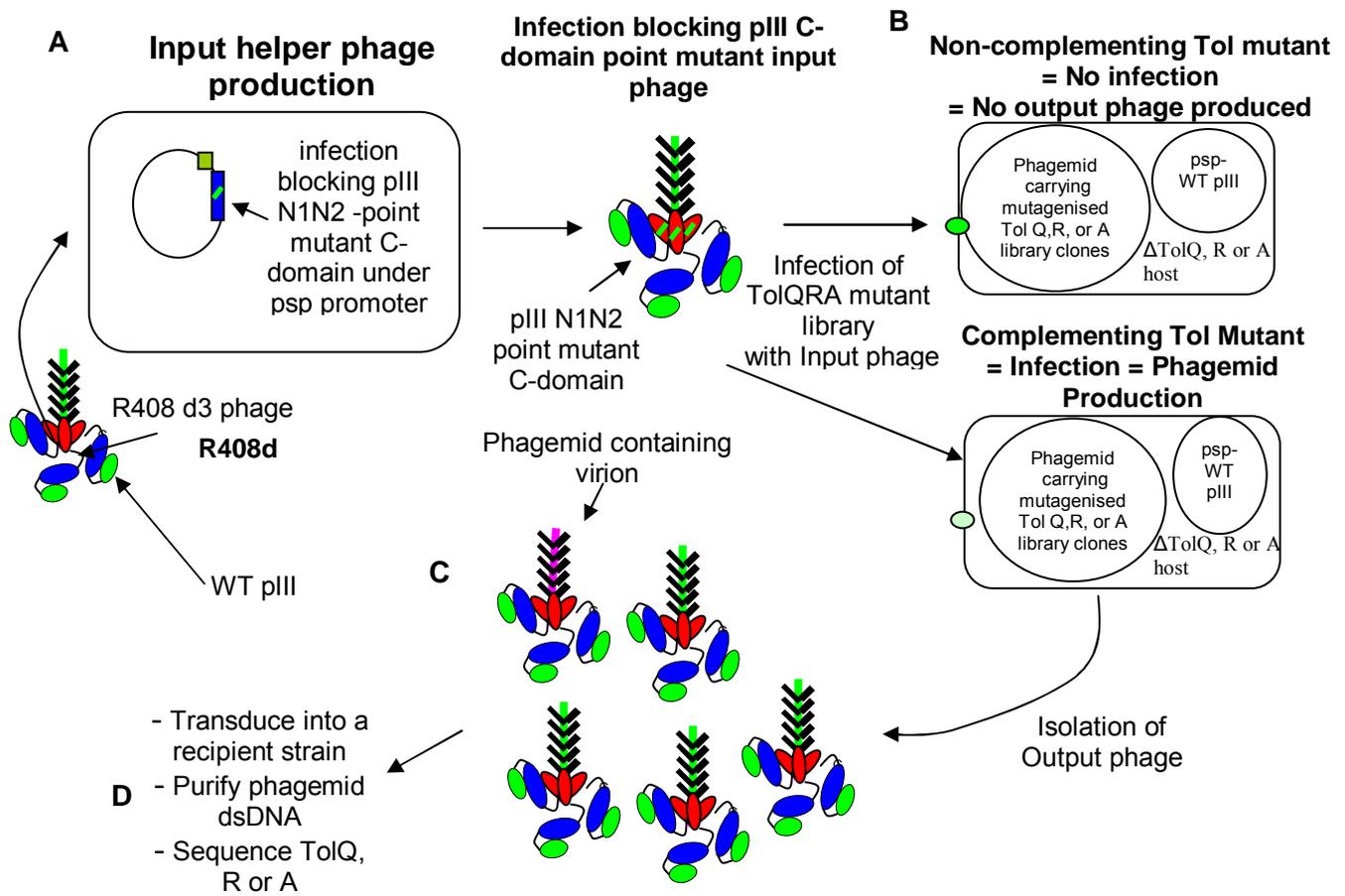
A second possible method to obtain C-domain point mutant that repress infection may be to use saturation mutagenesis of the pIII C-domain. This could be done by subjecting the C-domain to error-prone PCR to create a library of pIII C-domain mutants. This library would then be fused to the WT N1N2 domains and inserted into a phagemid expression vector, thus creating a library of C-domain mutants. By using the mass infection of cloned cells with a helper phage followed by plating, non-infective mutants could be isolated by scanning for infected colonies that fail to produce infective phagemid particles. This could be achieved using a filter based replica-plating technique. The non-infective phagemid could then be isolated and sequenced to identify the inactivating mutations within the pIII C-domain.

If mutations are found in the ICS, or elsewhere in the pIII C-domain, that block infection, a more accurate map of infection-mediating residues could be made.

Once residues essential for infection are identified then one of the mutants could be used to find intergenic suppressor mutations, and thereby be used to identify the host interacting partners of the C-domain during infection. It is the suggestion of this thesis that the most likely candidate for a cellular interaction partner is the TolQRA complex. This is because even though only TolA is required for the binding of N1 domain during infection, the host cells also become resistant to infection if either TolQ or TolR are deleted (Click & Webster, 1998; Sun & Webster, 1987). However, TolQ and R may be required to complex with TolA for the TolAIII domain to be available for phage infection. If this is the case then the deletion of either TolQ or R simply prevents complex formation making TolA unavailable for infection.

If TolQ, R or A are involved in forming a complex with the pIII C-domain during phage infection then an intergenic suppressor mutant may be isolated by using a genetic screen (Figure 37). Phage carrying an infection-blocking C-domain mutant will be produced in a growth experiment. These phage will then be used to screen a phagemid library of Tol Q, R or A mutants. The TolQ, R and A mutant library will be produced either by error-prone PCR amplification (Cadwell & Joyce, 1992) of the *tolQ*, *R* or *A* genes followed by cloning into a phagemid expression vector or by passaging the phagemid containing either *tolQ*, *R* or *A* through a mutator strain like XL1-Red (Greener & Callahan, 1994). The library of *tolQ*, *R* and *A* mutants will then be transformed into host strains that lacking the corresponding gene (obtained from the Keio collection (Baba et al., 2006; Datsenko & Wanner, 2000) that will also be expressing WTpIII from a compatible vector. If a mutation in *tol Q*, *R* or *A* suppresses the C-domain mutation of the helper phage stock, the infection of the host cell will occur and phagemid particles will be produced. The phagemid particles produced will contain WTpIII, expressed from plasmid containing *gIII* in the infected host cells and will also contain the corresponding *tol Q*, *R* or *A* mutant sequence contained on the phagemid genome. Thus, phagemid produced will be infective to standard host cells and will also carry the sequence of the intergenic TolQRA suppressor mutations that suppress the pIII C-domain mutant in the original helper phage stock. The phagemid particles produced then can be transduced

into a recipient strain and the suppressor mutants can then be sequenced. Because the initial helper phage infection will only occur if the recipient cells express a suppressor mutant of TolQRA, phagemid are going to be produced only in those cells that produce the correct suppressor mutation. If a cell does not produce a Tol Q, R or A mutant that suppresses phage infection then it will not be infected with the helper phage and will not be able to produce phagemid particles.



**Figure 37. Proposed genetic screen for identification of intergenic suppressor mutations in TolQRA.** **A.** Production of helper phage stock displaying an infection-incompetent pIII C point mutation. Cells carrying plasmid producing infection incompetent pIII C-domain mutant are infected with R408d3 helper phage, to produce the suppressor-screening helper phage stock. **B.** Scanning TolQRA mutant library. Host cells each contain a phagemid vector expressing mutagenized TolQ, R or A. This library is infected with the helper phage carrying infection-incompetent pIII C-domain mutant. If a TolQ, R or A library clone contains an intergenic suppressor of the pIII C non-infective mutation then it will compensate for pIII C-domain infection defect of the helper phage and hence become infected, producing phagemid particles that encapsulate suppressor mutant of Tol Q, R or A. If library does not contain intergenic suppressor mutation then no infection will occur and no phagemid particles will be produced. **C.** Phagemid particles produced. Suppressor mutant TolQ, R or A is encoded by the phagemid particle. **D.** Analysis of phagemid particles by sequencing of Tol Q, R or A, to identify the suppressor mutation.

## **6.4 Construction of a microphage variant of filamentous phage for structural studies**

In this thesis, an experimental system was outlined that produced a microphage variant of filamentous phage virions. In addition, a purification scheme to separate the microphage virions away from helper phage virions and other contaminants was established. Together, these two elements have allowed the production and purification of a high titre stock of microphage virions which were imaged using TEM and may be used in the future for structural studies of the filamentous phage termini.

## **6.5 Effect of pVIII on phage virion length**

While developing this system, some observations relevant to the control of virion length were made. The f1 phage, virion particle length is thought to be controlled by the size of the phage genome that is to be packaged. However, in the production of the microphage virions this was shown to not be the only factor controlling virion length. In the experiment using R408 as a helper phage for microphage production, the microphage ssDNA genomes were preferentially packaged particles that either contained multiple copies of the microphage genome packaged into one virion or contained co-packaged microphage/helper phage genomes. Our hypothesis is that this was primarily occurring due to the high concentration of the major phage coat protein pVIII at the site of phage assembly, favouring virion elongation over the termination and release of microphage. This hypothesis is based on the original microphage paper by L. Specthrie (1992). In this publication, it was found that only helper phage R474, which contains a promoter mutation that reduces *gVIII* expression to ~10% of WT levels, produced detectable quantity of microphage.

In this thesis, virions released during infection with two helper phage (R408 and R777 (R408 *gVIII<sup>am</sup>*), producing different amounts of pVIII, were compared. When helper phage R408 was used to induce microphage production and the virions produced were analysed it was found that the microphage genomes precipitated in the low PEG fraction that normally precipitates long helper virions. Furthermore, no microphage virions were

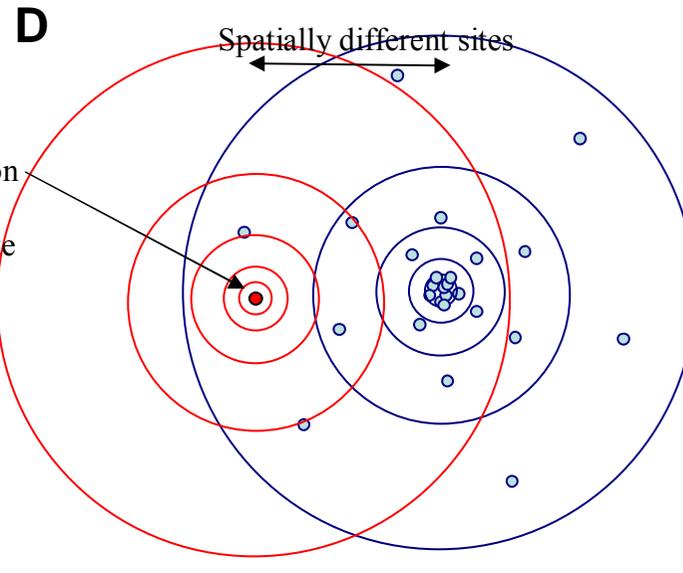
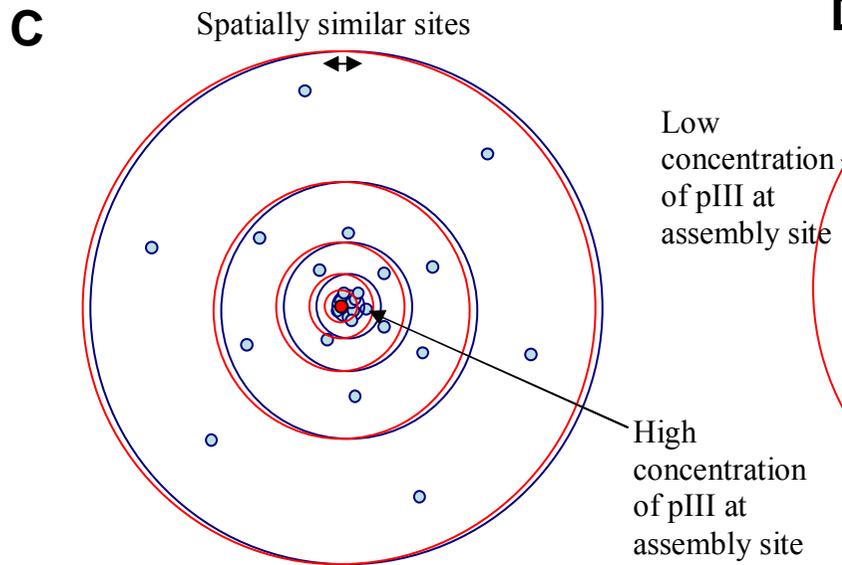
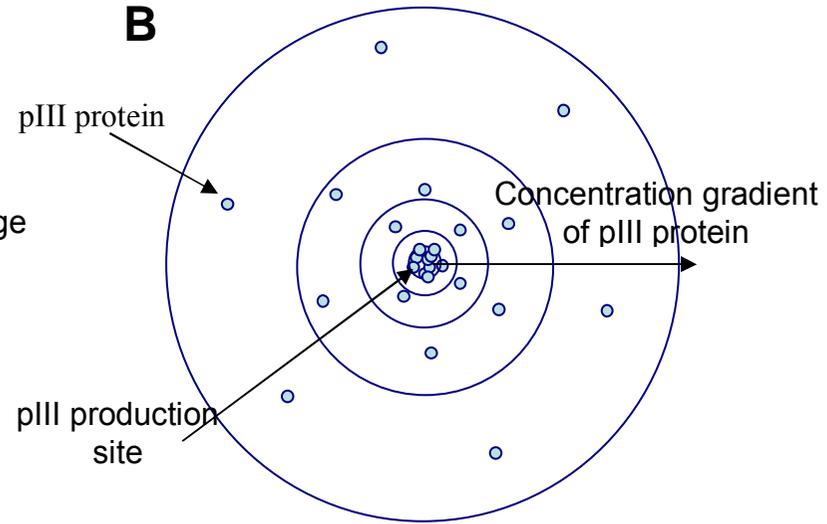
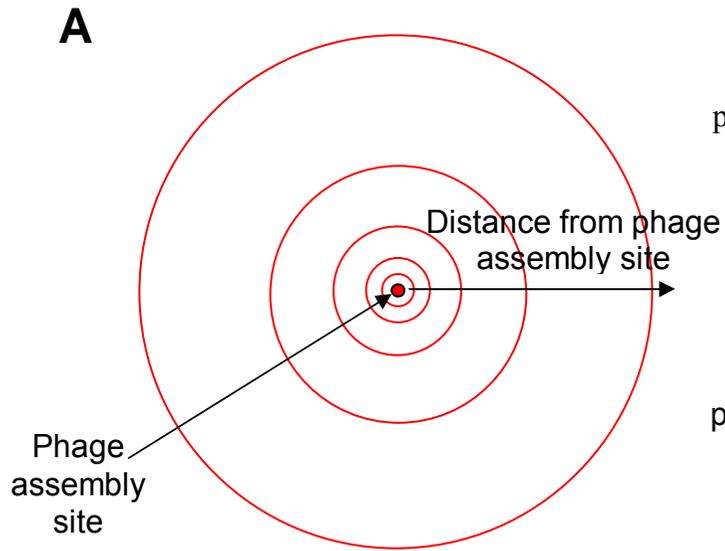
detected by native gel electrophoresis, suggesting that the microphage genomes were co-packaged with the helper phage in the same virions.

Helper phage R777 (R408 *gVIII<sup>am</sup>*) was then used to decrease the amount of pVIII available for phage assembly. In a suppressor strain (K1030, *supD*) the *gVIII<sup>am</sup>* mutant was translated at a level which is 10% of the WT pVIII transcription level of R408. When this *supD* suppressor strain carrying the microphage-producing plasmid was infected with this new helper phage, the majority of microphage genomes were packaged as individual virion particles and not as polymicrophage particles. Because of this, it can be concluded that in extreme situations of genome length it is not just a function of genome length that determines virion length. Instead, it is the rate of virion termination vs virion elongation that determines the virion length. The key determinant of this rate is the localised concentration of the phage virion coat proteins at the site of phage assembly. Thus if the length of the genome is reduced drastically, such as in the case with the microphage, phage genomes may become co-packaged to help alleviate the increasing concentration of phage pVIII within the host cell membrane, given the low abundance of termination proteins in the cell.

The reason why phage virion length can be dependent on localised concentrations of phage virion proteins and not overall cellular concentration is because of observations when growing phage that do not encode the minor virion coat protein pIII.

When the pIII is expressed *in-trans* to complement a pIII deletion mutant helper phage the average virion length increases though the increased production of polyphage. In normal WT fl culture, 5% of phage genomes are packaged into polyphage particles containing multiple genomes. This is in contrast to  $\Delta gIII$  phage where a major portion (>50%) of genomes are packaged into polyphage particles. This is likely due to the reduction in local concentration of pIII at the virion assembly site. In fl WT virion assembly, transcription, translation and membrane insertion are physically linked through mRNA and the SecAB-SecYEG membrane targeting apparatus (Chang, Model & Blobel, 1978). Virion coat proteins in WT fl are being expressed at sites spatially close to the sites of virion assembly. Therefore, the site of phage assembly there is a highly-localised concentration of coat proteins compared to anywhere else in the host cell membrane.

In  $\Delta gIII$  phage assembly a plasmid-based expression system is used to supply *gIII* *in-trans*, thus *gIII* expression is no longer physically linked to virion assembly site. Newly produced pIII proteins are therefore not inserted into the membrane close to the site of phage assembly and must diffuse through the membrane to the virion assembly site. Consequently, this displaced pIII targeting relative to the site of phage assembly, causes phage termination efficiency is reduced, thus causing production of mostly multiple-length virions (Figure 38). This leads to the conclusion that it is likely that there is no system in the inner membrane phage assembly complex (pI/pXI) that prevents a new phage genome entering the assembly complex once the previous genome is fully coated with pVIII. Thus if the minor coat proteins required for terminating a phage virion (pIII/pVI) are not available for termination, because the local concentration of minor coat protein is too low, it is possible for another genome to enter the assembly complex and for phage elongation to continue.



**Figure 38. Template-dependent spatial relations between pVIII and pIII concentration gradients in the host cells.** **A.** Top view of a phage assembly site within the cytoplasmic membrane of a host cell. Red dot in centre of the red rings is the site of phage assembly (pI/pXI/pIV complex); the red rings represent increasing distances from the phage assembly site in 2D space. **B.** A site of expression of pIII. In the centre of rings phage coat protein pIII is being transcribed, translated and inserted into the inner membrane. As pIII is inserted into the membrane it can begin to diffuse away from the insertion site. This site specific expression results in a concentration gradient of pIII, with a high concentration of pIII at the site of expression and reduced concentrations of pIII the further you get away from the phage expression site. **C.** WT fl phage assembly. Phage assembly and production of pIII are spatially linked. pIII is being expressed and inserted into the membrane close to phage assembly site, resulting in a high local concentration of pIII at the site of assembly. **D.** Phage assembly when pIII is supplied *in-trans* from a plasmid. Phage assembly is linked to other coat proteins, but not to pIII, which is expressed from a different replicon. pIII is being expressed independently of the phage assembly site and thus pIII has to diffuse to phage assembly site. This results in a lower localised concentration of pIII at the phage assembly site. Under low pIII concentration at the assembly site elongation by adding a new genome to growing filament is favoured over termination and release, hence polyphage particles are the predominant products of phage assembly.

In the WT fl life cycle a large amount of pVIII is produced by the host cell. This is to account for ~2700 copies required to form a single virion, and is sufficiently high for a cell to produce 1000 virion > per hour. Thus, the level of pVIII and other coat proteins is tuned for the assembly of full-length virions. This means that for microphage particles, which are 1/30th of the length of full length particles, the proteins required to terminate virion assembly are not present in sufficient levels, in comparison to pVIII, to allow efficient termination. Thus, to alleviate the host cell of the excess quantities of pVIII in comparison to the minor coat proteins, the microphage genomes become co-assembled into virions containing sequentially packaged helper phage and microphage genomes.

This effect most likely due to ~1/30 the amount of pVIII monomer required to form a microphage virion compared to a normal length virion. A full-length monophage virion contains ~2700 copies of pVIII, while a microphage particle contains 96 copies of pVIII, based on measurement made in Day L (1973). Therefore, to remove the same amount of pVIII from the membrane as one WT helper phage, around 28 microphage genomes would have to be packaged. In addition to this, 140 extra copies of each of the minor coat proteins, pVII, pIX, pIII and pVI, would also be required to complete these virions. Meaning that every time a microphage virion is produced, the amount of pVIII in the membrane increases relative to the amount of minor coat proteins. Therefore, phage elongation becomes increasingly favoured over phage termination, causing the production of poly-microphage and microphage/helper phage virion particles instead of mono-microphage particles.

It is possible that microphage virion production could be stimulated by increasing the level of expression of the other minor coat proteins. However, this is unlikely to succeed because the overexpression of pIII can cause it to form channels in the inner membrane (Glaser-Wuttke et al., 1989) and because the expression *in-trans* would still result in a low local concentration at the site of assembly.

## **6.6 Future modification of the microphage production system to enhance structural studies**

The current microphage-producing system can be modified to produce microphage that are better suited for structural analysis. The current microphage producing system releases virions that contain WT pVIII. WT pVIII is not suitable for structural analysis because, within one phage particle, it adopts a variety of conformations (Tan et al., 1999; Wang et al., 2006). However, if the 21st residue of the mature pVIII polypeptide is modified from a tyrosine to a methionine (Y21M) the conformation heterogeneity of pVIII is greatly reduced to only two possible conformations. Moreover, all pVIII within a phage virion will adopt only one of these two conformations (Tan et al., 1999; Wang et al., 2006). This reduces the structural heterogeneity of the virions and producing virions more suitable for structural analysis. A PCR construct containing the Y21M mutation and the amber mutation required for microphage production has been produced and has been subcloned. However, at the time of writing this thesis the mutant fragment has not been successfully cloned into phage R408.

## **6.7 Modification of the microphage producing system to facilitate purification**

Another change to the microphage-producing system that would facilitate purification is the use of a helper plasmid instead of a helper phage. A helper plasmid has a major advantage over a helper phage as it cannot be packaged into new virions, due to a lack of an *ori* origin of replication and a packaging signal. A helper plasmid was first described by Chasteen, Ayriss et al., (2006). It encodes all phage proteins required for replication and packaging of the microphage genome, as well as a plasmid origin of replication and an antibiotic resistance marker. The advantage of a helper-plasmid based microphage production system over a helper phage-based system is that long helper phage encapsidating virions would not be produced at all. Thus, purification of microphage virions away from longer helper phage virions would not be required. Microphage produced using this system could be directly precipitated with high PEG, or concentrated using tangential flow filtration (TIFF) and, if required, followed by a CsCl<sub>2</sub> gradient centrifugation to remove concentrated cellular debris.

The helper plasmid constructed in our lab, is incompatible with the current microphage producing plasmid; hence, construction of a new microphage producing plasmid is required, to test the helper plasmid for the production of microphage.

## **6.8 Comparison of f1 entry to other bacteriophage and viruses**

The pIII C-domain is required for both membrane release (exit) of the phage virion during phage extrusion and membrane insertion (entry) during infection (Bennett, 2005; Rakonjac et al., 1999; Rakonjac & Model, 1998). To my knowledge, this is the first time that a viral protein has been shown to be involved in both membrane exit and membrane entry. In the majority of known bacteriophage, virion exit is achieved by cellular lysis using a phage encoded protein (Young & Bläsi, 1995), Whereas phage infection is achieved by using either a contractile tail (e.g. T4) (Hershey & Chase, 1952; Kanamaru et al., 2002), forcing a channel through the host cellular membrane to allow genome entry, or via a non-contractile tail that excretes proteins which then form a channel through the host cellular membrane. In both mechanisms, after the entry of the phage genetic material into the host cell, the virion coat remains outside the cell (Hershey & Chase, 1952). This is a very broad generalisation of the mechanisms and there are known exceptions to this rule, such as bacteriophage  $\phi 6$  from the genus *Pseudomonas* (Hu et al., 2008) (Literature Review, Section 1.5).

The mechanism that filamentous phage use for membrane entry may be more a likened to the method used by eukaryotic membrane-enclosed viruses (e.g. HIV (Brasseur et al., 1988; Furuta et al., 1998; Lu et al., 1995; Root et al., 2001; Tan et al., 1997) and Influenza (Bullough et al., 1994; Korte et al., 1999)) to enter host cells. Eukaryotic membrane-enveloped viruses generally use a protein which contains a "fusion peptide" (White, 1992). This fusion peptide is highly hydrophobic sequence that is inserted into the host cellular membrane after a trigger such as binding to host cell receptors or a change in environmental pH (Wiley & Skehel, 1987). The protein carrying the fusion peptide then undergoes major structural rearrangements which results in viral/host cell membrane fusion, usually triggered by a secondary event e.g. a drop in environmental pH or the binding of a second host cell receptor (Colman & Lawrence, 2003). Like

eukaryotic membrane enclosed viruses, fl filamentous phage coat proteins re-insert into the host cell membrane. Thus, I propose the mechanism of insertion may be related to the mechanism used by eukaryotic membrane enclosed viruses, except in this case there is no membrane fusion event, rather a membrane-insertion event. I also propose that the pIII C-domain is required for this to occur, and that it is likely that the "fusion peptide" is the C-terminal membrane anchor.

Another interesting fact that supports this argument is that none of the fusion proteins found in eukaryotic viruses can be reused in the assembly of progeny virions (Colman & Lawrence, 2003). This is similar to pIII in that all current evidence shows that pIII cannot be recycled into new phage virions, unlike pVIII. The mechanism of virion integration mediated by pIII must however be very different from that of eukaryotic enveloped viruses given that the fl virion contains no lipids. Thus, the similarity to membrane enclosed viruses is probably limited to the membrane insertion step, but not the subsequent steps that are required in enveloped viruses for the fusion of viral and cellular membranes. Instead, there has to be some unknown step or event that triggers the removal or dissociation of the pIII/pVI complex from the tip of the reintegrated virion allowing the insertion the first layer of pVIII into the membrane, such as interaction with the TolQRA complex. From this step, the reintegration of the pVIII portion of the virion coat can proceed synchronously with the transporting the ssDNA genome into the host cell. At the present time, the mechanics of these events are also unknown. The discovery of the mechanisms involved in these steps would rely on the production of a high-resolution structure of pIII in various conformations corresponding to the steps of phage infection, allowing modelling of the proposed events.

## Chapter 7

### Conclusion

This thesis is comprised of two separate, but associated lines of investigation. The first line of research presented was on the molecular biology work studying the involvement of the C-domain and minimal requirement of the minor coat protein of pIII in the post-receptor binding events leading to phage infection. The second line of research presented was the work carried out to modify and improve a previously described microphage production system (Specthrie et al., 1992). Microphage are required for use in various structural studies of the filamentous phage virion, particularly in the analysis of the structure of the virion termini. These two lines of research are linked by the minor coat protein pIII, which is an integral part of the pIII/pVI phage cap complex. The phage cap complex is involved in processes at both ends of the phage lifecycle; the process of filamentous phage infection and the process of termination of phage assembly. The analysis of the C-domain of pIII and the phage termini using molecular biology has improved the current understanding of filamentous phage infection and phage assembly.

To increase understanding of the pIII C-domain-mediated entry of f1 phage into the host cells, this thesis mapped the minimal fragment of pIII C-domain required for infection. Mapping showed that a pIII C-terminal fragment of 121 residues was the minimal tested fragment that could effectively mediate phage infection. Fragments of the C-domain shorter than C121 residues were either partially infective (C111) or could not mediate infection (C93 and C83). This is a larger portion of the C-domain than that is required for the termination of assembly; a fragment containing only 93 C-terminal residues is sufficient for phage release. Therefore, the process of virion opening and merging into the host membrane is not directly functionally symmetrical to process of phage excision from the membrane at the end of assembly. This is most likely due to the difference in co-factors required for the two processes. The termination of assembly involves a phage-encoded complex (pI/pXI: pIV), whereas the process of infection requires the host encoded F pilus and the TolQRA complex. The sequence between infection-

competent and release-competent C-terminal fragments, 121-GKLDV VAVDY GAAID GFVGD VVGLA NGNG-93, was designated the "infection competence sequence" (ICS). I propose that the ICS region of the pIII C-domain is required to form important contacts and/or structural motifs that are required to mediate integration of the virion into the membrane, allowing infection to proceed. These important contacts are most likely formed with the periplasmic and/or transmembrane portion of the TolQRA complex.

Alanine scanning mutagenesis of the ICS was carried out, to identify which specific residue/s within the ICS are involved in infection. A total of 12 out of 28 residues of the ICS were mutated. None of the mutations affected the infection, suggesting that the key residues involved in infection are either among the remaining 16 residues or that the residues of ICS functions in conjunction with downstream sequences. An alternative explanation is that the ICS region is functioning as a structural unit and the modification of single residue did not modify the structure of the ICS region significantly enough for an observable effect. In conclusion, more extensive mutagenesis of the ICS and the rest of the C-domain will be required to identify residues involved in infection.

Ultimately, a high-resolution structure of the pIII C-domain before integration into the virion, within the virion and post infection will be required to fully understand the structural transitions involved in the process of pIII C-domain mediated infection. To begin the structural analyses of the C-domain within the virion, a novel microphage producing system and purification protocol were designed. This microphage production system was based on a published low-efficiency system (Specthrie, Bullitt et al., 1992). To obtain a high-efficiency system for producing microphage, the microphage producing origin of replication was transferred into a high copy number plasmid. In addition, a new helper phage was constructed that favours assembly of short virions, without compromising the replication cycle. A high-yield microphage purification protocol was also designed that eliminated the long helper phage virions, producing a pure stock of filamentous microphage at a concentration suitable for structural analyses of the virions. In the future, this microphage production and purification protocol will be used for the structural analysis of the phage termini using a range of technique including cryo-EM, x-ray crystallography and DXMS.

The pIII protein of filamentous phage is particularly interesting as it appears to be mediating the bi-directional integration and excision of the virion from the host cell's cellular membrane. Understanding of how pIII is involved in mediating this bi-direction transition during infection and phage assembly termination may provide further insights into the structural mechanisms involved in the processes of viral infection and the reintegration proteins into membranes.

## Appendix

### Appendix 1 - General Statistic of Ff Filamentous Phage

base	mol w	bases in f1 wt	total mw f1 wt	total mw f1d3	total mw VCSM13d3	total mw VCSM13d3
A	313.2	1574	492976.8	400896	1888	591321.6
T	304.2	2210	672282	551210.4	2399	729775.8
G	329.2	1325	436190	333479.6	1508	496433.6
C	289.2	1298	375381.6	286430	1597	461852.4

	f1	f1d3	VCSM13d3
Gram-mole (g)	1976830	1572016	2279383.4
Molecules (Avogadro's number)	$6.02 \times 10^{23}$	$6.02 \times 10^{23}$	$6.022 \times 10^{23}$
Molecules/gram DNA	$3.05 \times 10^{17}$	$3.83 \times 10^{17}$	$2.64 \times 10^{17}$
Molecules/mg DNA	$3.05 \times 10^{14}$	$3.83 \times 10^{14}$	$2.64 \times 10^{14}$
Genomes/ $\mu$ g DNA	$3.05 \times 10^{11}$	$3.83 \times 10^{11}$	$2.64 \times 10^{11}$

<b>Other physical measures (based on Day)</b>
---

Extinction coefficient (Day)  
 $e = 3.84 \text{ mL / (mg*cm)}$  at 269 nm (1mg/mL f1 in 1 cm cuvette  
has OD=3.84)

<b>Virion composition</b>
---------------------------

12% DNA
88% protein

DNA	proteins	Total mass (DNA+ proteins)	pVI	pIII
1 $\mu$ g	7.33 $\mu$ g	8.33 $\mu$ g	0.025 $\mu$ g	0.088 $\mu$ g

Mass ratio of pIII to pVIII in a monomeric  
virion is 0.012

Mass ratio of pVI to pVIII in a monomeric  
virion is 0.0034

**Appendix 2 - Position of all f1 *gIII* primers used in this thesis**

```

                NJB2
                ~~~~~
                Signal Sequence
                ~~~~~
V  K  K  L  L  F  A  I  P  L  V  V  P  F  Y  S  H  S  A  E  T  V  E  S  .
1  GTGAAAAAAT TATTATTCGC AATTCCTTTA GTTGTTTCCTT TCTATTCTCA CTCCGCTGAA ACTGTTGAAA
   CACTTTTTTTA ATAATAAGCG TTAAGGAAAT CAACAAGGAA AGATAAGAGT GAGGCGACTT TGACAACCTTT

                N1 Domain
                ~~~~~
.  C  L  A  K  P  H  T  E  N  S  F  T  N  V  W  K  D  D  K  T  L  D  R  .
71 GTTGTTTLAGC AAAACCCCAT ACAGAAAATT CATTACTAA CGTCTGGAAA GACGACAAA CTTTAGATCG
   CAACAAATCG TTTTGGGTA TGTCTTTTAA GTAAATGATT GCAGACCTTT CTGCTGTTTT GAAATCTAGC

                N1 Domain
                ~~~~~
.  Y  A  N  Y  E  G  C  L  W  N  A  T  G  V  V  V  C  T  G  D  E  T  Q
141 TTACGCTAAC TATGAGGGCT GTCTGTGGAA TGCTACAGGC GTTGTAGTTT GTACTGGTGA CGAAACTCAG
   AATGCGATTG ATACTCCCGA CAGACACCTT ACGATGTCCG CAACATCAAA CATGACCACT GCTTTGAGTC

```

Glycine linker 1

~~~~~

N1 Domain

~~~~~

211 C Y G T W V P I G L A I P E N E G G G S E G G G ·  
TGTTACGGTA CATGGGTTCC TATTGGGCTT GCTATCCCTG AAAATGAGGG TGGTGGCTCT GAGGGTGGCG  
ACAATGCCAT GTACCCAAGG ATAACCCGAA CGATAGGGAC TTTTACTCCC ACCACCGAGA CTCCCACCGC  
Glycine linker 1

~~~~~

N2 Domain

~~~~~

281 · S E G G G S E G G G T K P P E Y G D T P I P G ·  
GTTCTGAGGG TGGCGGCTCT GAGGGTGGCG G TACTAAACC TCCTGAGTAC GGTGATACAC CTATTCCGGG  
CAAGACTCCC ACCGCCGAGA CTCCCACCGC CATGATTTGG AGGACTCATG CCACTATGTG GATAAGGCC

N2 Domain

~~~~~

351 · Y T Y I N P L D G T Y P P G T E Q N P A N P N ·  
CTATACTTAT ATCAACCCTC TCGACGGCAC TTATCCGCCT GGTACTGAGC AAAACCCCGC TAATCCTAAT  
GATATGAATA TAGTTGGGAG AGCTGCCGTG AATAGGCGGA CCATGACTCG TTTTGGGGCG ATTAGGATTA

N2 Domain

~~~~~

421 P S L E E S Q P L N T F M F Q N N R F R N R Q G ·  
CCTTCTCTTG AGGAGTCTCA GCCTCTTAAT ACTTTCATGT TTCAGAATAA TAGGTTCCGA AATAGGCAGG  
GGAAGAGAAC TCCTCAGAGT CGGAGAATTA TGAAAGTACA AAGTCTTATT ATCCAAGGCT TTATCCGTCC

N2 Domain

```
~~~~~  
· A L T V Y T G T V T Q G T D P V K T Y Y Q Y T ·  
491 GGCATTAAC TGTTTATACG GGCACGTGTTA CTCAAGGCAC TGACCCCGTT AAAACTTATT ACCAGTACAC  
CCCGTAATTG ACAAATATGC CCGTGACAAT GAGTTCCGTG ACTGGGGCAA TTTTGAATAA TGGTCATGTG
```

NJB20

N2 Domain

```
~~~~~  
· P V S S K A M Y D A Y W N G K F R D C A F H S  
561 TCCTGTATCA TCAAAAGCCA TGTATGACGC TTAGTGGAAC GGTAATTCA GAGACTGCGC TTTCCATTCT  
AGGACATAGT AGTTTTTCGGT ACATACTGCG AATGACCTTG CCATTTAAGT CTCTGACGCG AAAGGTAAGA
```

NJB20

~~~~~

N2 Domain

```
~~~~~  
BamHI  
~~~~~  
G F N E D P F V C E Y Q G Q S S D L P Q P P V N ·  
631 GGCTTTAATG AGGATCCATT CGTTTGTGAA TATCAAGGCC AATCGTCTGA CCTGCCTCAA CCTCCTGTTA  
CCGAAATTAC TCCTAGGTAA GCAAACACTT ATAGTTCCGG TTAGCAGACT GGACGGAGTT GGAGGACAAT
```

Glycine linker 2

~~~~~

N2 Domain

~~~~~

· A G G G S G G G S G G G S E G G G S E G G G S ·  
701 ATGCTGGCGG CGGCTCTGGT GGTGGTTCTG GTGGCGGCTC TGAGGGTGGT GGCTCTGAGG GTGGCGGTTC  
TACGACCGCC GCCGAGACCA CCACCAAGAC CACCGCCGAG ACTCCCACCA CCGAGACTCC CACCGCCAAG

Glycine linker 2

~~~~~

C-domain

· E G G G S E G G G S G G G S G S G D F D Y E K  
771 TGAGGGTGGC GGCTCTGAGG GTGGCGGTTC CGGTGGTGGC TCTGGTCCG GTGATTTTGA TTATGAAAAG  
ACTCCCACCG CCGAGACTCC CACCGCCAAG GCCACCACCG AGACCAAGGC CACTAAAACT AATACTTTTC

~~~~~

NJB9

~~~~~

NJB3

```

NJB3121
~
NJB3120
~
NJB3119
~
NJB3118
~
NJB6
~
NJB4
~~~~~
NJB5
~~~~~
C-domain
~~~~~
M A N A N K G A M T E N A D E N A L Q S D A K G .
841 ATGGCAAACG CTAATAAGGG GGCTATGACC GAAAATGCCG ATGAAAACGC GCTACAGTCT GACGCTAAAG
TACCGTTTGC GATTATTCCC CCGATACTGG CTTTACGGC TACTTTTGCG CGATGTCAGA CTGCGATTTC
~~~~~
NJB3005
~~~~~
NJB3006

```

```

                NJB3117                NJB3109                NJB3098
                ~~~~~                    ~~~~~                    ~~~~~
                NJB3116                NJB3108                NJB3099
                ~~~~~                    ~~~~~                    ~~~~~
                NJB3115                NJB3107                NJB3100
                ~~~~~                    ~~~~~                    ~~~~~
                NJB3114                NJB3106                NJB3101
                ~~~~~                    ~~~~~                    ~~~~~
    NJB3121                NJB3113                NJB3102
    ~~~~~                    ~~~~~                    ~~~~~
    NJB3120                NJB3112                NJB3103
    ~~~~~                    ~~~~~                    ~~~~~
    NJB3119                NJB3111                NBJ3104
    ~~~~~                    ~~~~~                    ~~~~~
    NJB3118                NJB3110                NJB3105
    ~~~~~                    ~~~~~                    ~~~~~
    NJB6                    NJB7
    ~~~~~                    ~~~~~
                                C-domain
    ~~~~~
    . K L D S V A T D Y G A A I D G F I G D V S G L .
911 GCAAACCTGA TTCTGTCGCT ACTGATTACG GTGCTGCTAT CGACGGTTC ATTGGTGACG TTTCCGGCCT
    CGTTTGAAC T AAGACAGCGA TGACTAATGC CACGACGATA GCTGCCAAAG TAACCACTGC AAAGGCCGGA
    ~~~~~                    ~~~~~                    ~~~~~
                NJB3004                NJB3002                NJB3000
    ~~~~~                    ~~~~~                    ~~~~~
    NJB3005                NJB3003                NJB3001

```

NJB3093  
 ~~~~~  
 NJB3094  
 ~~~~~  
 NJB3095  
 ~~~~~  
 NJB3096  
 ~~~~~  
 NJB3097  
 ~~~~~  
 NJB3098  
 ~~~~~  
 NJB3099  
 ~~~~~  
 NJB3100  
 ~~~~~  
 NJB3101

NJB8

C-domain

981

· A N G N G A T G D F A G S N S Q M A Q V G D G  
 TGCTAATGGT AATGGTGCTA CTGGTGATTT TGCTGGCTCT AATCCCAA TGGCTCAAGT CGGTGACGGT  
 ACGATTACCA TTACCACGAT GACCACTAAA ACGACCGAGA TTAAGGGTTT ACCGAGTTCA GCCACTGCCA  
 ~  
 .

C-domain

~~~~~  
D N S P L M N N F R Q Y L P S L P Q S V E C R P ·  
1051 GATAATTCAC CTTTAATGAA TAATTTCCGT CAATATTTAC CTTCCCTCCC TCAATCGGTT GAATGTCGCC  
CTATTAAGTG GAAATTA CTT ATTAAGGCA GTTATAAATG GAAGGGAGGG AGTTAGCCAA CTTACAGCGG

C-domain

~~~~~  
· F V F G A G K P Y E F S I D C D K I N L F R G ·  
1121 CTTTTGTCTT TGGCGCTGGT AAACCATATG AATTTTCTAT TGATTGTGAC AAAATAAACT TATTCCGTGG  
GAAAACAGAA ACCGCGACCA TTTGGTATAC TTAAAAGATA ACTAACACTG TTTTATTTGA ATAAGGCACC

Transmembrane anchor

~~~~~  
C-domain NJB1  
~  
· V F A F L L Y V A T F M Y V F S T F A N I L R  
1191 TGTCTTTGCG TTTCTTTTAT ATGTTGCCAC CTTTATGTAT GTATTTTCGA CGTTTGCTAA CATACTGCGT  
ACAGAAACGC AAAGAAAATA TACAACGGTG GAAATACATA CATAAAAGCT GCAAACGATT GTATGACGCA  
NJB19

```

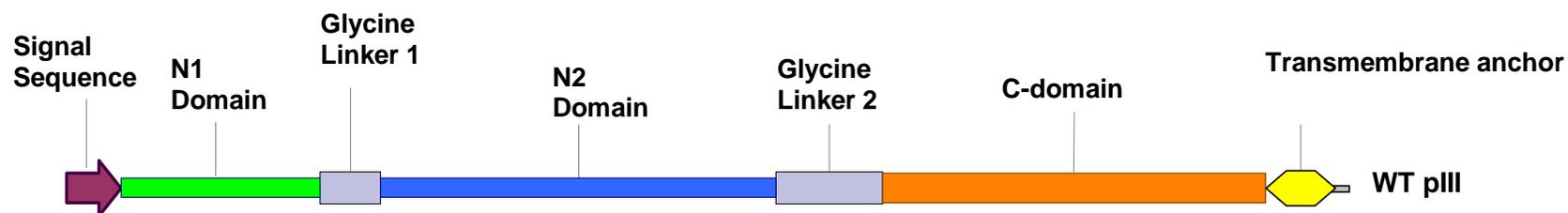
      NJB1
      ~~~~~
      N K E S *
1261  AATAAGGAGT CTTAA
      TTATTCCTCA GAATT
      ~~~~~
      NJB19

```

**Note for diagram:** Primer direction indicated by side of sequence. Primers above the sequence are 5'-3'. Primers below the sequence are 3'-5'

## Appendix 3 - Protein Sequence of WTpIII and NdC mutants.

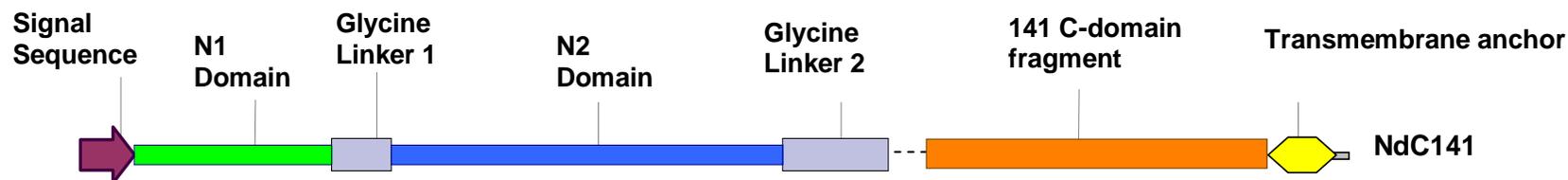
### f1 WT pIII (425 aa)



### f1 WT pIII sequence

|     |            |             |            |            |            |            |
|-----|------------|-------------|------------|------------|------------|------------|
| 1   | VKKLLFAIPL | VVPFYSHSAE  | TVESCLAKPH | TENSFTNVWK | DDKTLDRYAN | YEGCLWNATG |
| 61  | VVCTGDETQ  | CYGTWVPIGL  | AIPENEGGGS | EGGGSEGGGS | EGGGTKPPEY | GDTPIPGYTY |
| 121 | INPLDGTYPP | GTEQNAPANPN | PSLEESQPLN | TFMFQNNRFR | NRQGALTVYT | GTVTQGTDPV |
| 181 | KTYYQYTPVS | SKAMYDAYWN  | GKFRDCAFHS | GFNEDPFVCE | YQGQSSDLPQ | PPVNAGGGSG |
| 241 | GGSGGGSEGG | GSEGGGSEGG  | GSEGGGSGGG | SGSGDFDYEK | MANANKGAMT | ENADENALQS |
| 301 | DAKGLDSVA  | TDYGAAIDGF  | IGDVSGLANG | NGATGDFAGS | NSQMAQVGDG | DNSPLMNNFR |
| 361 | QYLPSLPQSV | ECRPFVFGAG  | KPYEFSIDCD | KINLFRGVFA | FLLYVATFMY | VFSTFANILR |
| 421 | NKES       |             |            |            |            |            |

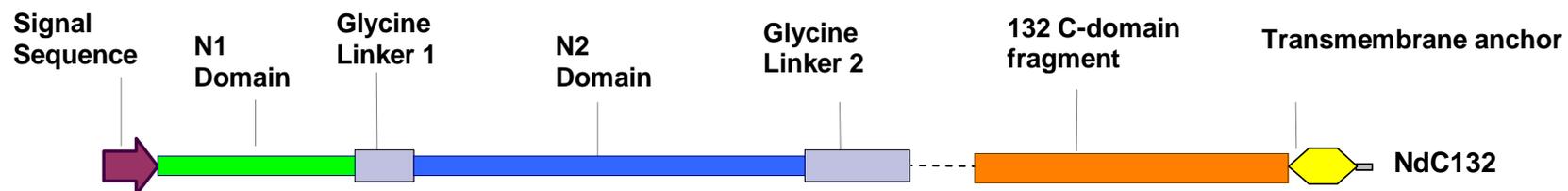
## NdC141 pIII (421aa)



## NdC141 pIII sequence

|     |             |            |            |            |            |            |
|-----|-------------|------------|------------|------------|------------|------------|
| 1   | MAKKLLFAIP  | LVVPFYSHSA | ETVESCLAKP | HTENSFTNVW | KDDKTLDRYA | NYEGCLWNAT |
| 61  | GVVVCTGDET  | QCYGTWVPIG | LAIPENEGGG | SEGGGSEGGG | SEGGGTKPPE | YGDTPIPGYT |
| 121 | YINPLDGTYP  | PGTEQNPANP | NPSLEESQPL | NTFMFQNNRF | RNRQGALTVY | TGTVTQGTDP |
| 181 | VKTYYYQYTPV | SSKAMYDAYW | NGKFRDCAFH | SGFNEDPFVC | EYQQQSSDLP | QPPVNAGGGS |
| 241 | GGGSGGGSEG  | GGSEGGGSEG | GGSEGGGSGG | GSGSGDFDYE | ANKGAMTENA | DENALQSDAK |
| 301 | GKLDSVATDY  | GAAIDGFIGD | VSGLANGNGA | TGDFAGSNSQ | MAQVGDGDNS | PLMNNFRQYL |
| 361 | PSLPQSVECR  | PFVFGAGKPY | EFSIDCDKIN | LFRGVFAFLL | YVATFMYVFS | TFANILRNKE |
| 421 | S           |            |            |            |            |            |

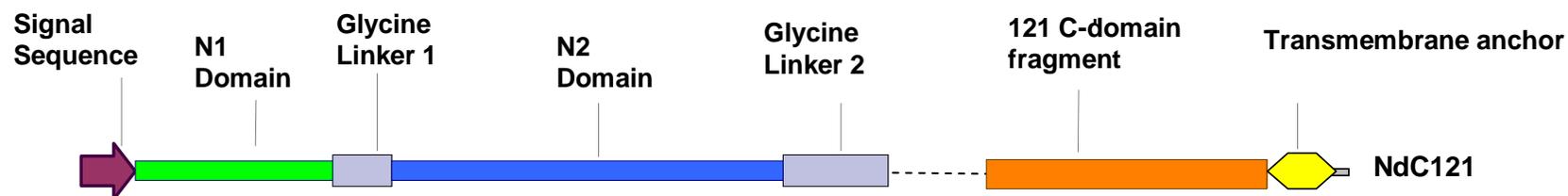
## NdC132 pIII (412 aa)



## NdC132 pIII sequence

|     |             |            |            |            |            |            |
|-----|-------------|------------|------------|------------|------------|------------|
| 1   | MAKKLLFAIP  | LVVPFYSHSA | ETVESCLAKP | HTENSFTNVW | KDDKTLDRYA | NYEGCLWNAT |
| 61  | GVVVCTGDET  | QCYGTWVPIG | LAIPENEGGG | SEGGGSEGGG | SEGGGTKPPE | YGDTPIPGYT |
| 121 | YINPLDGTYP  | PGTEQNPANP | NPSLEESQPL | NTFMFQNNRF | RNRQGALTVY | TGTVTQGTDP |
| 181 | VKTYYYQYTPV | SSKAMYDAYW | NGKFRDCAFH | SGFNEDPFVC | EYQGQSSDLP | QPPVNAGGGS |
| 241 | GGSGGGGSEG  | GGSEGGGSEG | GGSEGGGSGG | GSGSGDFDYE | ADENALQSDA | KGKLDSVATD |
| 301 | YGAAIDGFIG  | DVSGLANGNG | ATGDFAGSNS | QMAQVGDGDN | SPLMNNFRQY | LPSLPQSVEC |
| 361 | RPFVFGAGKP  | YEFSIDCDKI | NLFRGVFAFL | LYVATFMYVF | STFANILRNK | ES         |

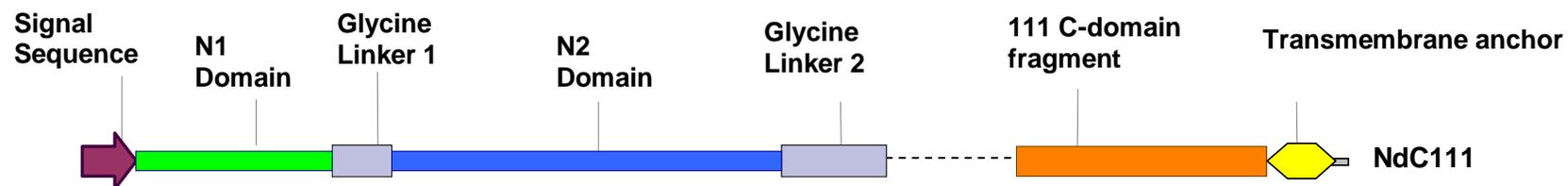
## NdC121 pIII (401 aa)



## NdC121 pIII Sequence

|     |             |            |            |            |            |            |
|-----|-------------|------------|------------|------------|------------|------------|
| 1   | MAKKLLFAIP  | LVVPFYSHSA | ETVESCLAKP | HTENSFTNVW | KDDKTLDRYA | NYEGCLWNAT |
| 61  | GVVVCTGDET  | QCYGTWVPIG | LAIPENEGGG | SEGGGSEGGG | SEGGGTKPPE | YGDTPIPGYT |
| 121 | YINPLDGTYP  | PGTEQNPANP | NPSLEESQPL | NTFMFQNNRF | RNRQGALTVY | TGTVTQGTDP |
| 181 | VKTYYYQYTPV | SSKAMYDAYW | NGKFRDCAFH | SGFNEDPFVC | EYQQQSSDLP | QPPVNAGGGS |
| 241 | GGGSGGGSEG  | GGSEGGGSEG | GGSEGGGSGG | GSGSGDFDYE | GKLDSVATDY | GAAIDGFIGD |
| 301 | VSGLANGNGA  | TGDFAGSNSQ | MAQVGDDNS  | PLMNNFRQYL | PSLPQSVECR | PFVFGAGKPY |
| 361 | EFSIDCDKIN  | LFRGVFAFLL | YVATFMYVFS | TFANILRNKE | S          |            |

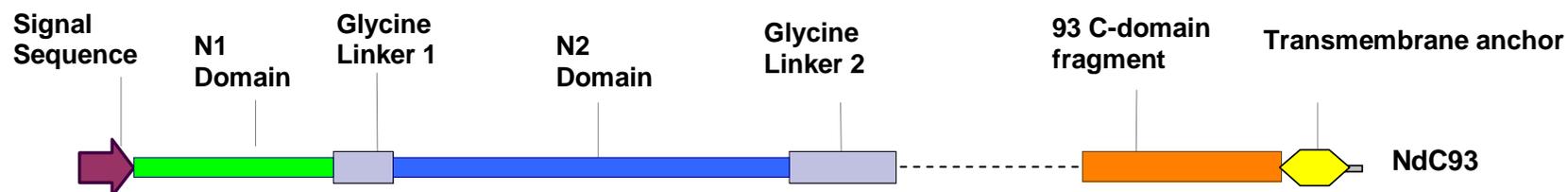
## NdC111 pIII (390 aa)



## NdC111 Sequence

|     |            |             |            |            |            |             |
|-----|------------|-------------|------------|------------|------------|-------------|
| 1   | MAKKLLFAIP | LVVPFYSHSA  | ETVESCLAKP | HTENSFTNVW | KDDKTLDRYA | NYEGCLWNAT  |
| 61  | GVVVCTGDET | QCYGTWVPIG  | LAIPENEGGG | SEGGGSEGGG | SEGGGTKPPE | YGDTPIPGYT  |
| 121 | YINPLDGTYP | PGTEQNPANP  | NPSLEESQPL | NTFMFQNNRF | RNRQGALTVY | TGTVTQGTDP  |
| 181 | VKTTYQYTPV | SSKAMYDAYW  | NGKFRDCAFH | SGFNEDPFVC | EYQGQSSDLP | QPPVNAGGGS  |
| 241 | GGSGGGSEG  | GGSEGGGSEG  | GGSEGGGSGG | GSGSGDFDYE | GAAIDGFIGD | VSGLANGNGA  |
| 301 | TGDFAGSNSQ | MAQVGDG DNS | PLMNNFRQYL | PSLPQSVECR | PFVFGAGKPY | EF SIDCDKIN |
| 361 | LFRGVFAFL  | YVATFMYVFS  | TFANILRNKE |            |            |             |

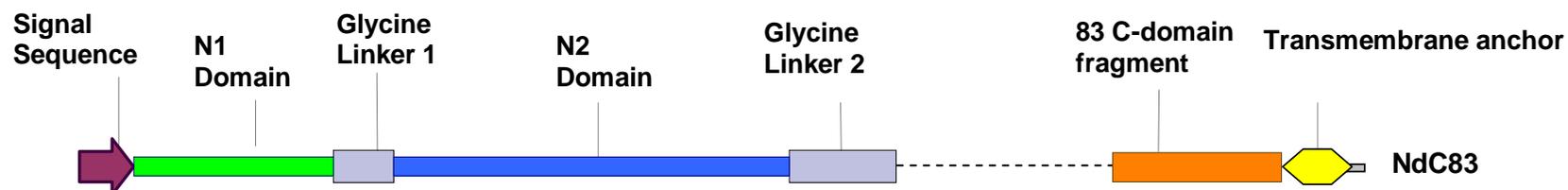
## NdC93 pIII (373 aa)



## NdC93 Sequence

|     |             |            |            |            |            |            |
|-----|-------------|------------|------------|------------|------------|------------|
| 1   | MAKKLLFAIP  | LVVPFYSHSA | ETVESCLAKP | HTENSFTNVW | KDDKTLDRYA | NYEGCLWNAT |
| 61  | GVVVCTGDET  | QCYGTWVPIG | LAIPENEGGG | SEGGGSEGGG | SEGGGTKPPE | YGDTPIPGYT |
| 121 | YINPLDGTYP  | PGTEQNPANP | NPSLEESQPL | NTFMFQNNRF | RNRQGALTVY | TGTVTQGTDP |
| 181 | VKTYYYQYTPV | SSKAMYDAYW | NGKFRDCAFH | SGFNEDPFVC | EYQGQSSDLP | QPPVNAGGGS |
| 241 | GGGSGGGSEG  | GGSEGGGSEG | GGSEGGGSGG | GSGSGDFDYE | GATGDFAGSN | SQMAQVGDGD |
| 301 | NSPLMNNFRQ  | YLPSLPQSVE | CRPFVFGAGK | PYEFSIDCDK | INLFRGVFAF | LLYVATFMYV |
| 361 | FSTFANILRN  | KES        |            |            |            |            |

## NdC83 pIII (364 aa)



## NdC83 sequence

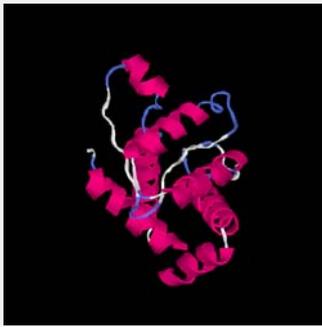
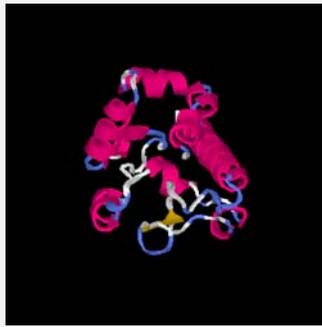
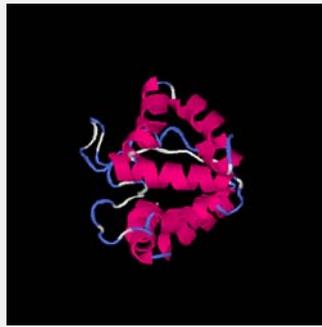
```

1   MAKKLLFAIP LVVPFYSHSA ETVESCLAKP HTENSFTNVW KDDKTLDRYA NYEGCLWNAT
61  GVVVCTGDET QCYGTWVPIG LAIPENEGGG SEGGGSEGGG SEGGGTKPPE YGDTPIPGYT
121 YINPLDGTYP PGTEQNPANP NPSLEESQPL NTFMFQNNRF RNRQGALTVY TGTVTQGTDP
181 VKTYYYQYTPV SSKAMYDAYW NGKFRDCAFH SGFNEDPFVC EYQGQSSDLP QPPVNAGGGS
241 GGGSGGGSEG GGSEGGGSEG GGSEGGGSGG GSGSGDFDYE SQMAQVGDGD NSPLMNNFRQ
301 YLPSLPQSVE CRPFVFGAGK PYEFSIDCDK INLFRGVFAF LLYVATFMYV FSTFANILRN
361 KES

```



Top 5 Models predicted by I-TASSER

|                                                                                   |                                                                                   |                                                                                    |                                                                                     |                                                                                     |
|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
|  |  |  |  |  |
| <a href="#">Download Model 1</a>                                                  | <a href="#">Download Model 2</a>                                                  | <a href="#">Download Model 3</a>                                                   | <a href="#">Download Model 4</a>                                                    | <a href="#">Download Model 5</a>                                                    |
| C-score=-3.27                                                                     | C-score=-4.25                                                                     | C-score=-4.88                                                                      | C-score=-4.70                                                                       | C-score=-4.79                                                                       |
| <a href="#">(more about C-score of generated models)</a>                          |                                                                                   |                                                                                    |                                                                                     |                                                                                     |

Top 10 templates used by I-TASSER

| Rank | PDB Hit               | Iden1 | Iden2 | Cov. | Norm. Z-score | Sec.Str Seq                                                                                                                                                                                                                                                                                                            |
|------|-----------------------|-------|-------|------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      |                       |       |       |      |               | CCHHHHHHCCCCCCCCCHHHHHHHHHCCHHHHHHCHHHHHHCCCCCCCCCCCCCCCCCCCC<br>DFDYKMANANKGAMTENADENALQSDAKGKLDVATDYGAAIDGFIGDVSGLANGNGATGDFAGS<br>HHHHHHCCCCCCHHHHCCCCCCCCCCCCCCCCSSSSCCSSSSSSCCHHHHHHHHHHHHHHHHH<br>NSQMAQVGDGDN SPLMNNFRQYLPSLPQSV ECRPFVFGAGKPYEFSIDCDKINLFRGVFAFLLYVATF<br>HHHHHHHHHHHHCCCC<br>MYVFSTFANILRNKES |
| 1    | <a href="#">1f01A</a> | 0.17  | 0.21  | 0.99 | 0.62          | NLDWDVIRDKTKESLKEHGPIKNKMSSEPNKSEEKAKQYLEEFHQTHPELSELKTVTGTNPVFAGANYA<br>AWAVAQVIDSETADNLEKTTAALSILPGIGSVMIADGAVHEISIAL-SSLMVAQAIPLVNFVESIINLF<br>QVVHNSYNRPAY                                                                                                                                                         |
| 2    | <a href="#">3cooA</a> | 0.09  | 0.15  | 0.67 | 0.63          | -----IAQEFSLRVEGDPD-----FYKPGTSY-----RVTL SAAPPSRGFTDKEEDHAGTFQIIDEE<br>EVAVTESTPRRR-TRIQVFWIAPPAGTGCVILKAQKRIIYFQDKLCEQD-----<br>-----                                                                                                                                                                                |
| 3    | <a href="#">2dbbA</a> | 0.08  | 0.13  | 0.94 | 0.48          | KLD--RVDMQLVKILSENSRLT--YRELADILNTTRQRIARRIDKLKKG-<br>IIRKFTIIPDIDKLGMYAIVLIKSKVPSDADK VISEISDIEYVKS--EKGVGRYNIIVRLLLPKIKD<br>AENLISEFLQRIK--NAENVEVILIRKFEI                                                                                                                                                           |
| 4    | <a href="#">1cy0A</a> | 0.09  | 0.15  | 0.79 | 0.75          | NIDSPLLWKKIARGLSAGRVQSVAVRLVVEREREIKAF-----VPKPPARFSEASLVKE<br>LEKRGIGRPSTYASII STIQDRGY-----VRVENRRFYAE-----KMGEIVTDRLEYDFTAQM<br>ENSLDQVANHET                                                                                                                                                                        |
| 5    | <a href="#">2fokA</a> | 0.14  | 0.17  | 0.96 | 0.61          | KSELEEKKSELRHKLKYELIEIARNSTQDRILEMKVMEFFMKVYGYRGTGSPIDYGAYSGGYNLPIGQ<br>ADEMQR-YVEENQTRNKHINPNYPSSVTEFKFLFVSGHFYK-----AQLTRLNHITNCNGAVLSVEELL<br>IGGEMIKAGTLT                                                                                                                                                          |
| 6    | <a href="#">2acvA</a> | 0.13  | 0.20  | 0.92 | 0.95          | QIQLIDLPEVEPPPLLKSPEFY--ILTFLES LI PHVKATIKTILKVGLV-----LDFFCVSMI<br>DVGNEFGIPSYLFLTNRQIEEVFDDSDRDHQLLNIPGISNQVPSNVLPDACFNKGGYIAYYNTFSDLE<br>QSSIDALYDHDE                                                                                                                                                              |

|    |                       |      |      |      |      |                                                                                                                                                                |
|----|-----------------------|------|------|------|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 7  | <a href="#">2beqA</a> | 0.22 | 0.05 | 0.24 | 0.72 | -----SIIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLS-----<br>-----<br>-----                                                                                               |
| 8  | <a href="#">3cooA</a> | 0.23 | 0.15 | 0.21 | 0.50 | -----KSEGY-CDPDFYKPGTSYRVTLAAPPYFRG-----<br>-----                                                                                                              |
| 9  | <a href="#">lit2A</a> | 0.11 | 0.14 | 0.93 | 0.41 | IIDQGPLP-----TLTD--GDKKAINKIWPKIYKEYEQYSLNILLRFLKCFSPKFSTKKSNI-LEQDPE<br>VKH--QAVVIFNKVNEIINSMDNQEIIKSLKDLSQKHKTVFKVDSIWFKELSSIFVSTIDGAEFEKLF<br>SIICILLRSAY-  |
| 10 | <a href="#">lglnA</a> | 0.13 | 0.15 | 0.90 | 0.62 | --DLEKLRWMNGKYIREVLSLEEVAERVKPFLREAEAYLRRAVELMRPRFDLKEFRYLFTEYDYPVSEKAQR<br>KLE----EGLPLLKELYPRLRAQEEWTEAALEA----LLRGFAAEKGVLGQVAQPLRAALT---LFEILA<br>LL--KERA |

- (a) All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are colored in dark shade.
- (b) Rank of templates represents the top ten threading templates used by I-TASSER.
- (c) Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.
- (d) Ident2 is the percentage sequence identity of the whole template chains with query sequence.
- (e) Cov. represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.
- (f) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.
- (g) Download Align. provides the 3D structure of the aligned regions of the threading templates.
- (h) The top 10 alignments reported above (in order of their ranking) are from the following threading programs:  
1: MUSTER    2: HHSEARCH    3: SP3    4: PROSPECT2    5: PPA-I    6: HHSEARCH I    7: FUGUE    8: HHSEARCH II    9: SPARKS    10: MUSTER

10 proteins in PDB which are structurally closest to the first I-TASSER model (identified by [TM-align](#))

| Rank | TM-score | RMSDa | IDENa | Cov. | PDB Hit                         | Structural alignment using TM-align                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|------|----------|-------|-------|------|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | 0.6325   | 3.78  | 0.12  | 0.93 | <a href="#">1f01A</a><br>Model1 | CINLDWDVIRDKTKTKIESLKEHGPIKNKMSSES-PNKTVSEEKAKQYLEEFH-QTALEHPELSELKTVTGTNPVAFAGANY<br>AAWAVNVAQVIDSETADNLEKTTAALSILPGI-GS-VMGIADG-----AVHHNTEEIIVAQSIALSSLMVAQAIPLVGEL<br>IGFAAYNFVESIINLFQVVHNSYNRPAY----DFD-YEK-MANANKGAMTENADENALQSDAKGKLD--VAT-DYGAAI<br>DGFIGD---VSSL-ANGNGATG-DFAGSNSQMAQVGDGDN SPLM-N--NFRQYLPSLPQSVECRPFV---GAGKPYEFS<br>I-DCDKINLFRGVFAFLLYV-----ATFM-YVFSTFANILRN-KES-----                                                                                                                                                           |
| 2    | 0.5113   | 4.32  | 0.04  | 0.80 | <a href="#">2oc3A</a><br>Model1 | DSASFLERLAVLAGEFSDIQACSAAWKADGVCSTVAGSRPEN--VRKNRYKDV-LPYDQTRVILSLLQEEGHSY-IN--<br>GN-F--IR----GVDGSLAYIATQGPLPHTLLDFWRLVWEFGVKV-IL--MACREIENGRKRCERYWAQEQEPLQTGLFC<br>ITLIKEKWL NEDIMLRTLKVTFQKESRSVYQLQYMSWPDR--GVPSS-PDHMLAMVEEARLQSGSPEPLCVHCSAGCGR<br>TGVLC TVDYVR-----QLLLTQMIPPDFS-LFDVVLKMRKORPAAV-QT--EEQYRFLYHTVAQMFC-----DF--<br>DYEKMANANKGAMTEN---ADEN-A-----LQS--DAKGKLDV-----ATDYGAAIDGFIGDVS<br>GLANGNGAT-----GDFAGSNSQMAQVGDG-----<br>-----NSPLMNNFRQYLPSL--PQ-SVE-----C--RPFV<br>FGAGKPYEFS----ID--CDKINLFRG-VF---AFLLYVATFMYVVFSTFANILRNKES |
| 3    | 0.5011   | 3.72  | 0.09  | 0.71 | <a href="#">2vebA</a><br>Model1 | IPGYTYGETENRAPFNLEDLKLKKEAVMFTAEDDEEYIQK-----AGEVLEDQVEEILD<br>TWYGFVGSHPHL--LYYFTSPDGTTPNEKYLA AVRKRFRSWILDTSNR-SYDQAWLDYQYEIGL----R-HHR----TKKN<br>QTDNVESVPNI-G--YRYLVAFIYPITATMKPFLARKGHTPEEVEKMYQAWFKATTLQVALWSYPYVKYGDF-----<br>-----DFDYEKMANANKGAMTENADENALQSDAKGKLDVATDYGAAIDGFIG<br>DVSGLANGNGA-----T--GDFAGSNSQMAQVGDGDN SPLMNN-FRQYLPSLPQSVECRPFVFGAGKPYE-----<br>-FSIDCDKINLFRGVFAFL-------YVATFMYVVFSTFANILRNKE--S                                                                                                                               |
| 4    | 0.5010   | 3.77  | 0.08  | 0.73 | <a href="#">1cgxA</a><br>Model1 | MLTQKTKDI-----VKATAPVLAEHGYDIIKCFYQRMFEAHPKLVFNMAHQEQGQQ<br>QQALARAVYAYAENIEDPNSLMAVLKNI-----ANK--HASLGVKPEQYPIVGEHL-LAAI KEVLGNAATDDIIS<br>AWAQAYGNLADVLMGMESELYERSAEQPGGWKGWRTFVIREKRPESDVITSFILEPADGGPVVNFEPGQYTSVAIDVPAL<br>GLQQIRQYSLSDMPNGRTYRISVKREGGGPQPPGYVSNLLHDHVNVDGQVKLAAPYGSFHIDVDAKTPIVLISGGVGLTP<br>MVSMLKVALQAPPRQVVFVHGARN SAVHAMRDLREAAKTYENLDLFFVFDQPLPEDVQGRDYDYPGLVDVKQIEKSILL<br>PDADYYICGPIPFMRMQHDALKNLGIHEARIHYEVFGPDLFAE-----DFDYEKMANANKGAMTENADENALQSDA                                                                           |

|   |        |      |      |      |                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|---|--------|------|------|------|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |        |      |      |      |                                 | KGKLDVATDYGAAIDGFIGDV--SGLANGN-GATG-D-F-AGSNSQMAQVGDGDN-SPLMNNFRQYLPQLPQSVECRP<br>FVFGAGKPYEFSIDCDKINLFRGVFAF--L-----L-YV-ATFMYVFSTFANILRN-KE-S-----<br>-----<br>-----<br>-----<br>-----                                                                                                                                                                                                                                                                                                                                                                                  |
| 5 | 0.4999 | 4.59 | 0.05 | 0.81 | <a href="#">2i75A</a><br>Model1 | SLRESMIQLAEGELITGTVLTQFDQ-LY-RKKPGMT-MSCAKLPQNIS-KNRYRDI--SPYDATRVILKGN DYIN--AN-<br>YI-NM---EIPSIINQYIACQGPLPHTCTDFWQMTWEGSSM-VV--MLTTQVERGRVKCHQYWPEPTGSSSYGQYQV<br>TCHSEEGNTAYIFRKMTLNFQEKNESRPLTQIQYIAWPDHGV PDDSSDFLDFVCHVRNKRAGKEEPVVVHCSAGIGRTG<br>VLITMETAM-----CLIECNQPVY-PLDIVRTMRDQRAMMI-QT--PSQYRFVCEAILKVYE-----DF-----<br>--DYEKMANANKGAMTENADENALQS---D-AKGKL--DSVA-----TDYGAAIDGFIGDV SGLANGNG<br>ATGD-----FAGSNSQMAQVGDGDN-----<br>-----SPLMNNFRQYLPQL-PQ-SVE-----C--RPFVFGAGKP<br>YEFS---IDCDKINLFRG-VF--AFLLYVATFMYVFSTFANILRNKES                      |
| 6 | 0.4994 | 4.65 | 0.10 | 0.83 | <a href="#">2h4vA</a><br>Model1 | YFQSMKQFVKHIGELYSNNQHGFSEDFEEVQRCTADMNITAEHSNHPENKHKNRYINILAYDHSRVKLRPLPHSDYINA<br>NYVDGYNKAKA-YIA-TQGPKS--TFEDFWRMIWEQNTGII VMI-----TN-LVEKGRRKCD--QY-WPTEN--SEE<br>YGNII-----VTLKSTKIHACYTVRRFSIRNTKERVVIQYHYT-QW-PDMG---V--PEYA-LPVLTF-VRRSS---<br>-----AARMPET-GPV--LV--HCSAGVGRGTGYIVIDSMLQIQIKDKSTVNVVLGFLKHIRTQRNYLVQTEEQYIFIHD<br>ALLEAILG-----<br>-----DFDYEKMANANKGAM---TEN--ADE-----NALQSDA-KGKL-DSVATDYGAAIDG<br>--FIG-----DVSGLANGNGATGDFA--GSN-----SQMAQVGDGDNSPLMNNFRQYLPQLPQSV<br>ECRPFVFGAGKPYEFSIDCDKINLFRGVFAFLY--VATFMYVFSTFANILRNKE-----<br>-----S-- |
| 7 | 0.4991 | 3.91 | 0.04 | 0.75 | <a href="#">1hbnA</a><br>Model1 | RARGENEP-----GGVPPGYL--ADICQS--SRVN-----Y-EDPVRVSLDVVATGAMLYDQIWLGSY<br>MSGGVGFTQYATAAYTDNILDFTYFGKEYVEDKYGLCEAPN-N--M--D-TVL-DVATE---V---TFYG---LEQY<br>EEYPALLEDQFGGSRAAVVAAAAGCSTAFATGNAQTGLSGWYLSMYLHKEQHSRLGFYYDLQDQGASNVFSIRGDEGLP<br>LELRGPNYPNYAMNVGHQGEYAGISQAPHAARGDAFVFNPLVKIAFADDNLVDFDTNVRGEFAKALREFEPAGERALI<br>TPA-----DFDYEKMANANK-GAMTENADENAL--QSD--AKGKLDVATDYGAAIDGFIG-DV-----<br>-----S-GLANGN-GAT-GD-FAGSNSQMAQVGDGDNSPLMNNFRQYLPQL-PQSVECRP FVFG-AGKPY<br>EFS-----I--D--CDKINLFRGVFAFL-L---YVATFMYVFSTFANILRNKES-----<br>-----                     |

|    |        |      |      |      |                                 | -----                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|----|--------|------|------|------|---------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 8  | 0.4980 | 3.61 | 0.09 | 0.67 | <a href="#">2ig3A</a><br>Model1 | -----M--KF-ETINQESI AKLMEIFYEKVRKDKDL--GPIFNNAIGTSDEEWKEHKAKIGN<br>FWAGMLLGEGD-----YN-GQP-----LKK-HLDLPPFPQEFFEIIWLKLF-EESLNIVYNEEMKNVILQRA<br>QMIASHFQNMLYKYGGH-DFDYEKMANANKGAMTENADENALQSDAKGKLDVATDYGAAIDGFIGDVSGLANGNGAT-<br>--G----DFAGSNSQMAQVGDG DNS---PLMNNFRQY-LPSLPQSVECRPFVFGAGKPYEFSIDCDKINLFRGVFAF-<br>LL-----YVATFMYVFSTFANILRNKE--S                                                                                                                                                                                                                                            |
| 9  | 0.4971 | 4.32 | 0.10 | 0.80 | <a href="#">2b3oA</a><br>Model1 | TRVNAADIENRFW-EEF-ES-LQKQEVKNLHQRLEGQRPENKGN-----RYKN-ILPFDHSRVILQGRDSNI<br>PGSDYINANYI--KN--Q-LLGPDENAKTYIASQGCLEATVNDWFQMAWQENS RVIV--MT-TREVEKGRNKCVPYWPE<br>VGMQRAYGYPYSVTNCGEHD TTEYKLR TLQVSPLDNGDLIREIWHYQYLSWPDHGVPSEPG-GVLSFLDQINQRQESLPH<br>AGPIIVHCSAGIGRTGTIIVIDMLMEN-----ISTKGLDCDIDIQKTIQMVRAQRSGMVQTEAQYKFIYVAIAQFI-ET<br>TKKKLEVL---DF--DY-E-KMANANKGAMTENA-DE-----NALQSDAKGKLDV-ATDY-----<br>-----GAAIDGFIGDVSGLANG-NGATGDF-----AGSNSQMAQVGD-----<br>-----GDNSPLMNNFRQYLP SLPQ-S-<br>--V-----E-----C--RPFVF-GAGKPYEFSI--DCDKINLFRGVFA----FLLYV-ATFMYVFST<br>FANILRN-KES----- |
| 10 | 0.4954 | 4.80 | 0.11 | 0.84 | <a href="#">2nlkA</a><br>Model1 | AIPVKQFVKHIGELYSNNQHGFSEDFEEVQRCTADMNITAEHSNHPENKHKKNRYINILA-----YDHSRVKLRPLPH<br>SDYINANYVDGYNKAKAYIATQGPKSTFEDFWRMIWEQNTGIIVMITN-LVEKGRRKCD--QYWPTENSEEYGNIIIVT<br>LKSTKIHACYTVR-RFSIR-NTKVKKGQKGNP--KGRQNERVVI--Q--YH--YT-QW-PDMG---V--PEYA-LPVL T<br>F-VRRSS-----AARMP-ETG-PV--LV--HCSAGVGRGTGTIYVIDSMLQIQKDKSTVNVLGFLKHIRTQRNYLVQ<br>TEEQYIFIHDALLEAILGK-----D<br>FDYEKMANANKGAM-----TENAD-----ENALQS-D-----A-KGKL-DSVATD<br>YGAAID-----G-----FIG-D-VS-----GLA---NGNGATGDFAGSNSQMAQVGD<br>GDNSPLMNNFRQYLP SLPQSVECRPFVFGAGKPYEFSIDCDKINLFRGVFAFLLY--VATFMYVFSTFANILRNKE-S-<br>-----      |

- (a) Ranking of proteins is based on TM-score of the structural alignment between Model1 and the PDB structures in our template library.
- (b) RMSDa is the RMSD between residues that are structurally aligned by TM-align.
- (c) IDENa is the percentage sequence identity in the structurally aligned region.

(d) Cov. represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the model.

#### Function Prediction

#### Predicted EC Numbers

| Rank | TM-score | RMSDa | IDENa | Cov. | EC-Score | PDB Hit               | EC No.                   |
|------|----------|-------|-------|------|----------|-----------------------|--------------------------|
| 1    | 0.6295   | 3.79  | 0.12  | 0.93 | 0.7454   | <a href="#">1toxA</a> | <a href="#">2.4.2.36</a> |
| 2    | 0.4835   | 4.64  | 0.12  | 0.81 | 0.5988   | <a href="#">2i5xA</a> | <a href="#">3.1.3.48</a> |
| 3    | 0.4631   | 4.98  | 0.12  | 0.79 | 0.5739   | <a href="#">2rdaF</a> | <a href="#">2.1.1.45</a> |
| 4    | 0.4671   | 5.28  | 0.11  | 0.83 | 0.5674   | <a href="#">2aazN</a> | <a href="#">2.1.1.45</a> |
| 5    | 0.4960   | 4.40  | 0.10  | 0.81 | 0.5647   | <a href="#">2b3oA</a> | <a href="#">3.1.3.48</a> |

- (a) Ranking is based on EC-score, which is a confidence score for the Enzyme Classification (EC) Number prediction.
- (b) RMSDa is the RMSD between models and the PDB structure in the structurally aligned regions by TM-align.
- (c) IDENa is percentage sequence identity in the structurally aligned region.
- (d) Cov. represents the coverage of the alignment and is equal to the number of structurally aligned residues divided by length of model.
- (e) EC-Score is defined based on the C-score of the structure prediction and similarity of the

model with known enzyme structures, as identified using both global and local structural alignment programs. The global similarity score uses TM-score, IDENa, RMSDa and Cov. of the structural alignment by TM-align, while the local match compares the structural and chemical similarity of local spatial motifs in the model with known catalytic site of enzymes. A prediction with a EC-score >1.1 signifies a prediction with high confidence (upto 3 digit numbers of EC) and vice versa (For details, see Ambrish, Kucukural and Zhang, Large scale benchmarking of protein function prediction using predicted protein structures, 2009, submitted).

#### Predicted GO terms

| Rank | TMscore | RMSDa | IDENa | Cov. | PDB Hit               | Fh-Score | Associated GO Terms                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------|---------|-------|-------|------|-----------------------|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | 0.6281  | 3.80  | 0.11  | 0.95 | <a href="#">1xdtT</a> | 0.52     | <a href="#">GO:0047286</a> <a href="#">GO:0009405</a> <a href="#">GO:0005576</a>                                                                                                                                                                                                                                                                                                                                                                                                 |
| 2    | 0.6281  | 3.80  | 0.11  | 0.95 | <a href="#">1xdtT</a> | 0.52     | <a href="#">GO:0047286</a> <a href="#">GO:0009405</a> <a href="#">GO:0005576</a>                                                                                                                                                                                                                                                                                                                                                                                                 |
| 3    | 0.4539  | 4.38  | 0.14  | 0.73 | <a href="#">1dejA</a> | 0.45     | <a href="#">GO:0005515</a> <a href="#">GO:0005524</a> <a href="#">GO:0030554</a> <a href="#">GO:0005622</a> <a href="#">GO:0005623</a> <a href="#">GO:0005737</a><br><a href="#">GO:0005856</a> <a href="#">GO:0043229</a>                                                                                                                                                                                                                                                       |
| 4    | 0.4049  | 5.41  | 0.14  | 0.74 | <a href="#">2q9oB</a> | 0.43     | <a href="#">GO:0005507</a> <a href="#">GO:0008471</a> <a href="#">GO:0043169</a> <a href="#">GO:0006725</a> <a href="#">GO:0009056</a> <a href="#">GO:0009808</a><br><a href="#">GO:0009987</a> <a href="#">GO:0019439</a> <a href="#">GO:0030243</a> <a href="#">GO:0030245</a> <a href="#">GO:0042219</a> <a href="#">GO:0043283</a><br><a href="#">GO:0043285</a> <a href="#">GO:0046274</a> <a href="#">GO:0055114</a> <a href="#">GO:0005576</a> <a href="#">GO:0043245</a> |
| 5    | 0.4879  | 4.66  | 0.11  | 0.80 | <a href="#">2h02A</a> | 0.43     | <a href="#">GO:0004725</a> <a href="#">GO:0004872</a> <a href="#">GO:0006470</a> <a href="#">GO:0009987</a> <a href="#">GO:0016311</a> <a href="#">GO:0034960</a><br><a href="#">GO:0043412</a>                                                                                                                                                                                                                                                                                  |
| 6    | 0.4631  | 4.98  | 0.12  | 0.79 | <a href="#">2rdaF</a> | 0.42     | <a href="#">GO:0004799</a> <a href="#">GO:0006231</a> <a href="#">GO:0009157</a> <a href="#">GO:0009162</a> <a href="#">GO:0009219</a> <a href="#">GO:0009221</a><br><a href="#">GO:0009262</a> <a href="#">GO:0009263</a> <a href="#">GO:0009987</a> <a href="#">GO:0044237</a> <a href="#">GO:0046073</a>                                                                                                                                                                      |
| 7    | 0.4900  | 3.76  | 0.12  | 0.73 | <a href="#">2g3hA</a> | 0.42     | <a href="#">GO:0005506</a> <a href="#">GO:0005515</a> <a href="#">GO:0019825</a> <a href="#">GO:0020037</a> <a href="#">GO:0043169</a> <a href="#">GO:0015671</a>                                                                                                                                                                                                                                                                                                                |
| 8    | 0.4863  | 3.75  | 0.12  | 0.73 | <a href="#">2bk9A</a> | 0.42     | <a href="#">GO:0005506</a> <a href="#">GO:0019825</a> <a href="#">GO:0020037</a> <a href="#">GO:0043169</a> <a href="#">GO:0015671</a>                                                                                                                                                                                                                                                                                                                                           |
| 9    | 0.4066  | 5.00  | 0.14  | 0.71 | <a href="#">1j2bB</a> | 0.42     | <a href="#">GO:0003723</a> <a href="#">GO:0008270</a> <a href="#">GO:0008479</a> <a href="#">GO:0043169</a> <a href="#">GO:0006399</a> <a href="#">GO:0008616</a><br><a href="#">GO:0009058</a> <a href="#">GO:0009451</a> <a href="#">GO:0009987</a> <a href="#">GO:0010467</a> <a href="#">GO:0034960</a> <a href="#">GO:0042278</a><br><a href="#">GO:0042451</a> <a href="#">GO:0043412</a> <a href="#">GO:0044249</a> <a href="#">GO:0046116</a> <a href="#">GO:0046118</a> |
| 10   | 0.3415  | 4.55  | 0.18  | 0.55 | <a href="#">2gwgA</a> | 0.41     | <a href="#">GO:0003824</a> <a href="#">GO:0008152</a>                                                                                                                                                                                                                                                                                                                                                                                                                            |

### Consensus Prediction of Gene Ontology terms

| Molecular Function         |          | Biological Process         |          | Cellular Location          |          |
|----------------------------|----------|----------------------------|----------|----------------------------|----------|
| GO term                    | GO-Score | GO term                    | GO-Score | GO term                    | GO-Score |
| <a href="#">GO:0003824</a> | 0.314    | <a href="#">GO:0008152</a> | 0.211    | <a href="#">GO:0005737</a> | 0.286    |
| <a href="#">GO:0005488</a> | 0.213    | <a href="#">GO:0044237</a> | 0.170    | <a href="#">GO:0044424</a> | 0.286    |
| <a href="#">GO:0016740</a> | 0.188    | <a href="#">GO:0009987</a> | 0.170    | <a href="#">GO:0005623</a> | 0.286    |
| <a href="#">GO:0043169</a> | 0.168    | <a href="#">GO:0044238</a> | 0.170    | <a href="#">GO:0044464</a> | 0.286    |
| <a href="#">GO:0043167</a> | 0.168    | <a href="#">GO:0043170</a> | 0.127    | <a href="#">GO:0005622</a> | 0.286    |
| <a href="#">GO:0046914</a> | 0.168    | <a href="#">GO:0043283</a> | 0.127    | <a href="#">GO:0005576</a> | 0.186    |
| <a href="#">GO:0046872</a> | 0.168    | <a href="#">GO:0015671</a> | 0.125    | <a href="#">GO:0044444</a> | 0.162    |
| <a href="#">GO:0016763</a> | 0.145    | <a href="#">GO:0051234</a> | 0.125    | <a href="#">GO:0032991</a> | 0.121    |
| <a href="#">GO:0016757</a> | 0.145    | <a href="#">GO:0006810</a> | 0.125    | <a href="#">GO:0044445</a> | 0.121    |
| <a href="#">GO:0047286</a> | 0.103    | <a href="#">GO:0015669</a> | 0.125    | <a href="#">GO:0005833</a> | 0.121    |

- (a) Ranking in the first table is based on a function prediction score (Fh-score), which is calculated based on the C-score of the structure prediction and the TM-score, IDENa, RMSDa and Cov. of the structural alignment by TM-align between the predicted model and the PDB structures (For details, see Ambrish, Kucukural and Zhang, Large scale benchmarking of protein function prediction using predicted protein structures, 2009, submitted).
- (b) RMSDa is the RMSD between models and the PDB structure in the structurally aligned regions by TM-align.
- (c) IDENa is the percentage sequence identity in the structurally aligned region.
- (d) Cov. represents the coverage of the alignment and is equal to the number of structurally aligned residues divided by length of model.
- (e) A consensus prediction of GO terms is derived from the structural analogs that have an Fh-score of  $\geq 1.0$ . The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on the Fh-score of the template from which the GO term is derived. A prediction with a GO-score  $> 0.5$  signifies a prediction with high confidence and vice versa.

## **Appendix 5 - Reference CD**

## References

- Arnold, K., Bordoli, L., Kopp, J., Schwede, T. 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22**:195-201
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knock-out mutants -- the Keio collection. *Mol Systems Biol* **2**
- Bamford, D.H., Palva, E.T., Lounatmaa, K. 1976. Ultrastructure and life cycle of the lipid-containing bacteriophage phi 6. *J Gen Virol* **32**:249-59
- Bamford, D.H., Romantschuk, M., Somerharju, P.J. 1987. Membrane fusion in prokaryotes: bacteriophage phi 6 membrane fuses with the *Pseudomonas syringae* outer membrane. *EMBO* **6**:1467-73
- Barbas, C.F., Burton, D.R., Scott, J.K., Silverman, G.J. 2004. Phage Display: A Laboratory Manual. CSHL Press
- Barbas III, C.F., Burton, D.R., Scott, J.K., Silverman, G.J. 2001. Phage display: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Beck, E., Zink, B. 1981. Nucleotide sequence and genome organisation of filamentous bacteriophages fl and fd. *Gene* **16**:35-58
- Bennett, N. 2005. Unlocking the M13 (f1) Virion. Investigation into the role of pIII C domain of F specific Filamentous Bacteriophage in Infection. *In*: IMBS. Massey University, Palmerston North
- Bennett, N.J., Rakonjac, J. 2006. Unlocking of the filamentous bacteriophage virion during infection is mediated by the C domain of pIII. *J Mol Biol* **356**:266-73
- Blake, M.S., Johnsten, K.H., Russel-Jones, G.J., Gotschlich, E.C. 1984. A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on Western Blots. *Anal. Biochem.* **136**:175-179
- Blumer, K.J., Ivey, M.R., Steege, D.A. 1987. Translational control of fl gene expression by differential activities of the gene V, VII, IX and VIII initiation sites. *J. Mol. Biol.* **197**:439-451

- Boeke, J.D., Model, P. 1982. A procaryotic membrane anchor sequence: Carboxyl terminus of bacteriophage  $\phi$ 1 gene III protein retains it in the membrane. *Proc. Natl. Acad. Sci. USA* **79**:5200-5204
- Boeke, J.D., Model, P., Zinder, N.D. 1982. Effects of bacteriophage  $\phi$ 1 gene III protein on the host cell membrane. *Mol. Gen. Genet.* **186**:185-192
- Bradley, D.E., Whelan, J. 1989. *Escherichia coli* tolQ mutants are resistant to filamentous bacteriophages that adsorb to the tips, not the shafts, of conjugative pili. *J. Gen. Microbiol.* **135**:1857-1863
- Brasseur, R., Cornet, B., Burny, A., Vandenbranden, M., Ruyschaert, J.M. 1988. Mode of Insertion into a Lipid Membrane of the N-Terminal HIV gp41 Peptide Segment. *AIDS Res Hum Retroviruses* **4**:83-90
- Brissette, J.L., Russel, M., Weiner, L., Model, P. 1990. Phage shock protein, a stress protein of *Escherichia coli*. *PNAS* **87**:862-866
- Brissette, J.L., Weiner, L., Ripmaster, T.L., Model, P. 1991. Characterization and sequence of the *Escherichia coli* stress-induced *psp* operon. *Journal of Molecular Biology* **220**:35-48
- Bullough, P.A., Hughson, F.M., Skehel, J.J., Wiley, D.C. 1994. Structure of Influenza Haemagglutinin at the pH of Membrane fusion. *Nature Structural and Molecular Biology* **371**:37-43
- Cadwell, R.C., Joyce, G.F. 1992. Randomization of Genes by PCR mutagenesis. *Genome Research* **2**:28-33
- Caldentey, J., Bamford, D.H. 1992. The lytic enzyme of the *Pseudomonas* phage  $\phi$ 6. Purification and biochemical characterization. *Biochim Biophys Acta* **1159**:44-50
- Calendar, R. 2005. The bacteriophages. Oxford University Press, New York
- Caro, L.G., Schnös, M. 1966. The attachment of the male-specific bacteriophage  $\phi$ 1 to sensitive strains of *E. coli*. *Proc. Natl. Acad. Sci. USA* **56**:126-132
- Carter, P., Bedouelle, H., Winter, G. 1985. Improved oligonucleotide site-directed mutagenesis using M13 vectors. *Nucleic Acids Res* **13**:4431-4443
- Chan, T.S., Model, P., Zinder, N.D. 1975. In vitro protein synthesis directed by separated transcripts of bacteriophage  $\phi$ 1 DNA. *J Mol Biol* **99**:369-82
- Chang, C.N., Model, P., Blobel, G. 1978. Detection of prokaryotic signal peptidase in *Escherichia coli* membrane fraction: endoproteolytic cleavage of nascent  $\phi$ 1 pre-coat protein. *Proc. Natl. Acad. Sci. USA* **75**:361-365

- Chopin, M.-C., Rouault, A., Ehrlich, S.D., Gautier, M. 2002. Filamentous Phage Active on the Gram Positive Bacterium *Propionibacterium freudenreichii*. *Journal of Bacteriology* **184**:2032-2033
- Click, E.M., Webster, R.E. 1997. Filamentous phage infection: Required interactions with the TolA protein. *J. Bacteriol.* **179**:6464-6471
- Click, E.M., Webster, R.E. 1998. The TolQRA proteins are required for membrane insertion of the major capsid protein of the filamentous phage  $\phi$ 1 during infection. *J. Bacteriol.* **180**:1723-1728
- Coburn, B., Sekirov, I., Finlay, B.B. 2007. Type III secretion systems and disease. *Clin Microbiol Rev* **20**:535-49
- Cole, C., Barber, J.D., Barton, G.J. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Research* **36**:W197-W201
- Collins, R.F., Frye, S.A., Balasingham, S., Ford, R.C., Tonjum, T., Derrick, J.P. 2005. Interaction with type IV pili induces structural changes in the bacterial outer membrane secretin PilQ. *J Biol Chem* **280**:18923-30
- Colman, P.M., Lawrence, M.C. 2003. The structural biology of type I viral membrane fusion. *Nature Reviews Molecular Cell Biology* **4**
- Craig, L., Pique, M.E., Tainer, J.A. 2004. Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* **2**:363-78
- Cramer, R., Jaussi, R., Menz, G., Blaser, K. 1994. Display of expression products of cDNA libraries on phage surfaces. A versatile screening system for selective isolation of genes by specific gene-product/ligand interaction. *Eur J Biochem* **226**:53-8
- Crissman, J.W., Smith, G.P. 1984. Gene-III protein of filamentous phages: evidence for a carboxyl-terminal domain with a role in morphogenesis. *Virology* **132**:445-455
- Cunningham, B.C., Wells, J.A. 1989. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244**:1081-1085.
- Datsenko, K.A., Wanner, B.L. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci* **97**:6640-6645
- Davis, B.M., Lawson, E.H., Sandkvist, M., Ali, A., Sozhamannan, S., Waldor, M.K. 2000. Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTX $\Phi$ . *Science* **288**:664-670
- Davis, B.M., Waldor, M.K. 2003. Filamentous phages linked to virulence of *Vibrio cholerae*. *Curr Opin Microbiol* **6**:35-42

- Davis, N.G. 1985. The membrane integration of gene III protein: design of the anchor domain. The Rockefeller University
- Davis, N.G., Boeke, J.D., Model, P. 1985. Fine structure of a membrane anchor domain. *J. Mol. Biol.* **181**:111-121
- Davis, N.G., Model, P. 1985. An artificial anchor domain: hydrophobicity suffices to stop transfer. *Cell* **41**:607-614
- Day, L.A., Marzec, C.J., Reisberg, S.A., Casadevall, A. 1988. Dna packing in filamentous bacteriophages *Ann. Rev. Biophys. Biophys. Chem.* **17**:509-539
- Day, L.A., Wiseman, R.L., Marzec, C.J. 1979. Structure models for DNA in filamentous viruses with phosphates near the center. *Nucleic Acids Research* **7**:1393-1403
- DeLano, W. 2006. PyMol
- Deng, L.W., Perham, R.N. 2002. Delineating the site of interaction on the pIII protein of filamentous bacteriophage fd with the F-pilus of Escherichia coli. *J Mol Biol* **319**:603-14
- Deprez, C., Lloubes, R., Gavioli, M., Marion, D., Guerlesquin, F., Blanchard, L. 2005. Solution Structure of the E.coli TolA C-terminal Domain Reveals Conformational Changes upon Binding to the Phage g3p N-Terminal Domain. *Journal of Molecular Biology* **346**:1047-1057
- Dotto, G.P., Enea, V., Zinder, N.D. 1981. Functional analysis of bacteriophage intergenic region. *Virology* **114**:463-473
- Dotto, G.P., Horiuchi, K., Zinder, N. 1984. The origin of DNA replication of Bacteriophage fl and its interaction with the phage gene II protein. *In: Proteins involved in DNA replication.* H. U and S. S, editors. pp. 185-191. Plenum Publishing Corporation
- Dotto, G.P., Horiuchi, K., Zinder, N.D. 1982. Initiation and Termination of Phage F1 plus-strand synthesis. *Proc Natl Acad Sci U S A* **79**:7122-7126
- Dotto, G.P., Zinder, N.D. 1983. The morphogenetic signal of bacteriophage fl. *Virology* **130**:252-256
- Dotto, G.P., Zinder, N.D. 1984. Reduction of the minimal sequence for initiation of DNA synthesis by qualitative or quantitative changes of an initiator protein. *Nature* **311**:279-280
- Dyson, H.J., Wright, P.E. 2005. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* **6**:197-208

- Eckert, B., Martin, A., Balbach, J., Schmid, F.X. 2007. Prolyl Isomerization as a Molecular Timer in Phage Infection. *Nature Structural and Molecular Biology* **12**:619-623
- Eckert, B., Schmid, F.X. 2007. A conformational unfolding reaction activates phage fd for the infection of Escherichia coli. *Journal of Molecular Biology* **373**:452-461
- Efron, B. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7**:1-26
- Efron, B. 1981. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika* **68**:589-599
- Endemann, H., Model, P. 1995. Location of filamentous phage minor coat proteins in phage and in infected cells. *J. Mol. Biol.* **250**:496-506
- Enea, V., N.D Zinder. 1982. Interference resistant mutants of phage fl. *Virology* **112**:222-226
- Feng, J.-n., Model, P., Russel, M. 1999. A trans-envelope protein complex needed for filamentous phage assembly and export. *Mol. Microbiol.* **34**:745-755
- Feng, J.N., Russel, M., Model, P. 1997. A permeabilized cell system that assembles filamentous bacteriophage. *Proc. Natl. Acad. Sci. USA* **94**:4068-4073
- Furuta, R.A., Wild, C.T., Weng, Y., Weiss, C.D. 1998. Capture of an early fusion-active conformation of HIV-1 gp41. *Nat. Struct. Biol.* **5**:276-279
- Gailus, V., Rasched, I. 1994. The adsorption protein of bacteriophage fd and its neighbour minor coat protein build a structural entity. *Eur. J. Biochem.* **222**:927-931
- Galan, J.E., Collmer, A. 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**:1322-1328
- Gallaher, W.R. 1987. Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. *Cell* **50**:327-8
- Garner, E., Romero, P., Dunker, A., Brown, C., Obradovic, Z. 1999. Predicting binding regions within disordered proteins. *Genome Informatics* **10**:41-50
- Glaser-Wuttke, G., Keppner, J., Rasched, I. 1989. Pore-forming properties of the adsorption protein of filamentous phage fd. *Biochim. Biophys. Acta* **985**:239-247
- Goldsmith, M.E., Konigsberg, W.H. 1977. Adsorption protein of bacteriophage fd - isolation, molecular-properties, and location in virus. *Biochemistry* **16**:2686-2694

- Gonzalez, M.D., Lichtensteiger, C.A., Caughlan, R., Vimr, E.R. 2002. Conserved Filamentous Prophage in *Escherichia coli* O18:K1:H7 and *Yersinia pestis* Biovar orientalis. *J Bacteriol* **184**:6050-6055
- Grant, R., Lin, T., Webster, R., Konigsberg, W. 1980. Structure of filamentous bacteriophage: isolation, characterization, and localization of the minor coat proteins and orientation of the DNA. *In: Bacteriophage assembly*. M. DuBow, editor. pp. 413-428. Alan. R. Liss, Inc., New York
- Gray, C.P., Sommer, R., Polke, C., Beck, E., Schaller, H. 1978. Structure of origin of dna-replication of bacteriophage fd *Proceedings Of The National Academy Of Sciences Of The United States Of America* **75**:50-53
- Gray, C.W., Brown, R.S., Marvin, D.A. 1979. Direct visualization of adsorption protein of fd phage. *J Supramol Str* **1979**:91-91
- Gray, C.W., Brown, R.S., Marvin, D.A. 1981. Adsorption complex of filamentous fd virus. *J. Mol. Biol.* **146**:621-627
- Greener, A., Callahan, M. 1994. An Efficient Random Mutagenesis Technique Using an *E. coli* Mutator Strain *In: Methods in Molecular Biology*. Humana Press
- Guex, N., Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* **18**:2714-2723
- Heery, D.M., Gannon, H.F., Powell, R. 1990. A simple method for subcloning DNA fragments from gel slices. *Trends in Genetics* **6**:173
- Heilpern, A.J., Waldor, M.K. 2000. CTXphi infection of *Vibrio cholerae* requires the tolQRA gene products. *J Bacteriol* **182**:1739-47
- Hershey, A.D., Chase, M. 1952. Independent Functional of Viral Proteins and Nucleic Acid in Growth of Bacteriophage. *Journal of Gen Physiol* **36**:39-56
- Hill, D.F., Petersen, G.B. 1982. Nucleotide sequence of bacteriophage f1 DNA. *Journal Of Virology* **44**:32-46
- Hoffman-Berling, H., Durwald, H., Beulke, I. 1963. Ein fadiger DNS-phage (fd) und ein spharischer RNS-phage (fr) wirtsspezifisch für männliche Stämme von *E. coli*, III. Biologisches Verhalten von fd und fr. *Fadiger DNS-Phage (fd) und Spharischer RNS-Phage (fr) III*:893-898
- Hofschneider, P.H. 1963. Untersuchungen über "kleine" *E. coli* K12 Bakteriophagen. *Z. Naturforschung* **18B**:203-210

- Holliger, P., Riechmann, L., Williams, R.L. 1999. Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 Å: evidence for conformational lability. *J. Mol. Biol.* **288**:649-657
- Hu, G.B., Wei, H., Rice, W.J., Stokes, D.L., Gottlieb, P. 2008. Electron cryotomographic structure of cystovirus phi 12. *Virology* **1**:1-9
- Huber, K.E., Waldor, M.K. 2002. Filamentous phage integration requires the host recombinases XerC and XerD. *Nature* **417**:656-659
- Hunter, G.J., Rowitch, D.H., Perham, R.N. 1987. Interactions between DNA and coat protein in the structure and assembly of filamentous bacteriophage fd. *Nature* **327**
- ICTVdB. 2008. ICTVdB Index of Viruses
- Jacob, E., Hofschneider, P.H. 1969. Replication of the single-stranded DNA bacteriophage M13: messenger RNA synthesis directed by M13 replicative form DNA. *J Mol Biol* **46**:359-63
- Jankovic, D. 2008. Direct Selection and Phage Display of the *Lactobacillus rhamnosus* HN001 secretome. *In: Institute of Molecular Biosciences. Massey University, Palmerston North*
- Johnson, T.L., Abendroth, J., Hol, W.G., Sandkvist, M. 2006. Type II secretion: from structure to function. *FEMS Microbiol Lett* **255**:175-86
- Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F., Rossmann, M.G. 2002. Structure of the Cell-Puncturing Device of Bacteriophage T4. *Nature Structural and Molecular Biology* **415**
- Kazmierczak, B., Mielke, D.L., Russel, M., Model, P. 1994. pIV, a filamentous phage protein that mediates phage export across the bacterial cell envelope, forms a multimer. *J Mol Biol* **238**:187-198
- Khatoon, H., Iyer, R.V., Iyer, V.N. 1972. A new filamentous bacteriophage with sex-factor specificity. *Virology* **48**:145-155
- Kim, A.Y., Blaschek, H.P. 1991. Isolation and characterization of a filamentous viruslike particle from *Clostridium acetobutylicum* NCIB 6444. *J Bacteriol* **173**:530-5
- Kim, M.H., Hines, J.C., Ray, D.S. 1981. Viable deletions of the M13 complementary strand origin. *Proc Natl Acad Sci* **78**:6784-6788

- Konings, R.N.H. 1973. Synthesis of phage M13 specific proteins in a DNA-dependent cell-free system. *FEBS Letters* **35**:155-160
- Kopp, J., Schwede, T. 2004. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res* **32**:D230-D234
- Korte, T., Ludwig, K., Booy, F., Blumenthal, R., Herrmann, A. 1999. Conformational intermediates and fusion activity of influenza virus hemagglutinin. *J Virol.* **73**:4567-74
- Kremser, A., Rasched, I. 1994. The adsorption protein of filamentous phage fd: assignment of its disulfide bridges and identification of the domain incorporated in the coat. *Biochemistry* **33**:13954-13958
- Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galan, J.E., Aizawa, S.I. 1998. Supramolecular structure of the *Salmonella* typhimurium type III protein secretion system. *Science* **280**:602-5
- Kyte, J., Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**:105-132
- La Farina, M., Izzo, V., Duro, G., Barbieri, R., Assunta Costa, M., Mutolo, V. 1987. Intragenomic recombination between homologous regions of genes II and IV promotes formation of bacteriophage  $\phi$ 1 miniphages. *Nucl. Acids Res.* **15**:7190
- Laemmli, U. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680-685
- Lee, S., Mao, C., Flynn, C., Belcher, A. 2002. Ordering of quantum dots using genetically engineered viruses. *Science* **296**:892-895
- Lessl, M., Balzer, D., Lurz, R., Waters, V.L. 1992. Dissection of IncP conjugative plasmid transfer: definition of the transfer region Tra2 by mobilization of the Tra1 region *in trans*. *J. Bacteriol.* **174**:2493-2500
- Li, X., Romero, P., Rani, M., Dunker, A., Obradovic, Z. 1999. Predicting protein disorder for N-, C-, and internal regions,. *Genome Informatics* **10**:30-40
- Linderoth, N.A., Model, P., Russel, M. 1996. Essential role of a sodium dodecyl sulfate-resistant protein IV multimer in assembly-export of filamentous phage. *J. Bacteriol.* **178**:1962-1970
- Linderoth, N.A., Simon, M.N., Russel, M. 1997. The filamentous phage pIV multimer visualized by Scanning Transmission Electron Microscopy. *Science* **278**:1635-1638

- Liu, D.J., Day, L.A. 1994. Pfl virus structure: helical coat protein and DNA with paraxial phosphates. *Science* **265**:671-4
- Loeb, T. 1960. Isolation of a bacteriophage specific for the F<sup>+</sup> and H<sup>r</sup> mating types of *Escherichia coli* K12. *Science* **131**:932-933
- Lopez, J., Webster, R.E. 1983. Morphogenesis of filamentous bacteriophage  $\phi$ 1: orientation of extrusion and production of polyphage. *Virology* **127**:177-193
- Lowman, H.B., Bass, S.H., Simpson, N., Wells, J.A. 1991. Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* **30**:10832-8
- Lu, M., Blacklow, S.C., Kim, P.S. 1995. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nat Struct Biol* **2**:1075-82
- Lubkowski, J., Hennecke, F., Pluckthun, A., Wlodawer, A. 1998. The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. *Nat. Struct. Biol.* **5**:140-7
- Lubkowski, J., Hennecke, F., Pluckthun, A., Wlodawer, A. 1999. Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure Fold. Des.* **15**:711-272
- Maeda, T., Ohnishi, S. 1980. Activation of influenza virus by acidic media causes hemolysis and fusion of erythrocytes. *FEBS letters* **122**:283-7
- Makowski, L. 1993. Structural constraints on the display of foreign peptides on filamentous bacteriophages. *Gene* **128**:5-11
- Marciano, D.K., Russel, M., Simon, S.M. 2001. Assembling filamentous phage occlude pIV channels. *Proc. Natl. Acad. Sci. USA* **98**:9359-9364
- Martin, A., Schmid, F.X. 2003a. The folding mechanism of a two-domain protein: folding kinetics and domain docking of the gene-3 protein of phage  $\phi$ d. *J. Mol. Biol.* **329**:599-610
- Martin, A., Schmid, F.X. 2003b. A Proline Switch Controls Folding and Domain Interactions in the Gene-3-protein of the Filamentous Phage  $\phi$ d. *J Mol Biol* **331**:1131-1140
- Marvin, D., Hoffmann-Berling, H. 1963. Physical and chemical properties of two new small bacteriophages. *Nature* **197**:517-518
- Marvin, D.A. 1998. Filamentous phage structure, infection and assembly. *Curr Opin Struct Biol* **8**:150-8

- Marvin, D.A., Hale, R.D., Nave, C., Helmer-Citterich, M. 1994. Molecular models and structural comparisons of native and mutant class I filamentous bacteriophages Ff (fd, fl, M13), Ifl and IKe. *J Mol Biol* **235**:260-86
- Marvin, D.A., Welsh, L.C., Symmons, M.F., Scott, W.R., Straus, S.K. 2006. Molecular Structure of fd(fl, M13) filamentous bacteriophage refined with resects to X-ray fibre diffraction and solid-state NMR data supports specific models of phage assembly at the bacterial membrane. *Journal of Molecular Biology* **355**:294-309
- McCafferty, J., Griffiths, A.D., Winter, G., Chiswell, D.J. 1990. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**:552-554
- McLeod, S.M., Kimsey, H.H., Davis, B.M., Waldor, M.K. 2005. CTXphi and *Vibrio cholerae*: exploring a newly recognized type of phage-host cell relationship. *Mol Microbiol* **57**:347-56
- Meyer, Geider. 1979. Bacteriophage fd gene II-protein. II. Specific cleavage and relaxation of supercoiled RF from filamentous phages. *J. Biol. Chem.* **252**:12642-12646
- Meyer, T.F., Geider, K. 1982. Enzymatic synthesis of bacteriophage fd viral DNA. *Nature* **296**:828-32
- Meynell, G.G., Lawn, A.M. 1968. Filamentous phages specific for the I sex factor. *Nature* **217**:1184-6
- Michel, B., Zinder, N.D. 1989. Translational repression in bacteriophage fl: characterization of the gene V protein target on the gene II mRNA. *Proc Natl Acad Sci* **86**:4002-6
- Model, P., Jovanovic, G., Dworkin, J. 1997. The *Escherichia coli* phage shock protein operon. *Molecular Microbiology* **24**:255-261
- Model, P., McGill, C., Mazur, B., Fulford, W.D. 1982. The replication of bacteriophage fl gene V protein regulates the synthesis of gene II protein *Cell* **29**:329-335
- Model, P., Russel, M. 1988. Filamentous Bacteriophage. *In: The Bacteriophages*. R. Calendar, editor. pp. 375-456. Plenum Publishing, New York
- Model, P., Zinder, N.D. 1974. In vitro synthesis of Bacteriophage fl Proteins. *J. Mol. Biol.* **83**:231-251

- Mooij, M.J., Drenkard, E., Llamas, M.A., Vandenbroucke-Grauls, C.M., Savelkoul, P.H., Ausubel, F.M., Bitter, W. 2007. Characterization of the integrated filamentous phage Pf5 and its involvement in small-colony formation. *Microbiology* **153**:1790-8
- Nambudripad, R., Start, W., Opella, S.J., Makowski, L. 1991. Membrane-mediated assembly of filamentous bacteriophage Pfl coat protein. *Science* **252**:1305-1308
- Nasu, H., Iida, T., Sugahara, T., Yamaichi, Y., Park, K.S., Yokoyama, K., Makino, K., Shinagawa, H., Honda, T. 2000. A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. *J Clin Microbiol* **38**:2156-61
- Nelson, F.K., Friedman, S.M., Smith, G.P. 1981. Filamentous phage DNA cloning vectors - a non-infective mutant with a non-polar deletion in gene III. *Virology* **108**:338-350
- O'Neil, E.K., Grisham, C.M. 1997. Java Helical Wheel Applet. pp. Helical Wheel drawing Java Applet. University of Virginia, Charlottesville
- O'Neil, K., Hoess, R.H. 1995. Phage display: protein engineering by directed evolution. *Current Opinion in Structural Biology* **5**:443-449
- Okamoto, T., Sugimoto, K., Sugisaki, H., Takanami, M. 1977. DNA regions essential for the function of a bacteriophage fd promoter. *Nucleic Acids Res* **4**:2213-22
- Opalka, N., Simon, M.N., Linderoth, N.A., Model, P., Darst, S.A., Russel, M. 2002. Organization of filamentous phage pIV multimers. *In preparation*
- Overman, S.A., Thomas, G.J. 1995. Raman Spectroscopy of filamentous virus Ff (fd, fl and M13): Structural interpretations for coat proteins aromatics. *Biochemistry* **34**:5440-5451
- Papavoine, C.H., Christiaans, B.E., Folmer, R.H., Konings, R.N., Hilbers, C.W. 1998. Solution structure of the M13 major coat protein in detergent micelles: a basis for a model of phage assembly involving specific residues. *J Mol Biol* **282**:401-19
- Peitsch, M.C. 1995. Protein modeling by E-mail. *Bio/Technology* **13**:658-660
- Pratt, D., Tzagoloff, H., Beaudoin, J. 1969. Conditional lethal mutants of the small filamentous coliphage, II. Two genes for coat proteins. *Viol.* **39**:42-53
- Pratt, D., Tzagoloff, H., Erdahl, W.S. 1966. Conditional lethal mutants of the small filamentous coliphage M13. I. Isolation, complementation, cell killing, time of cistron action. *Virology* **30**:397-410

- Rakonjac, J. 1998. The roles of pIII in termination of filamentous phage assembly. Rockefeller University, New York
- Rakonjac, J. 2006. Personal Communication
- Rakonjac, J. unpublished research.
- Rakonjac, J., Feng, J.-n., Model, P. 1999. Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. *J. Mol. Biol.* **289**:1253-1265
- Rakonjac, J., Jovanovic, G., Model, P. 1997. Filamentous phage infection-mediated gene expression: construction and propagation of the *gIII* deletion mutant helper phage R408d3. *Gene* **198**:99-103
- Rakonjac, J., Model, P. 1998. The roles of pIII in filamentous phage assembly. *J. Mol. Biol.* **282**:25-41
- Rapoza, M.P., Webster, R.E. 1995. The products of gene I and the overlapping in-frame gene XI are required for filamentous phage assembly. *J. Mol. Biol.* **248**:627-638
- Reichmann, L., Holliger, P. 1997. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. *Cell* **90**:351-360
- Romantschuk, M., Olkkonen, V.M., Bamford, D.H. 1988. The nucleocapsid of bacteriophage Phi6 penetrates the host cytoplasmic membrane. *EMBO J.* **7**:1821-1829
- Romero, P., Obradovic, Z., Dunker, A. 1997. Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Informatics* **8**:110-124
- Romero, P., Obradovic, Z., Li, X., Garner, E., Brown, C., Dunker, A. 2001. Sequence complexity of disordered protein. *Proteins: Struct. Funct. Gen* **42**:38-48
- Root, M.J., Kay, M.S., Kim, P.S. 2001. Protein design of an HIV-1 entry inhibitor. *Science* **291**:884-888
- Russel, M. 1993. Protein-protein interactions during filamentous phage assembly. *J. Mol. Biol.* **231**:689-697
- Russel, M., Kazmierczak, B. 1993. Analysis of the structure and subcellular location of filamentous phage pIV. *J Bacteriol* **175**:3998-4007
- Russel, M., Kidd, S., Kelley, M.R. 1986. An improved filamentous helper phage for generating single-stranded plasmid DNA. *Gene* **45**:333-338
- Russel, M., Model, P. 1982. Filamentous phage pre-coat is an integral membrane protein: analysis by a new method of membrane preparation. *Cell* **28**:177-184

- Russel, M., Model, P. 1989. Genetic analysis of the filamentous bacteriophage packaging signal and of the proteins that interact with it. *J. Virol.* **63**:3284-3295
- Salstrom, J.S., Pratt, D. 1971. Role of coliphage M13 gene 5 in single-stranded DNA production. *J Mol Biol* **61**:489-501
- Sambrook, J., Fritsch, E.F., Maniatis, T. 1989. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Samuelson, J., Chen, M., Jiang, F., Möller, I., Wiedmann, M., Kuhn, A., Phillips, G., Dalbey, R. 2000. YidC mediates membrane protein insertion in bacteria. *Nature* **406**:637-41
- Schägger, H., von Jagow, G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamid gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166**:168-379
- Schwede, T., Kopp, J., Guex, N., Peitsch, M.C. 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* **31**:3381-3385
- Smilowitz, H. 1974a. Bacteriophage fl infection and colicin tolerance. *J. Virol.* **13**:100-106
- Smilowitz, H. 1974b. Bacteriophage fl infection: fate of the parental major coat protein. *J Virol.* **13**:94-9
- Smilowitz, H., Carson, J.A., Robbins, P.W. 1972. Association of newly synthesized major fl coat protein with infected host cell inner membrane. *J. Supramolec. Structure* **1**:8-18
- Smith, D.E., Tans, S.J., Smith, S.B., Grimes, S., Anderson, D.L., Bustamante, C. 2001. The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* **413**:748-52
- Smith, G. 2005. Large Scale Purification of Wild Type Virions using Sarkosyl. *In: Bio Online Site. bioProtocol, editor. www.bio.com*
- Smith, G.P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**:1315-1317
- Speckthrie, L., Bullitt, E., Horiuchi, K., Model, P., Russel, M., Makowski, L. 1992. Construction of a microphage variant of filamentous bacteriophage. *J. Mol. Biol.* **228**:720-724
- Stengele, I., Bross, P., Garces, X., Giray, J., Rasched, I. 1990. Dissection of functional domains in phage fd adsorption protein - discrimination between attachment and penetration sites. *J. Mol. Biol.* **212**:143-149

- Stiasny, K., Allison, S., Schalich, J., Heinz, F. 2002. Membrane interactions of the tick-borne encephalitis virus fusion protein C at low pH. *J Virol.* **76**:3784-3790
- Sun, T.-P., Webster, R.E. 1987. Nucleotide sequence of a gene cluster involved in entry of E colicins and single-stranded DNA of infecting filamentous bacteriophages into *Escherichia coli*. *J.Bacteriol.* **169**:2667-2674
- Takeya, K., Amako, K. 1966. A rod-shaped *Pseudomonas* phage. *Virology* **28**:163-165
- Tan, K., Liu, J., Wang, J., Shen, S., Lu, M. 1997. Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc Natl Acad Sci* **94**:12303-8
- Tan, W.M., Jelinek, R., Opella, S.J., Malik, P., Terry, T.D., Perham, R.N. 1999. Effects of temperature and Y21M mutation on conformational heterogeneity of the major coat protein (pVIII) of filamentous bacteriophage fd1. *Journal of Molecular Biology* **286**:787-796
- Trenkner, E., Bonhoeffer, F., Gierer, A. 1967. The fate of the protein component of bacteriophage fd during infection. *Biochem Biophys Res Commun* **28**:932-9
- van Wezenbeek, P.M.G.F., Hulsebos, T.J.M., Schoenmakers, J.G.G. 1980. Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. *Gene* **11**:129-148
- Vieira, J., Messing, J. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic primers. *Gene* **19**:259-268
- Waldor, M., Mekalanos, J.J. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**:1910-1914
- Wang, Y.A., Yu, X., Overman, S., Tsuboi, M., Jr, G.J.T., Egelman, E.H. 2006. The Structure of a Filamentous Bacteriophage. *Journal of Molecular Biology* **361**:209-215
- Wanner, B., Kodaira, R., Neidhardt, F. 1977. Physiological regulation of a decontrolled lac operon. *J Bacteriol* **130**:212-22
- Webb, J.S., Lau, M., Kjelleberg, S. 2004. Bacteriophage and phenotypic variation in *Pseudomonas aeruginosa* biofilm development. *J Bacteriol* **186**:8066-73
- Webster, R.E. 1996. Biology of the filamentous bacteriophage. *In*: Phage display of peptides and proteins. B.K. Kay, J. Winter, and J. McCafferty, editors. pp. 1-20. Academic Press, London, New York
- Weiner, L. 1993. Regulation and function of the phage shock proteins of *Escherichia coli*. The Rockefeller University

- Weiner, L., Brissette, J.L., Model, P. 1991. Stress-induced expression of the *Escherichia coli* phage shock protein operon is dependent on  $\sigma_{54}$  and modulated by positive and negative feedback mechanisms. *Genes & Development* **5**:1912-1923
- Weiss, G.A., Roth, T.A., Baldi, P.F., Sidhu, S.S. 2003. Comprehensive mutagenesis of the C-terminal domain of the M13 gene-3 minor coat protein: the requirements for assembly into the bacteriophage particle. *J Mol Biol* **332**:777-82
- Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J., Wiley, D.C. 1997. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**:346-8
- White, J.M. 1992. Membrane Fusion. *Science* **258**:917-24
- Wickner, W. 1975. Asymmetric orientation of a phage coat protein in cytoplasmic membrane of *E. coli*. *Proc. Natl. Acad. Sci. USA* **72**:4749-4753
- Wiley, D.C., Skehel, J.J. 1987. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.* **56**:365-94
- Williams, K.A., Glibowicka, M., Li, Z., Li, H., Khan, A.R., Chen, Y.M., Wang, J., Marvin, D.A., Deber, C.M. 1995. Packing of coat protein amphipathic and transmembrane helices in filamentous bacteriophage M13: role of small residues in protein oligomerization. *J Mol Biol* **252**:6-14
- Woolford, J.L., Steinman, H.M., Webster, R.E. 1977. Adsorption Protein of Bacteriophage  $\phi$ 1: Solubilization in Deoxycholate and Localization in the  $\phi$ 1 Virion. *Biochemistry* **16**:2694-2700
- Young, R., Bläsi, U. 1995. Holins: form and function in bacteriophage lysis. *FEMS Microbiol Rev* **17**:191-205
- Zacher, A.N., Stock, C.A., Golden, J.W., Smith, G.P. 1980. A new filamentous phage cloning vector- $\phi$ d-tet  
*Gene* **9**:127-140
- Zenkin, N., Naryshkina, T., Kuznedelov, K., Severinov, K. 2006. The mechanism of DNA replication primer synthesis by RNA polymerase. *Nature* **439**:617-20
- Zhang, Y. 2007. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins Suppl vol 8*:108-117
- Zhang, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**
- Zinder, N.D., Boeke, J.D. 1982. The filamentous phage (F $\phi$ ) as vectors for recombinant DNA—a review. *Gene* **19**:1-10

- Zinder, N.D., Horiuchi, K. 1985. Multiregulatory elements of filamentous bacteriophages. *Microbiol. Rev.* **49**:101-106
- Zwick, M.g., Shen, J., Scott, J.K. 1998. Phage-displayed peptide libraries. *Curr. Opin. Biotech.* **9**:427-436