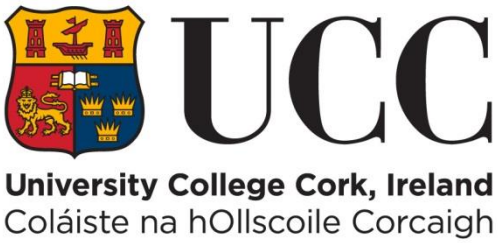
CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Mutational and molecular analyses of lactococcal bacteriophages TP901-1 and Tuc2009
Authors	Stockdale, Stephen R.
Publication date	2013
Original Citation	Stockdale, S. R. 2013. Mutational and molecular analyses of lactococcal bacteriophages TP901-1 and Tuc2009. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2013, Stephen R. Stockdale - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-11-02 04:14:06
Item downloaded from	https://hdl.handle.net/10468/1500

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh



Mutational and Molecular Analyses of Lactococcal Bacteriophages TP901-1 and Tuc2009

By

Stephen R. Stockdale, BSc

A thesis presented in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy.

School of Microbiology,

National University of Ireland, Cork.

December 2013

Research Supervisor: Prof. Douwe van Sinderen

Head of School: Prof. Gerald Fitzgerald

DECLARATION

I hereby declare that the content of this thesis is the result of my own work and has not been submitted for another degree, either at University College Cork or elsewhere.

_____

Stephen R. Stockdale

Date: _____

Dedicated to my parents and to Amy,

you made this possible

There's a divinity that shapes or ends,

Rough-hew them how we will

-Shakespeare, *Hamlet*

CONTENTS

ABSTRACT		1
CHAPTER I	Literature Review	3
CHAPTER II	The Lactococcal Phages Tuc2009 and TP901-1 Incorporate Two Alternate Forms of their Tail Fibre into their Virions for Infection Specialization	101
CHAPTER III	Mutational and Molecular Analysis of Lactococcal Phage TP901-1	140
CHAPTER IV	Mutational Analysis of a Lactococcal Tape Measure Protein Reveals its Multi-Functionality in Tail Morphogenesis and DNA Injection	215
CHAPTER V	General Discussion	258
ACKNOWLEDGEMENTS		271

ABSTRACT

Bacteriophages belonging to the *Siphoviridae* family represent viruses with a non-contractile tail that function as extremely efficient bacterium-infecting nanomachines. The *Siphoviridae* phages TP901-1 and Tuc2009 infect *Lactococcus lactis*, and both belong to the so-called P335 species. As P335 phages are typically capable of a lytic and lysogenic life cycle, a number of molecular tools are available to analyse their virions. This doctoral thesis describes mutational and molecular analyses of TP901-1 and Tuc2009, with emphasis on the role of their tail-associated structural proteins. Several novel and intriguing findings discovered during the course of this study on the nature of *Siphoviridae* phages furthers a basic molecular understanding of their virions, and the role of their virion proteins, during the initial stages of infection.

While *Siphoviridae* virions represent complex quaternary structures of multiple proteins and subunits thereof, mutagenic analysis represents an efficient mechanism to discretely characterize the function of individual proteins, and constituent amino acids, in the assembly of the phage structure and their biological function. However, as always, more research is required to delve deeper into the mechanisms by which phages commence infection. This is important to advance our understanding of this intricate process and to facilitate application of such findings to manipulate phage infections. On the one hand, we may want to prevent phages from infecting starter cultures used in the dairy industry, while on the other hand it may be desirable to optimize viral infection for the application of phages as bacterial parasites and therapeutic agents.

CHAPTER I

Literature Review

Part of this chapter has been peer reviewed and accepted for publication:

Mahony, J., Ainsworth, S., Stockdale, S., & van Sinderen, D. (2012). Phages of lactic acid bacteria: The role of genetics in understanding phage-host interactions and their co-evolutionary processes. *Virology*, **434** (2), 143-150.

CONTENTS

SECTION I	A History of Bacteriophages		5
SECTION II	Lactic Acid Bacteria and Bacteriophages		14
	II.I	LAB Cell Envelope	14
	II.II	Disruption of LAB Fermentations	20
	II.III	Phage Diversity and Classification	21
	II.IV	Phage Replication	28
	II.V	LAB Phage-Resistance Mechanisms	39
SECTION III	*Siphoviridae* Tail Structure and Function		50
	III.I	Tail Length Determination	50
	III.II	Initiating Tail Assembly	52
	III.III	Tail Tube Formation	54
	III.IV	Phage Host-Recognition	58
	III.V	Host Penetration and DNA Injection	68
SECTION IV	Concluding Remarks		71
SECTION V	Bibliography		73

SECTION I A History of Bacteriophages

In 1896, the British bacteriologist Ernest H. Hankin noted an unidentified substance in the Ganges and Jumna Indian rivers with antibacterial activity against the pathogenic bacterium *Vibrio cholera*. Unbeknown at the time, Hankin's description of a small, heat-labile, antibacterial agent was possibly the first scientific observation of a bacterial virus (1, 2). However, the discovery of viruses infecting bacteria is accredited to Frederick W. Twort (1915) and Felix d'Herelle (1917) (3, 4). Twort described an uncultivable 'glassy' bacterial appearance that when filtered could induce a glassy phenotype in other bacterial cells. Although uncertain in his conclusions of the filterable infectious agent, Twort deliberated over the possibility of viruses, and this work was enough to credit him with the co-discovery of bacteriophages. The other co-discoverer was Felix d'Herelle, who, independent of Twort, unravelled the mystery behind bacterial plaques. Perceiving these filterable infectious agents to be bacterial viruses, d'Herelle coined the term bacteriophage from a combination of the words 'bacteria' and the ancient Greek 'phagein' (meaning 'to eat') (5).

Since the discovery of phages, they have been a part of many important milestones in molecular biology. This is attributed to their small genome size, ability to work with their bacterial hosts, ease of experimental manipulation and relatively minimal replication complexity. Outlined below are several examples of how phages have contributed to our understanding of molecular biology.

Evolution, a concept now taken for granted (by many), previously required scientific debunking. Salvador Luria and Max Delbruck (1943) demonstrated evolution by challenging a population of bacteria, all originating from a common

ancestor, with phages. Within a few hours, the media, cloudy from bacterial growth, returned clear. However, if Luria and Delbruck incubated the now-cleared media further, the culture sometimes regrew. This was due to the growth of a sub-population of phage-resistant bacteria that arose within the population of bacteria. Through mathematical modelling, Luria and Delbruck calculated these heritable changes, or mutations, in the bacterial phage-sensitivity phenotype over time. These mutations occurred independent of the phage, during bacterial growth, as the number of resistant bacteria did not increase in stationary phase cells (6). Natural selection, at a microscopic level, was therefore first demonstrated by Luria and Delbruck using phages.

In 1952, Alfred Hershey and Martha Chase used phages to demonstrate, beyond any reasonable doubt, that nucleic acid encoded proteins. By analysing radioactively labelled phage DNA and phage proteins, Hershey and Chase demonstrated that phage proteins directed DNA injection, while only radio-labelled DNA was passed on to progeny phage (7).

Norton Zinder and Joshua Lederberg (1952) demonstrated that a *Salmonella*-infecting phage could carry DNA, encoding a selectable marker, between bacteria. This process, different from sexual recombination in bacteria, was termed transduction (8). Many years following on from this discovery, transduction has remained an important tool in molecular biology and represents one of the main routes by which bacteria achieve horizontal gene transfer.

Three years after James D. Watson and Francis Crick deciphered the structure of DNA, two scientists, Elliot Volkin and Lazarus Astrachan (1956) discovered the missing link between DNA and protein. At this time, it was known

that there had to be some correlation between hereditary genetic material and proteins, however, conclusive experimental evidence was lacking. Volkin and Astrachan noted that soon after bacterial infection by a phage, the infected bacterium started to produce phage proteins. While the majority of the cellular RNA remained unchanged, a small portion of the cellular RNA was observed that had a similar base composition to that of the infecting phage DNA. They called their discovery 'DNA-like-RNA' (9). Like most scientists who accidentally discover something monumental, they did not realize the significance of their find. It was only later that François Jacob and Jacques Monod realized this 'DNA-like-RNA' was in fact a cytoplasmic message for protein synthesis; hence its name as messenger RNA (mRNA). Jacob and Monod later received a Nobel Prize (1965) for their contributions to the understanding of certain regulatory processes that occur in cells, while Volkin and Astrachan were never acknowledged for their contribution to molecular biology (10).

Bacteriophage T4 has served as an important molecular tool for understanding the triplet nature of the genetic code. Francis Crick, Leslie Barnett, Sydney Brenner and R.J. Watts-Tobin (1961) used the acridine dye proflavin to introduce or remove bases into T4's gene *rII*. They demonstrated that the addition or deletion of nucleotide bases, and complementary additions or deletions of bases, were deleterious, but not if they were in multiples of three. Overall Crick and colleagues used phage T4 to demonstrate that three nucleotide bases, a codon, encode one amino acid (Fig. 1). Crick *et al.* further speculated at the degeneracy of the genetic code, as there were 64 possible codons and only 20 amino acids (11). However, it was Marshall Warren Nirenberg (1961) who deciphered the first codon,

UUU, which encodes phenylalanine, marking a major milestone in understanding molecular biology (12).

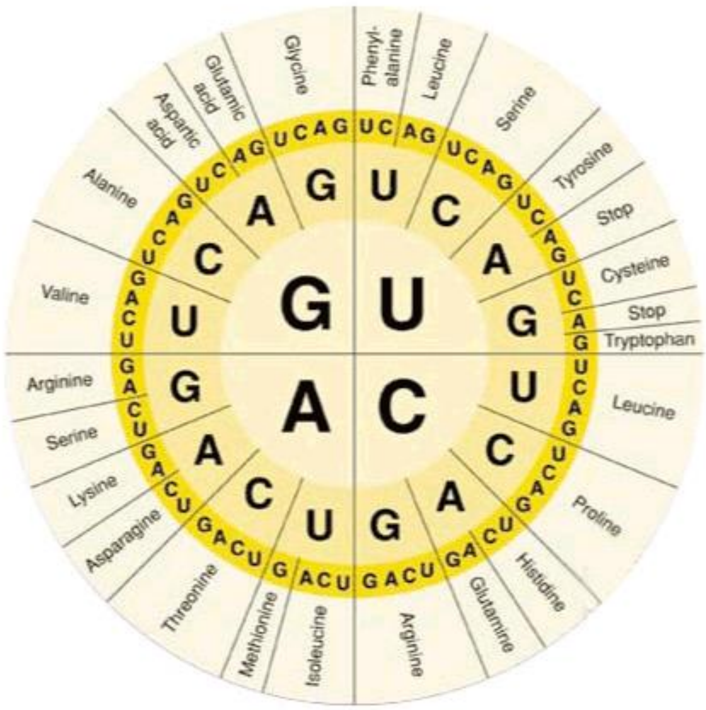


Figure 1. Circular genetic code table. The four nucleotide bases of mRNA are: A, Adenine; U, Uracil; G, Guanine; C, Cytosine. An individual codon, as demonstrated by Crick *et al.* (1961), is composed of 3 nucleotides. The circular codon table above shows the first codon position in the centre, with the names of the 20 encoded amino acids radially located around their respective third codon positions. As predicted by Crick and colleagues, the 20 amino acids are encoded by 61 codons, in addition to three codons that represent translational termination signals (11). This can be seen with the amino acid phenylalanine, deciphered by Nirenberg as UUU (12), and which is also encoded by UUC (figure sourced from the following web link: http://www.millerandlevine.com/).

During the 1950's and 60's, it became apparent that the transcription of genetic material was regulated. Jacques Monod, who was studying the *lac* operon, showed that the lactose-degrading enzyme β-galactosidase was produced following the addition of lactose to the media. When the scientists Andre Lwoff, François Jacob and Jacques Monod (1961) teamed up, they also began to look at the genetic regulation of phage λ lysogeny (λ was discovered by Joshua Lederberg's wife, Esther). In genetic maps of *E. coli*, they showed lysogenized λ always occurred at a defined place. Also, when the DNA encoding a prophage λ entered a female *E. coli* through conjugation, the λ lysogen was induced. Lwoff *et al.* speculated that it was a cytoplasmic factor of the male *E. coli* maintaining the dormant phage (13). The scientific discoveries by Lwoff, Jacob and Monod, ultimately leading to the operon model of genetic regulation (Fig. 2), whereby regulatory genes encode proteins that interact with operator sequences to control sets of co-located structural genes, earned them a Nobel Prize (1965).

A few years later, Mark Ptashne (1967) demonstrated that phage λ repressor protein maintained the lysogenized phage within the chromosome (Fig. 2). The λ repressor also fulfilled Jacob and Monod's operon model, as it was shown to bind tightly to the operator sequences, blocking transcription and thus maintaining the lysogenic phage (14).

Allan Campbell (1962) demonstrated DNA integration into genomes using phage λ. While phages such as Mu are capable of transposition (random genome insertion), λ integrates by site-specific recombination. Campbell focused on the chromosomal sequences specific for recombination, identifying the need for specific sequences and enzymes to catalyse the reaction (15). Campbell's model went a long

way towards understanding the insertion of foreign DNA into a genome and also the ability of viruses to remain dormant until activated.

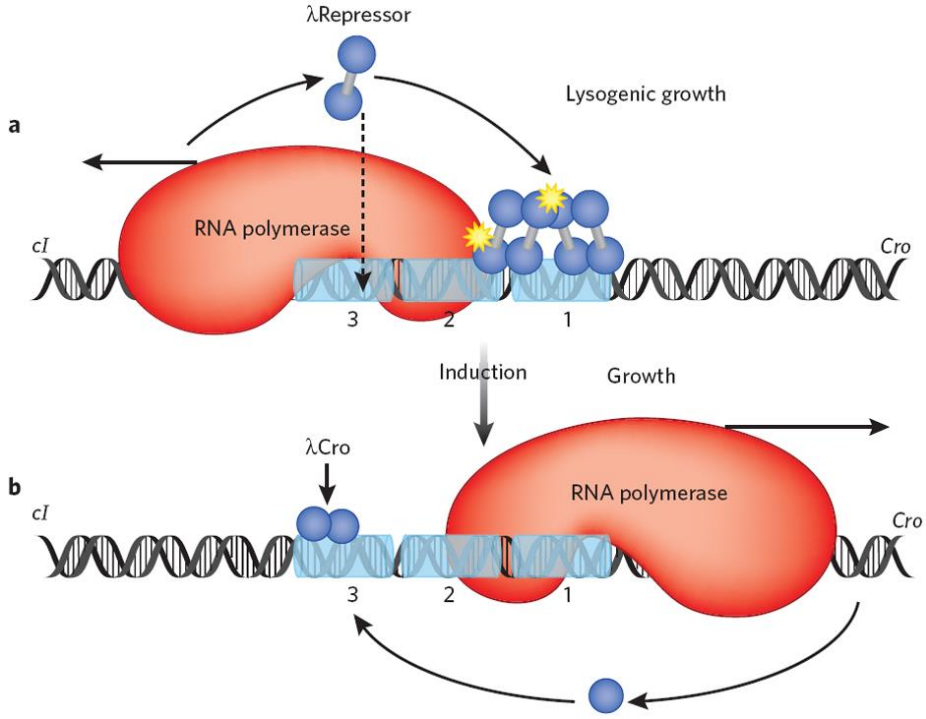


Figure 2. Phage λ genetic switch obeys the operon model of Lwoff, Jacob and Monod. (A) Leftward transcription from the divergent genetic switch of phage λ produces the repressor protein, cI. Protein cI forms a homodimer and binds to operator sequence 1. Repressor binding to operator sequence 1 recruits a second cI dimer to bind operator sequence 2, which in turn enhances production of additional repressor proteins. (B) Induction conditions result in the degradation of cI protein and rightward transcription from the genetic switch producing the anti-repressor protein, Cro. The anti-repressor protein stimulates the transcription of factors required for λ lytic growth (figure sourced from (16)).

Those working with phages observed a phenomenon whereby phages infecting a particular strain did so more effectively in the second generation of propagation. It was Daisy Dussoix and Werner Arber (1962) who proposed a model by which a host endonuclease could degrade susceptible foreign phage λ DNA, restricting infection (hence the name restriction enzyme). The host strain, by modifying its own DNA, was protected from this endonuclease. Dussoix and Arber postulated that, at low frequency, susceptible phage λ DNA during infection also became modified, which subsequently increased their infection efficiency in the second generation (17). Several years later, Martin Gellert (1967) demonstrated that *E. coli* extracts could close the cohesive ends of λ phage DNA into a closed circle (18). When taken together, the ability to digest DNA at specific sequences, by restriction enzymes, and repair these cuts, via a ligase, are amongst the most important tools in DNA manipulation and molecular engineering.

Costa Georgeopoulos, Roger W. Hendrix, Sherwood R. Casjens and A. D. Kaiser (1973) isolated host mutants incapable of supporting bacterial colonies at 43 ^{o}C and assembling the phage λ capsid. These mutants were mapped to *groE* and could be complemented (19). However, it was years after the description of *groE* that its function as a chaperone molecule became apparent. Reginald John Ellis (1987), who was working on Rubisco in chloroplasts, demonstrated a chaperone function of a Rubisco-binding protein. Ellis argued molecular chaperones were common in nature, assisting in folding and unfolding of macromolecular structures, a notion that was strongly supported by the Rubisco-binding protein's homology to GroEL, known to be important for the production of various different bacteriophages (20).

Recent advances in DNA sequencing have resulted in biologists requiring massive computational power to interpret the generated data. However, the first attempts to sequence genetic material (such as Maxam and Gilbert's chemical degradation) were laborious, required numerous dangerous chemicals, and generated relatively little data (21). The major breakthrough in sequencing was the method of chain-termination developed by Frederick Sanger (1977). The chain-termination method uses dideoxynucleoside triphosphates (ddNTPs) capable of being incorporated into an extending DNA sequence, yet upon incorporation they themselves cannot facilitate further DNA extension. By analyzing four nucleotide reactions in parallel, each containing a small proportion of a particular ddNTP in with the corresponding dNTPs, a number of different molecular weight products are produced. By comparing the sizes of terminated DNA sequences and correlating them to the incorporation of the dideoxy -adenines, -guanines, -thymines, and -cytosines, the genetic sequence could be determined (22). The single-stranded DNA genome of phage φX174 (5,386 nt), the first genome to be sequenced by chain-termination, represented a major milestone in molecular biology. The ability to sequence DNA is an important analytic tool and has been pivotal in advances in multiple life-science fields.

Phages continue to contribute to discoveries in molecular biology and the fundamental nature of DNA, not all of which have been discovered despite modern advances. Mark Ptashne, who demonstrated the repressor protein and lysogenic switch of phage λ, has shown that genetic regulation is a much more complex system than initially perceived. Ptashne's book 'A Genetic Switch' (2004) demonstrated epigenetic control of phage λ DNA by the persistent expression of λ lysogeny genes in the absence of the original signal, without a change in DNA sequence (23). The

scientific advances that have been made through phage research, such as epigenetics, have had major consequences in molecular biology and other branches of science, and phage research will hopefully continue to significantly contribute to our understanding of molecular biology.

SECTION II Lactic Acid Bacteria and Bacteriophages

Lactic acid bacteria (LAB) represent a diverse group of Gram-positive, low-GC genomic content, non-spore forming, coccoid or rod-shaped bacteria that produce lactic acid as the major end-product of fermentation. Species of *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Pediococcus* and *Streptococcus thermophilus* form the core LAB representatives and are economically important in the dairy industry for their fermentation of milk to produce dairy products, such as yogurts and cheeses, with desirable tastes and textures (24, 25). For example, in 2012 the Irish dairy industry exported fermented dairy products and ingredients with an estimated value of €2.66 billion (26), making it one of Ireland's most important agri-food sectors. The metabolic activities of LAB have been exploited for thousands of years to ferment carbohydrates in foods for human consumption. With few exceptions, such as *Enterococcus faecalis*, LAB are deemed generally regarded as safe (GRAS) organisms and their presence in food is considered harmless or even adventitious as many are claimed to possess health-promoting probiotic properties (27).

SECTION II.I LAB Cell Envelope

The exposed surface of a microorganism represents a rather unique molecular signature that allows interactions with its environment. Particular surface components of certain probiotic LAB are known to interact with the human gastrointestinal tract and have been demonstrated to exclude or inhibit pathogens through enhancing the intestinal epithelial barrier and modulating the host immune

response (28). The cell envelope of LAB is composed of an inner phospholipid cell membrane surrounded by a thick cell wall of peptidoglycan (PG) decorated with various proteins, teichoic acids and carbohydrates.

The majority of cellular life is covered in a cell wall to resist turgor; animal and protozoal cells being notable exceptions (29). PG (peptidoglycan or murein) of bacterial cells forms a macromolecular sacculus, composed of long glycan strands of repeating *N*-acetylglucosamine (GluNAc) and *N*-acetylmuramic acid (MurNAc) sugars cross-linked to neighbouring glycan strands through short peptide stems protruding from the D-lactyl moiety of MurNAc. The consensus peptide stem of most Gram-positive bacteria, including LAB, is a pentapeptide of $_{L}$-Ala-γ-$_{D}$-Glu-$_{L}$-Lys-$_{D}$-Ala-$_{D}$-Ala (Fig. 3). Interspecies variations in PG are observed in the glycan strands, peptide stems, and interpeptide bridges which can be present to link peptide stems. Variations in murein also occur in a given bacterium depending on growth conditions and phase of growth, thus highlighting its dynamic nature (30, 31).

The glycan strands of most Gram-positive bacteria are cross-linked indirectly via an interpeptide bridge in their peptide stems. The size of the interpeptide bridge can range from one to seven amino acids, and its composition can be quite diverse. The interpeptide bridge of lactococci and several lactobacilli are composed of a single dicarboxylic amino-amino acid, typically $_{D}$-Asp (31). As peptide stems are theoretically distributed regularly along the glycan strand, the number of interpeptide bridges directly correlates with glycan strand length. However, the thickness of PG in the cell envelope and glycan strand length do not necessarily correlate, as thick cell-walled *Staphylococcus aureus* and *Bacillus subtilis* have short and long glycan strands, respectively (32).

A PG hydrolase (PGH) enzyme exists for every chemical bond that is present in the PG muropeptides (Fig. 3). Muramidase and glucosaminidase enzymes cleave glycosidic bonds between MurNAc-GluNAc and GluNAc-MurNAc, respectively. Amidase enzymes (more specifically *N*-Acetylmuramyl-$_L$-alanine amidases) hydrolyze the peptide stem from the MurNAc of the glycan strand, while certain cell wall-active peptidases cut within the peptide-stem and peptide-bridge of PG (33). Bacterial PGHs are necessary for normal bacterial growth as the cell wall is constantly recycled and expanded to accommodate cell growth and division. Lysis of bacteria by their own PGHs, termed autolysis, is particularly important with regards LAB, as it facilitates the production and ripening of cheeses (34).

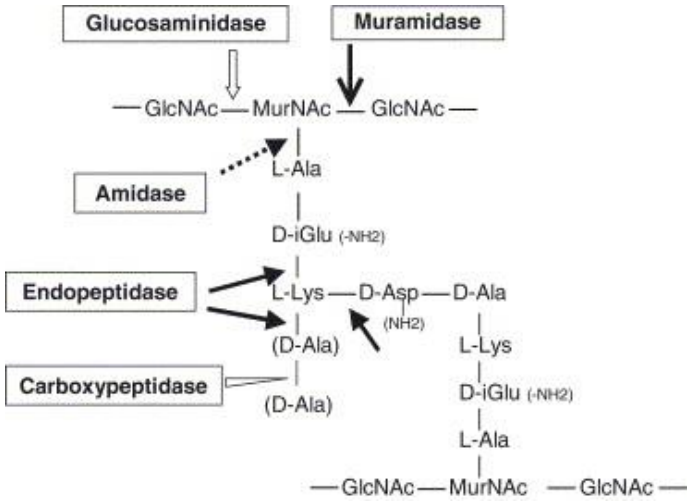


Figure 3. Peptidoglycan structure of lactic acid bacteria with the chemical bonds targeted by various hydrolase enzymes indicated (figure sourced from (35)).

Three architectural models of peptidoglycan are debated amongst biologists (Fig. 4). Either the PG glycan strands run parallel to the bacterial membrane, in the

classical layered model or the disorganised layered model, or the PG is perpendicular to the bacterial surface and the cell wall cross-linkage is parallel to membrane, as in the 'scaffolding' model (36). Gan *et al.*'s (2008) study of frozen-hydrated *Escherichia coli* cell walls using cryo-tomographic electron microscopy showed a single 2.5 nm layer of PG comprising the sacculus, supporting the disorganized layered model (37).

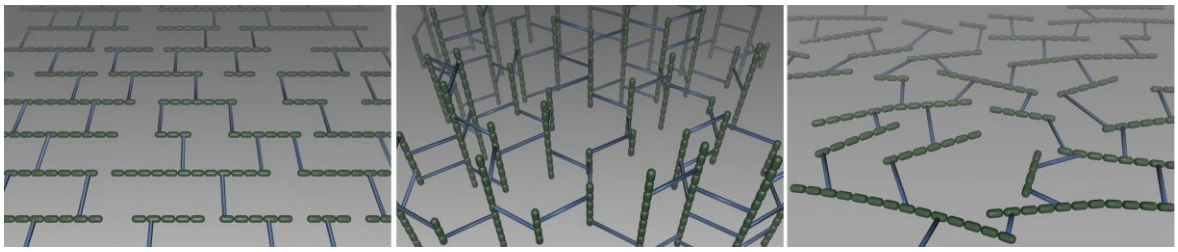


Figure 4. Models for PG structure. (A) The layered model, (B) the scaffold model, and (C) the disorganized layered model. MurNAc and GluNAc disaccharides are in green while PG proteinaceous cross-linkages are in blue (figure sourced from (37)).

The PG architecture of *B. subtilis* is possibly the best studied Gram-positive bacterium and (predominantly) supports the layered model. The following paragraph is a summary of a detailed description by Hayhurst *et al.* (2008) regarding the current understanding of the Gram-positive cell wall (38). Muropeptide glycan strands have a right-hand helical twist, and under certain physiological conditions *B. subtilis* grows helically. A number of glycan strands are cross-linked to form a helical PG 'rope', and PG ropes are then coiled into a helix (of 50 nm width) to form 'cable' structures. Turgor pressure on the cell causes PG cables to flatten, resulting

in characteristic 25 nm cross striations observed in the PG cables when observed by atomic force microscopy (AFM). The PG of *B. subtilis* runs circumferential to the rod shaped bacterium, and PG cables are stabilized by inter/intra glycan strand linkages. New peptidoglycan cables are inserted into the cell wall by cross-linkage with existing cables by enzymes protruding from the cell membrane into the bacterial periplasmic space.

AFM images of live lactococcal cells show a smooth surface, while mutants defective in a surface-coating cell wall polysaccharide (CWP) show 25 nm periodic bands running parallel across the bacterium (Fig. 5) (39). Whether these bands are analogous to the coiled cables of the glycan strands observed in *B. subtilis* is yet to be determined, but it strongly suggests a common structural arrangement of PG amongst these Gram-positive bacteria.

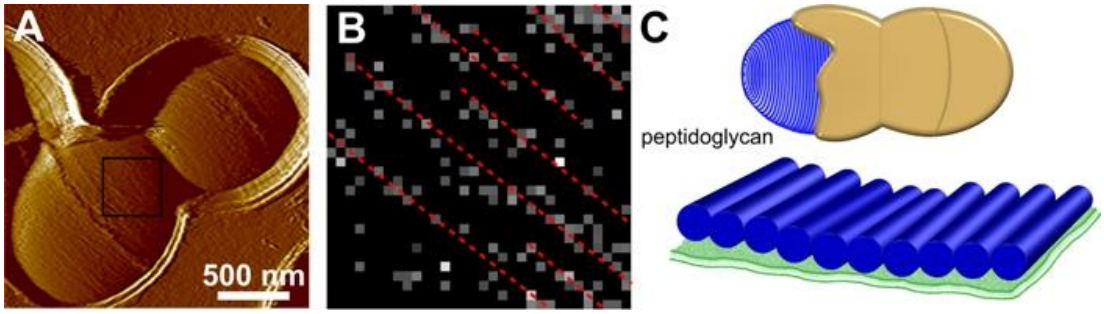


Figure 5. Atomic force microscopy showing the nanoscale organization of lactococcal PG. (A) Division of an *L. lactis* mutant lacking its CWP. (B) Detection of peptidoglycan molecules found they were arranged parallel around the short cell axis (red lines). (C) Cartoon representation of PG periodic cables (blue) lying on top of the cell membrane (green) and underneath the natural CWP coating (brown) (figure sourced from (39)).

Proteins, teichoic acids and carbohydrates present on the surface of LAB fulfil multiple functions during the complex life cycle of these microorganisms; therefore, their discussion in this review is limited to their pertinence with regard phage infection (for more information, see reviews (40-43)). Proteins on the surface of bacteria represent specific targets for phage adsorption. Surface-layer (S-layer) proteins found on the surface of some lactobacilli have been reported to act as targets for phage to facilitate adsorption and subsequent infection (44) and the phage infection protein (Pip) on the surface of lactococcal cells serves as a receptor for c2 species phages (45).

Teichoic acids are long anionic polymers composed largely of glycerol phosphate, glucosyl phosphate, or ribitol phosphate repeats. Teichoic acids can be associated with the bacterial cell wall (wall teichoic acids) or anchored to the bacterial phospholipid membrane (lipoteichoic acids) (46). Teichoic acids were demonstrated as the surface-recognised receptor of *Lb. delbrueckii* and *Lb. plantarum* phage LL-H and Phi2, respectively (47, 48). Teichoic acid is also a candidate receptor for *S. thermophilus* phage 0BJ, as adsorption to its host is strongly inhibited by the carbohydrate ribose (49).

A surface-coating CWP of *L. lactis* MG1363 was recently described (50). This CWP is composed of hexasaccharide repeating units, containing rhamnose, glucose, galactose and GluNAc, linked by phosphodiester bonds. The saccharidic composition of lactococcal CWPs differs amongst strains, and appears to correlate with the observed genetic diversity of their encoding regions (51) (Ainsworth *et al.*, unpublished results). CWPs on the surface of *S. thermophilus* have also been reported, although they have not been studied in as much detail as their lactococcal

counterparts. Similar to the CWPs of *Lactococcus*, *S. thermophilus* CWPs were demonstrated to contain carbohydrates rhamnose, glucose and GluNAc, and affect the adsorption of streptococcal phages (49, 52). LAB CWPs are believed to be covalently connected to PG and contribute to the ecological properties of the strain, as they were demonstrated to affect the recognition of lactococci by phagocytes and phages (50) (Ainsworth *et al.*, unpublished results).

SECTION II.II Disruption of LAB Fermentations

Prokaryotic viruses are ubiquitous and abundant in seemingly all ecological niches (53), industry notwithstanding. The first report of phages affecting an LAB-mediated fermentation was by Whitehead and Cox in 1935 (54). Since then, numerous procedures have been implemented to prevent, control or reduce the incidence of failed fermentations as a result of phage infection, as slow or dead fermentations result in disruption of the production schedule and low quality produce (for reviews, see (55-57)). However, infection of starter cultures remains the single greatest cause of fermentation failure and a significant challenge to the fermentation industry.

Strategies used by the dairy industry to limit phages within their production facilities include: thermal and biocidal treatments, strain selection and rotation, and the application of bacteriophage-insensitive mutant strains (BIMs) (58, 59). However, phages appear to have become more resistant to pasteurisation temperatures and sanitation procedures (60). Therefore, phages enter, and contaminate, dairy plants through incoming milk. To further complicate matters,

robust phages which manage to enter dairy factories were demonstrated to persist on various work surfaces and spread through the production facility as airborne particles (61).

Strain rotation is a simple, yet efficient, phage control procedure. Phages of LAB typically have narrow host ranges (51), and the rotation of starter cultures prevents the amplification of a specific infecting phage (56). In addition to strain rotation, the selection of starter cultures with phage-resistant properties is important to avoid fermentation problems. For instance, *L. lactis*-infecting phages of the c2 species require a proteinaceous receptor, Pip, to successfully recognise and initiate infection of their hosts. The generation of lactococcal BIMs lacking the Pip-encoding gene, through evolutionary pressure and natural selection, results in strains resistant to phages belonging to the virulent c2 species (62-64).

SECTION II.III Phage Diversity and Classification

The high cost associated with phage-mediated problems in the dairy industry has in turn resulted in significant scientific research efforts, which is crucial to the understanding of the phage problem and to finding solutions for the affected industrial processes. In addition, with cheaper and more advanced DNA sequencing technologies available, the number of fully sequenced genomes of LAB and their phages has dramatically increased (Table 1). Taken together, a considerable wealth of knowledge has been generated regarding the diversity and genomic content of dairy phages, their ecology, their replication cycle, the responses of bacteria to phages, and host-encoded phage-defence mechanisms (for reviews, see (65-67)).

Taxonomic classification is an important procedure to create a compilation of members with shared characteristics. Categories can be scrutinised for many purposes, including: (i) the determination of inter-/intra-relatedness amongst collections, (ii) identification of similarities and unique aspects, (iii) and predicting the ecological and evolutionary significances. The International Committee on the Taxonomy of Viruses (ICTV) is primarily charged with the classification of viral orders, families, and genera, and then designating the particular level of nomenclature in the form of the suffixes -virales, -viridae, and -virus, respectively (68).

Table 1. Overview of the number of fully sequenced and published LAB and LAB-infecting phage genomes. Best approximate of bacterial and phage genomes were sourced from (www.genomesonline.org) and (www.ebi.ac.uk/genomes/phage.html), respectively. Table updated from Mahony *et al.* (2012) and accurate as of October 2013 (69).

	No. of sequenced genomes
Bacteria	
Lactobacillus spp.	48
Lactococcus spp.	10
Leuconostoc spp.	8
Pediococcus spp.	2
S. thermophilus	6
Phages	
Lactobacillus spp.	21
Lactococcus spp.	77
Leuconostoc spp.	5
Pediococcus spp.	1
S. thermophilus	12

In contrast to the success of bacterial classification by a universally present taxonomic unit, 16S rRNA, determining the phylogenetic relationship of phages is not as straightforward. Comparative genomic analysis on the ever-increasing number of sequenced phages is facilitating the determination of evolutionary relationships to a greater resolution. However, one of the major requirements for classifying viruses is the application of the 'polythetic species concept'. This means a species may be defined by a set of properties some of which may be absent in a given member (70). As a consequence, the primary feature for classification of phages remains the morphology of phages as defined by electron microscopy.

Bradley (1967) referred to several types of phages as resembling a 'tadpole', possessing a head-structure, called a capsid, attached to a tail that is capable of changing shape (71). The phylogenetic order *Caudovirales* was consequently established to represent the ca. 96 % of phage virions that possess a tail structure (72). The capsids of these phages can be icosahedral (round), prolate (slightly extended), or elongated in shape. Based on tail morphology, replication and assembly, it is recognised that the order *Caudovirales* can be subdivided in three familial groups of tailed, double-stranded DNA phages: (i) *Myoviridae*, for phages with long contractile tails, (ii) *Podoviridae*, phages with short non-contractile tails, and (iii) *Siphoviridae*, phages with long non-contractile tails. The remaining 4 % of non-tailed phages possess polyhedral, filamentous, and pleomorphic morphologies (Fig. 6). These phages have no taxonomic rank at the level of order assigned, but 10 families are recognised (73).

Phages infecting LAB typically possess a double-stranded DNA genome encased in a virion with a long non-contractile tail, characteristic of the phage family

Siphoviridae from the order *Caudovirales* (72, 74). However, a small number of tailed *Podoviridae* and *Myoviridae* phages, also of the order *Caudovirales*, have been found to infect *L. lactis* species and the genus *Lactobacillus*, respectively. As lactococcal phages represent one of the best studied groups of bacterial viruses infecting a single bacterial species, 10 distinct phage species have been characterized (75). The most frequently isolated *L. lactis* phages that disrupt milk fermentations are *Siphoviridae* belonging to the 936, c2 and P335 species (75). While 936 and c2 species phages are strictly virulent, P335 species phages can be virulent or temperate. Two *Podoviridae* and five additional *Siphoviridae* species are known to infect *L. lactis*; however, these species are considered rare (Table 2).

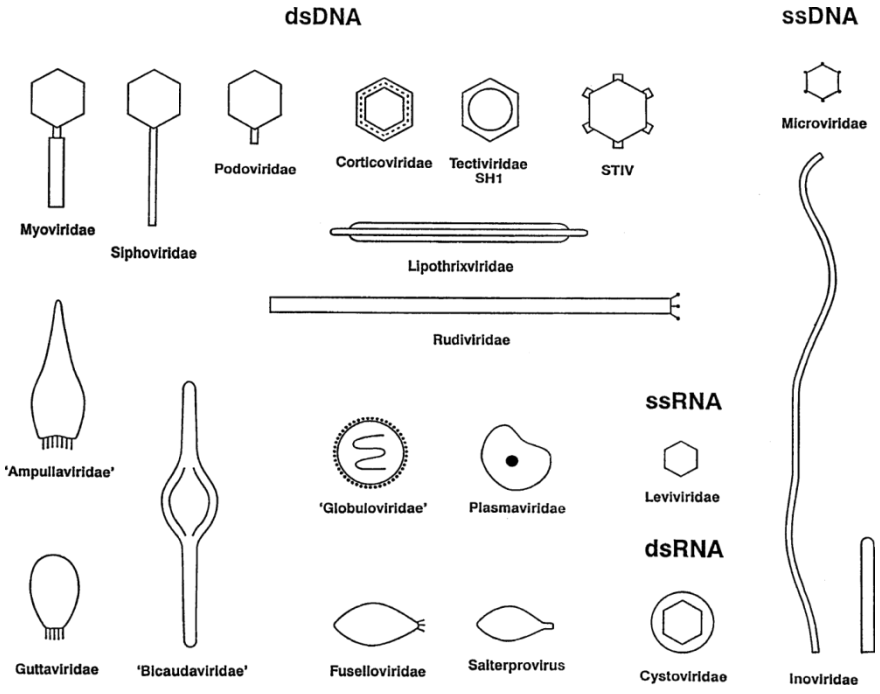
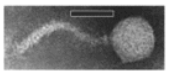
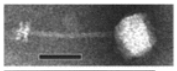
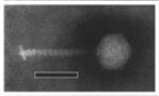
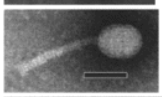
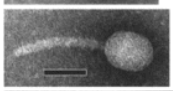
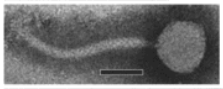
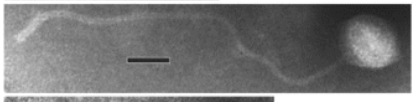
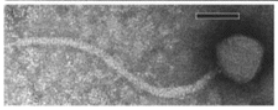
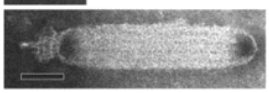


Figure 6. Phage classification as based (primarily) on the morphological characteristics of phages as observed under an electron microscope. While 96 % of all phages to date display only three morphologies (*Myoviridae*, *Podoviridae* and

Siphoviridae), many other phage morphologies have been described (figure sourced from (73)).

Investigators Rohwer and Edwards (2002) proposed an alternative system of classifying phages compared to that currently employed by the ICTV. This was termed the phage 'Proteomic Tree', based on DNA sequencing and comparative analyses of the predicted proteome. The advantages over the ICTV mechanisms are as follows: firstly, ICTV requires visualization of free phage particles under an EM, limiting classification only to those that can be propagated; secondly, prophage sequences are omitted; thirdly, metagenomic studies that generate large amounts of sequence data, by their very nature, do not attempt to work with the host bacterium or phage in question. Due to the limitations imposed by the ICTV, the number of unclassified phages will increase as more (complete) phage genome sequences become available (76). While no single conserved sequence is present in the genomes of all phages, the phage proteomic tree created by Rohwer and Edwards produced a phage phylogenetic tree similar to that produced by the ICTV-supported classification scheme (Fig. 7).

Table 2. Diversity and classification of lactococcal phages. [a] Scale bar represents 50 nm (table sourced from (75)).

Family Species	Phage	Capsid diameter (nm)	Tail width (nm)	Tail length (nm)	Electron micrograph[a]
Siphoviridae					
936	bIL170	50	11	126	
P335	ul36	49	7	104	
1358	1358	45	10	93	
c2	c2	54 X 41	10	95	
Q54	Q54	56 X 43	11	109	
P087	P087	59	14	163	
949	949	70	12	490	
1706	1706	58	11	276	
Podoviridae					
P034	P369	57 x 40	5	19	
KSY1	KSY1	223 X 45	6	32	

Rohwer and Edwards (2002) also propose a new nomenclature system, where the suffix '-phage' replaces '-viridae' when referring to computational classification and not morphological. Other suggestions to phage nomenclature provided by Rowher and Edwards include that the first mention of the phage should reference the host organism (or else left as 'unknown') and using the symbol φ (phi for phage) followed by the first designation of the phage from its original source (the example used was *E. coli* φX174) (76).

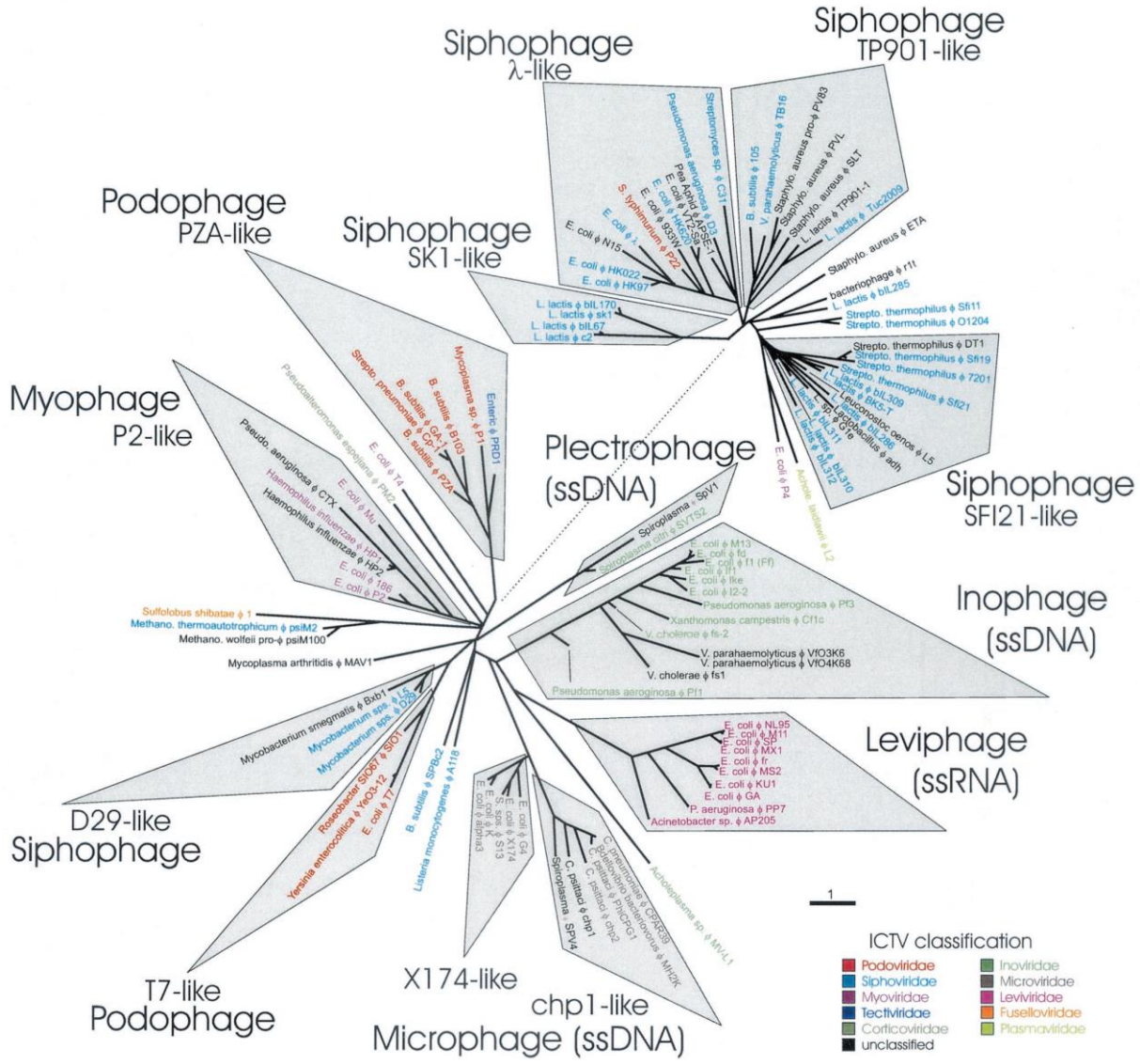


Figure 7. Phage phylogenetic tree constructed using the predicted proteome of phage sequences. The tree consists of 105 completed genome sequences, generated by length-corrected protein distances, with a penalty score for missing proteins. Each genome is coloured according to their ICTV classification as shown in the key. For ease of reading the large Siphophage group has been moved away from other phage groups (as indicated by the dotted line) (figure sourced from (76)).

SECTION II.IV Phage Replication

Incapable of self-proliferation, phages require the inadvertent contribution of a host's DNA replication and protein synthesizing machinery to complete their life cycle. The lytic cycle of phages is frequently divided into 5 main stages; adsorption, penetration, synthesis, assembly and release. In addition, temperate phages can integrate their genome into the genome of their host and replicate *in situ*, before excising and proceeding to replicate through the lytic cycle (Fig. 8).

Phage recognition of a cognate bacterial host has important consequences, as it will determine the fate of both the phage and the bacterium. Several phages seem to target two different receptors on the surface of the bacterium to adsorb to: a reversible adsorption to a bacterial generic moiety, allowing the phage to 'walk' along the bacterium surface in order to find a second, specific receptor that mediates irreversible adsorption (77). As discussed above, a variety of surface-exposed moieties are adsorption targets for LAB-infecting phages, such as proteins, teichoic acids and carbohydrates.

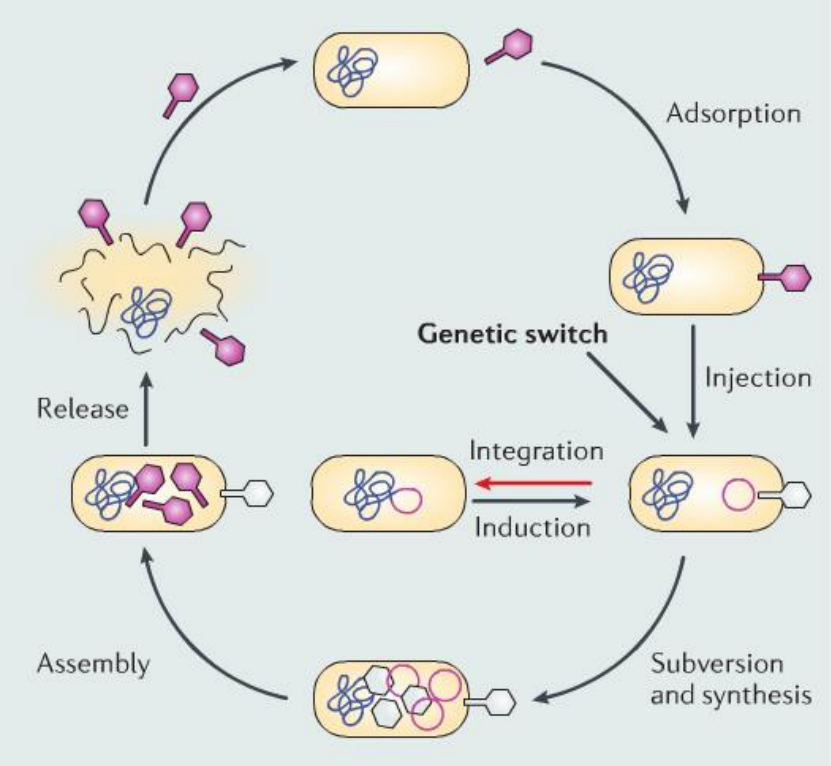


Figure 8. A simplified schematic representation of a phage replication cycle. The phage lytic cycle follows: adsorption, penetration (injection), synthesis, assembly and release. The lysogenic switch marks the 'decision' by temperate phages to integrate into, or induce from, their host (figure sourced from (78)).

Irreversible adsorption commits the phage to infection, whereby its genetic material must gain entry into the host's cytoplasm. However, the bacterial cell envelope consists of multiple barriers preventing easy access of the phage genetic material into the bacterium. Extracellular bacteriophage must penetrate the PG-containing cell envelope to infect a bacterium and produce progeny that is subsequently released into the environment. The parasitic replication cycle of many phages is facilitated by virion-associated PGH (VAPGH) enzymes to locally degrade PG at the initiation of infection (for morphologically distinct phage-type examples,

see (79-83)). The thicker PG sacculus of Gram-positive bacteria, relative to Gram-negative bacteria, is likely a more considerable physical barrier to phage infection. In accordance with this, Moak and Molineux (2004) found that five LAB-infecting *Siphoviridae*, representative of four different lactococcal-infecting phage species, possess a virion-associated PGH activity (84). VAPGH domains are often associated with large, essential phage structural proteins, often encumbering functional characterisation of such proteins. The covalent association of VAPGH domains with these large key proteins has been proposed to prevent diffusion of the potentially lethal, PG-degrading enzymes (84). Several phages require PGH domains only under certain physiological conditions (for examples, see (85-87)), and thus studying PGH domains and their localization often reflects on the infection mechanism of such particular phages.

The challenge of traversing the bacterial cell envelope, including lipid membranes and the peptidoglycan cell wall, may well have been a major catalyst in driving the observed diversity of phage morphologies. The virion structures of specific phage classes, such as *Myoviridae*, *Podoviridae* and *Siphoviridae*, represent extremely efficient DNA-delivering nano-machines, and members of each morphology appear to utilize a unique mechanism to penetrate the bacterial cell envelope and inject their genetic material into their particular host (88). Discussed below are prototypical examples of *Myo-*, *Podo-* and *Sipho-viridae* phages; LAB-infecting phages with these morphologies are expected to penetrate their hosts in a similar manner.

Myoviridae, with long contractile tails, are often described to employ an injection mechanism similar to a ‘syringe’ (89). After irreversible adsorption,

contraction of the *Myoviridae* tail sheath drives the tail tube through the bacterial envelope. The genetic material of the phage is then ejected from the phage capsid, directed through the phage tail tube, thereby facilitating its exit at the tail tube tip and directly in the bacterial cytoplasm. The best studied *Myoviridae* and accepted prototype of this phage family is T4 (Fig. 9). Comprehensive reviews of the structure and function of T4's virion are available (90-92). T4 adsorbs to the terminal glucose of lipopolysaccharide (LPS), in combination to the outer membrane protein OmpC, on the surface of its enterobacterial host through its long tail fibres (93, 94). T4 infection can occur, albeit at lower efficiency, in the absence of either the terminal glucose of the LPS or OmpC (95). However, binding by at least three long tail fibres is required for T4 to be infectious (96). The signal generated by T4's long tail fibres adsorbing to the host's receptors results in tail tube contraction, which in turn pushes the cell-puncturing device, composed of a trimer of gp5, through the bacterial cell envelope (97). The gp5 syringe-complex of phage T4 also possesses lysozyme domains facilitating the degradation of the bacterial cell wall during host penetration (98).

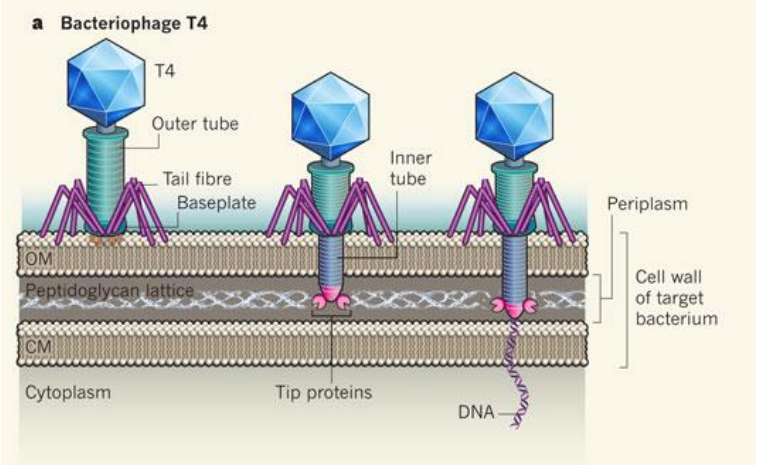


Figure 9. Host penetration by bacteriophage T4. Long tail fibres reversible adsorb to LPS and OmpC on the surface of *Enterobacteriacaea*. If three long tail fibres recognise the bacterial receptors, T4's tail sheath contracts, driving the cell puncturing syringe through the bacterial envelope. The phage DNA can then exit the phage tail tip directly into the host's cytoplasm. Penetration of the bacterial cell wall is facilitated by VAPGH domains associated with the cell-puncturing device's tip (figure sourced from (89)).

Podoviridae, with their short non-contractile tails, have been shown to eject proteins from their capsids that form a channel to bridge the cell envelope. The channel from the phage virion to the bacterial cytoplasm both directs the phage genetic material and protects it from nucleases present in the bacterial cell-envelope's periplasmic space (99). The penetration of the bacterial cell envelope by *Podoviridae* has been demonstrated for several phages, however, the best example is phage T7 (Fig. 10). Phage T7 ejects the capsid core proteins corresponding to gene products 14 (gp14), gp15 and gp16 from its capsid to span the bacterial cell envelope. Protein gp16 also possesses a lytic transglycosylase domain to degrade the peptidoglycan cell wall of its enterobacterial host, facilitating the penetration stage of phage T7's infection process (86, 100).

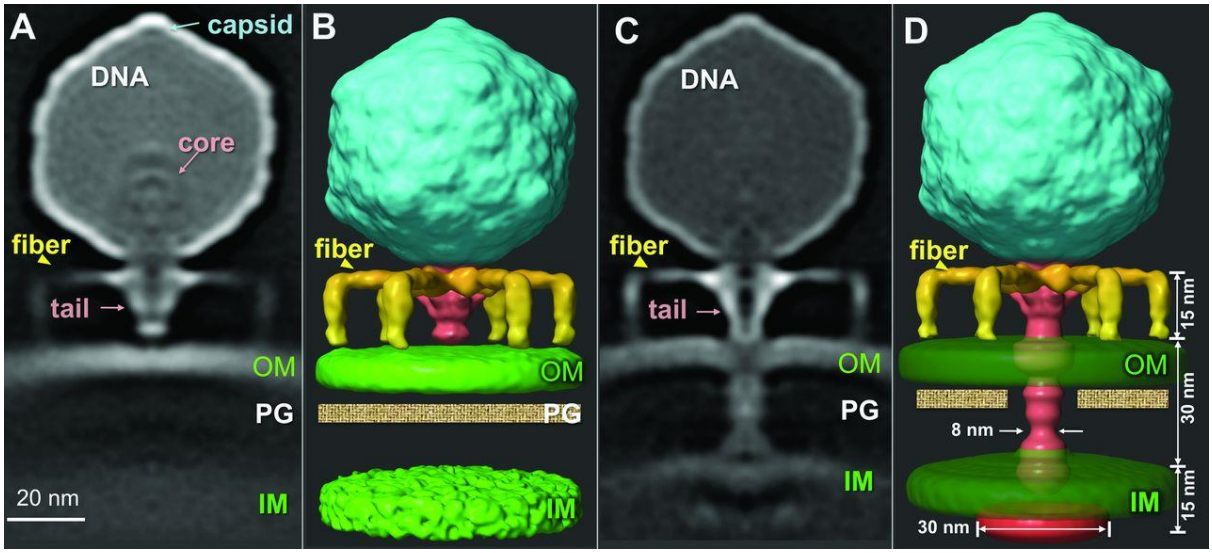


Figure 10. Host penetration by phage T7. Phage T7 binds to its host through side tail fibres, before ejecting the capsid core, composed of proteins gp14, gp15 and gp16, to span the bacterial envelope. Protein gp14 anchors in the outer membrane, while the lytic activity of gp16 degrades the bacterial PG layer. The phage DNA is partially ejected into the host, whereby the host polymerase inadvertently pulls the remainder of the phage DNA into the bacterial cytoplasm resulting in phage infection (figure sourced from (99)).

Siphoviridae, with their long non-contractile tails, are presumed to create a channel traversing the bacterial cell envelope by ejecting the Tape Measure Protein (TMP) from the core of their tail tube (101-103). While several experiments have indicated that the TMP of *Siphoviridae*, which dictates the length of their tails, also creates a channel for DNA injection, there is as yet no concrete evidence to support this hypothesis. This is due to the large diversity of *Siphoviridae* and complexity surrounding this stage of replication. The most compelling example for *Siphoviridae* TMPs forming the channel for DNA ejection into their host is from phage T5 (Fig.

11). The tip of *Enterobacteriaceae* phage T5 TMP protein, Pb2, was shown to have pore-forming activity *in vitro* (102). Further work, using cryo-tomographic electron microscopy, demonstrated T5 DNA injection into membrane vesicles possessing T5's proteinaceous receptor FhuA (104). In addition, Pb2 of T5 possesses a C-terminally located muramidase domain which facilitates degradation of the host PG layer (105).

Upon penetrating a bacterium, the phage genome replicates either through the lytic cycle, during which phage-encoded functions force the host to dedicate its metabolic and biosynthetic activities to the procreation of progeny phages, or to establish lysogeny, whereby the phage genome integrates into the bacterial genome and replicates *in situ* as a consequence of the bacterial division process. Phages capable of a lytic life cycle only are considered virulent phages, while phages that can undergo a lysogenic life cycle (in addition to lytic propagation) are referred to as temperate phages. Lysogenic phages are referred to as prophages when they are present in a dormant state within the genome of their bacterial host.

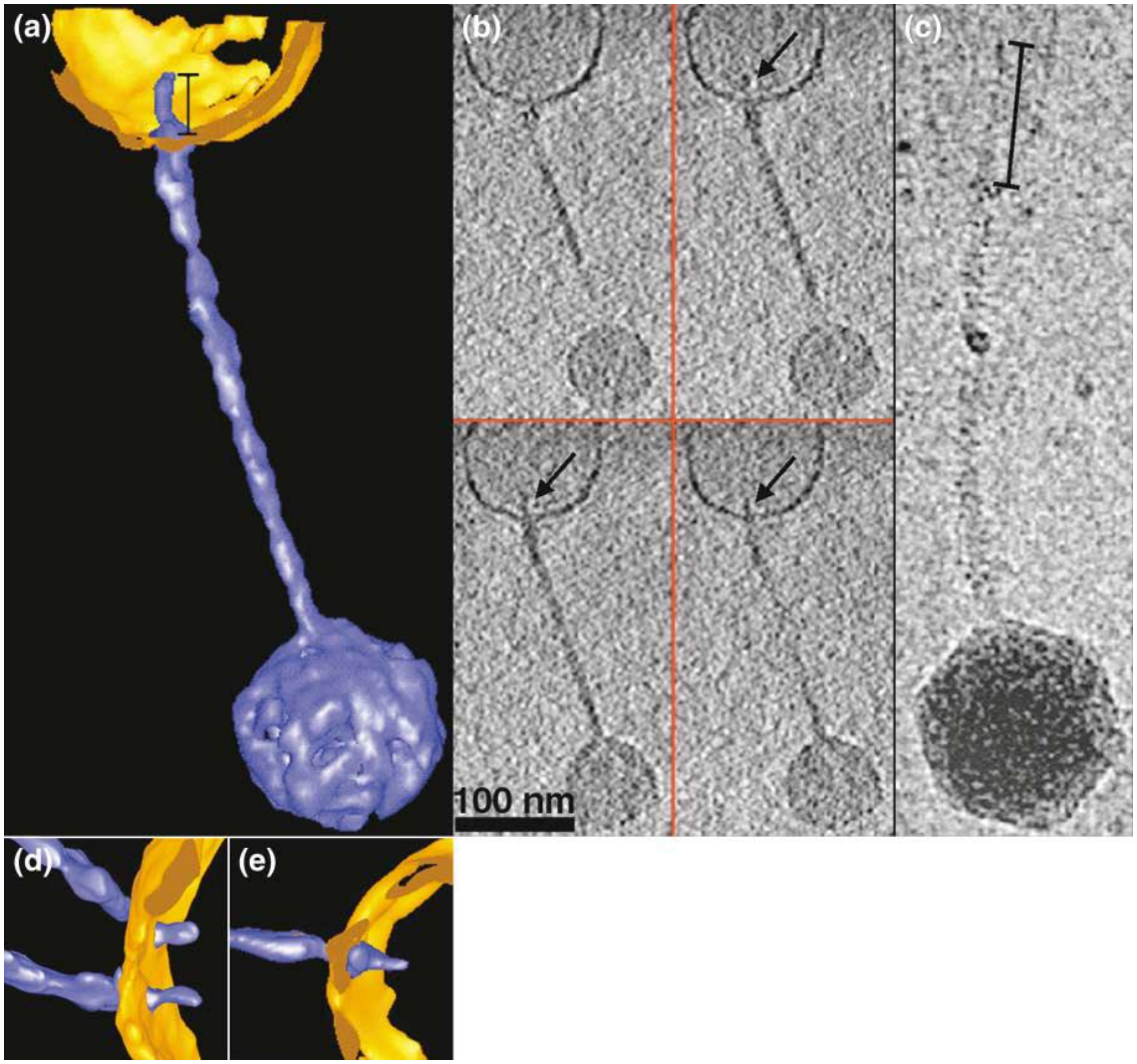


Figure 11. Adsorption and penetration of phage T5 into proteoliposomes containing FhuA, the phage's proteinaceous receptor. Panel (c): electron micrograph of phage T5, displaying the tail tip, which is approximately 50 nm long and 2 nm wide; Panel (b), multiple subpanels: upon adsorption to FhuA containing proteoliposomes, T5's tail tip decreases in length to approximately 23 nm and widens to 4 nm Panels (a), (d) and (e): de-noised single particle analysis reconstructions showing the tail tip of phage T5 penetrating the lipid vesicle (figure sourced from (104)).

Phages which possess the necessary genes to establish lysogeny can integrate into a bacterial chromosome via random transposition (eg. phage Mu) or via site-specific recombination (e.g. phage λ). Transposition is catalysed by transposase proteins that bind specific sequences at the transposon ends, and then coalesce these ends into a transposome complex. Mu phage transposition into the host genome occurs via a non-replicative integration mechanism, whereby the transposome complex catalyses DNA cleavage of the host genome and the strand transfer steps inserting Mu's genome randomly into the host chromosome (106, 107). Site-specific recombination by phages such as λ occurs through particular attachment sequences present on the bacterium (*attB*) and on the phage (*attP*). The recombination reaction is driven by an integrase or recombinase enzyme (108).

The genetic switch controls the transcription of genes required to establish and maintain phage lysogeny or proceed to produce phage progeny via the lytic life cycle. One promotor of this genetic switch controls the expression of a repressor protein, which binds to, and interferes with, a second promotor sequence that controls transcription of genes required to proceed to lytic replication (Fig. 2). Upon entering a host bacterium, the initial phage 'decision' to establish lysogeny or proceed through a lytic life cycle is based on the host cell's physiological state, influenced by various environmental conditions (23). When a temperate phage is integrated as a prophage, it often imparts beneficial traits to its host. Prophages can encode phage defence mechanisms to prevent super-infection (109), and even pathogenicity factors to increase the host's fitness (110). The expression of a repressor protein, a lysogenic conversion factor, also prevents homologous phages in the environment from 'super'-infecting the bacterium.

Prophage induction, such as phage λ, occurs following DNA damage. *E. coli* RecA protein recognises DNA damage and becomes activated, which stimulates autoproteolytic activity in the LexA repressor that results in the *E. coli* SOS response. LexA repressor and λ repressor protein (cI) are structural homologues and RecA also stimulates λ cI proteolysis, signalling to λ that physiological conditions within its host are unfavourable (111).

The chromosomes of phages are organized into specific functional modules and their expression proceeds in a controlled manner (112). Each module controls a distinct functional activity of the phage, and commonly encountered modules include the lysogenic module (as discussed above), synthesis module, structural module, and lysis module. This specific order ensures careful control of lysogeny (if applicable), production of multiple copies of the phage genome, packaging and protection of the phage chromosome within a capsid, production of the completed virion (that is specialised in finding a new host), and production of proteins to lyse the bacterial host cell and release phage progeny into the environment.

Phage progeny production and assembly is a complex process which varies considerably between different phages. Many phage genomes encode the necessary factors required to completely synthesize a new virion, while other phages are heavily dependent on host factors (113). The assembly of tailed phages, for which much is known, predicts that the phage's capsid structure and tail structure assemble independently and are subsequently joined to form a functional virion (114).

Proteins encoded by the phage lysis module, typically a holin and endolysin, are responsible for rupturing the host bacterium and releasing mature phage progeny (Fig. 12). Endolysins accumulate into the cytoplasm of the host bacterium, and are

conformationally ready to degrade the bacterial cell wall. However, holins are produced and accumulate in the membrane until a specific 'programmed' time. At a critical point, holin oligomers form, creating pores that function to permeabilize the bacterial membrane and release the endolysins to complete the phage infection cycle (115).

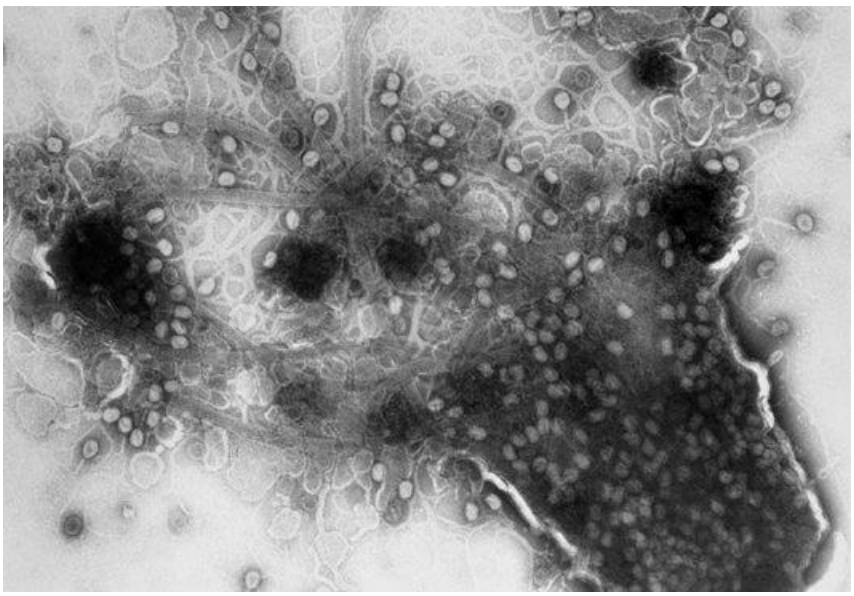


Figure 12. Example of bacterial lysis and phage progeny release (figure sourced from R. Bijlenga, Biozentrum, University of Basel though web link: www.sciencephoto.com/media/249780/view).

Phages released from the infected bacterium typically kill the host, releasing newly formed infectious virions. These phage particles are inert, non-living entities, but are primed to adsorb to a host bacterium should they collide, initiating another round of the phage infection process.

SECTION II.V LAB Phage-Resistance Mechanisms

Advances in our understanding of LAB and their infecting phages have helped in the rational selection of phage starter cultures, their rotational implementation and management, and the development and design of dairy factories. These modern practices have industrialized the dairy industry from what was traditionally an artisan process. Probably the clearest examples demonstrating why phages remain problematic for the dairy industry is demonstrated by studies examining bacterial-encoded phage-resistance mechanisms and subsequent phage circumvention thereof.

Bacterial acquisition of a phage-resistance mechanism represents a key innovation, which can render the majority of phages unable to replicate. For a period, this emergent bacterial population multiplies unchecked by phage infection. However, while phage replication produces predominantly homogeneous descendants, a minority of the population will accrue point mutations, genomic deletions, insertions, or combinations thereof. Many, if not most, of these mutations may negatively affect phage replication fitness; however, if one of the mutated phage descendants is capable of infecting a previously phage-resistant bacterium, this genetic variant becomes the fittest and thus repopulates the ecological niche. As before, this phage produces a heterogeneous mix of descendants whose evolutionary trajectory is determined, in part, by numerous complex phage-host interactions.

Bacteria have evolved systems to block phage attack at numerous stages of the infection process, while phages in turn have developed solutions to evade such bacterial defences (116). Naturally occurring *L. lactis* phage-defence mechanisms have been identified and characterized that act at virtually every stage of phage

development: from cell surface modification that prevents phage adsorption (50, 63, 69), superinfection exclusion (Sie) systems that prevent phage DNA injection (117, 118), restriction-modification (R/M) systems which target the phage DNA for degradation (59, 119), clustered regularly interspaced short palindromic repeat (CRISPR) systems which provide a memory of, and immunity to, previous phage infections, to abortive infection (Abi) systems that protect the bacterial population via suicide of the infected bacterium (119, 120). The section below focuses on phage-resistance mechanisms encoded by LAB, highlighting, where pertinent, systems that target the different stages of phage replication.

The first stage of phage infection, as discussed above, requires a phage virion to successfully recognise and adsorb to a cognate host bacterium. The selective pressure of phage infection often results in natural bacterial mutants that alter their cell surface composition, and such LAB mutants are commonly utilized by dairy industries as bacteriophage-insensitive mutant, or BIM, starter cultures. However, lactococcal BIMs with changes in their cell surface are often negatively affected in their growth properties compared to the parental strain (69) (Ainsworth *et al.*, unpublished results).

During the 1980's and 90's, several plasmid-encoded phage adsorption inhibition systems were identified in *L. lactis* strains (121-124). Strains harbouring these plasmids were reported as producing a more 'fluffy' pellet upon centrifugation. This was attributed to a cell surface coating, which could be observed under electron microscopy, which apparently blocked or masked the phage receptor and prevented phages from efficiently adsorbing to the surface of lactococcal strains. Recently, cell surface coatings of lactococcal strains conferring phage resistance were described as

an exopolysaccharide (EPS) layer (125, 126). Whether the original lactococcal adsorption inhibition systems were due to EPS production is currently unknown, as there exists a number of other mechanisms utilized by bacteria to avoid phage adsorption (for a comprehensive review, see (119)). However, similar to EPS layers, the LAB surface-coating layers described in the initial phage adsorption inhibition system studies could be easily removed by mild alkali treatment. When strain SK110, possessing a plasmid which was shown to encode a phage adsorption inhibition system, was given a mild alkali treatment the phage adsorption efficiency returned to levels comparable to the plasmid-free strain (123).

Host-encoded defence mechanisms have been characterized in several LAB that prevent phage host-penetration. The best studied example in a lactococcal strain is the Sie protein of phage Tuc2009 (Sie_{2009}), produced during lysogenic replication of the P335 species phage. Sie_{2009} is encoded by Tuc2009's lysogenic module and situated between the genes encoding the repressor and integrase proteins, which is a common gene constellation found in genomes of lactococcal phages that belong to the P335 species. Sie_{2009} is a membrane-associated protein that was shown to interfere with phage DNA injection, but not phage adsorption. Sie_{2009}, expressed *in vitro*, is capable of protecting its host bacterium from superinfection by several 936 species phages, and was shown to prevent phage infection even when the host bacterium was challenged with 10^9 phage per ml (118).

Similar to Sie_{2009}, the temperate *S. thermophilus* phage TP-J34 encodes a lipoprotein, Ltp, within its lysogeny module which is believed to be directed to the cell surface through signal sequence secretion. This is consistent with its detection on the surface of *L. lactis* protoplasts, which had been stripped of their cell wall

layer, by immunogold labelling. Interestingly, when Ltp is expressed in an *L. lactis* host, it protects *L. lactis* from infection by a 936 species phage P008 (127). Recently, the crystal structure of $Ltp_{TP\text{-}J34}$ was solved. While the N-terminus of Ltp anchors the superinfection exclusion protein to the cytoplasmic membrane, the soluble C-terminus is formed by a tandem of three-helix helix-turn-helix (HTH) domains. Mutants of phage P008 which could overcome the protective effect of Ltp were sequenced, and mutations were identified within the phage TMP. The TMP of phage P008 has a strong positive charge (pI > 9.0), while the C-terminal surface charge of Ltp is highly negatively charged. These results are congruent with Ltp interacting with P008 TMP, thereby interfering with normal phage host-penetration and DNA injection processes (128).

R/M systems are ubiquitous in bacteria and are an important mechanism employed by bacteria to control the acquisition of foreign DNA (129). A typical R/M system requires three components: (i) a specific DNA sequence, (ii) an restriction enzyme which recognizes the specific DNA sequence if it is unmethylated thereby targeting such DNA for its endonuclease activity, and (iii) a cognate methylase which methylates a base in the recognition sequence, thereby preventing the corresponding DNA from being recognized and cut by the endonuclease (130). R/M systems are designed to degrade foreign DNA entering the cytoplasm of a bacterium. Bacterial DNA is protected from its own R/M system, even after synthesis of a new DNA strand, as at least one strand is methylated (i.e. hemi-methylated). Methylases are more specific for hemi-methylated DNA. This is important as both the methylase and restriction enzyme are active in the cell at the same time, and their relative activity is important in determining the fate of foreign DNA, as phage DNA which

becomes methylated before an endonuclease degrades it could result in phage replication and bacterial death.

As soon as a phage successfully infects an R/M system-possessing bacterium, the DNA of its progeny becomes methylated. Phages infecting a bacterial population possessing an R/M system may also evolve point mutations to avoid endonuclease-recognition (131, 132), while phages have also been described that have acquired a methylase gene from the host chromosome and can thus alter between infecting bacteria possessing or lacking a specific R/M system (133, 134).

As highlighted above, the protection provided by R/M systems is fundamentally dependent on specific DNA sequences. Phages which evolve point mutations in their recognised sequences become resistant and avoid nuclease degradation. The mutation rates for dsDNA organisms were observed by Drake *et al.* (1998) as remarkably similar in diverse microbes. A mutation rate per genome per replication (m/g/r) of approximately 0.0034, or 1 bp change per 10^{10} nt, was observed for bacteria (and temperate phages) (135). However, lytic dsDNA phages possessed an m/g/r under laboratory conditions of ~0.004 (~1 bp change per 10^{8} nt). The increased mutation rate is probably because a lytic virus is replicated numerous times per infection (135, 136). This explains how, under successive rounds of replication, phages so quickly acquire the necessary point mutations to overcome sequence-specific bacterial defence mechanisms such as R/M systems.

The accumulation of point mutations throughout a phage genome to avoid R/M systems is often evident by the lack of the relevant restriction sites in their genome. For example, *L. lactis* 936-type phages jj50 and sk1 possess 14 and 11 LlaAI restriction sites (GATC) in their 27.5 kb and 28.5 kb genomes, respectively,

while the rare emerging lactococcal phage P087 possesses only 2 LlaAI sites in its entire 60.1 kb genome (reference, this work). This requires evolutionary selection over multiple generations, which is facilitated by the inherent leakiness of R/M systems (130).

The genes encoding many R/M systems of LAB have been identified on plasmids (137). While LAB plasmids encoding R/M systems are still being identified and characterized (138-140), the number of chromosomally encoded R/M systems detected in LAB is increasing due to advances in genome sequencing (141-143).

Approximately 40 % of all bacteria possess one or more CRISPR loci and CRISPR-associated (*cas*) genes (144). The following outline of the CRISPR/Cas system is taken from the extensive reviews on this bacteriophage-resistance system (145, 146), except where indicated. A CRISPR locus is composed of 2 to 375 DNA repeat sequences (21-48 bp), each separated from the next by a unique DNA sequence called a spacer (21-72 bp). Following entry of foreign genetic material into the bacterial cell, such as an unsuccessful phage infection, part of the invading DNA can be incorporated into the 5' end of the CRISPR locus as a new spacer. The incorporation of such a new spacer, flanked by repeat sequences, is facilitated by the so-called Cas proteins.

The CRISPR locus is transcribed from a promoter in an upstream leader region (20-534 bp). This leader region is also believed to be important in the acquisition of new spacer sequences, as new spacers were observed to be orientated in a conserved manner next to the leader region (147). The CRISPR transcripts are processed into CRISPR RNA (crRNA) at the repeat sequences to yield individual

spacer sequences with part of a repeat sequence at its 5' end. The spacer sequences of the crRNA direct Cas proteins to protect the bacterium against invasion by foreign genetic material containing complementary sequences to the crRNA. Both DNA and RNA have been observed as the target of different CRIPSR/Cas systems, highlighting the mechanistic diversity of the system (148, 149).

The various CRISPR/Cas subtypes all contain two universal *cas* genes (*cas1-2*), four conserved *cas* genes (*cas3-6*) and one or more of nine subtype-specific *cas* gene-sets, that may function as nucleases, helicases, RNA binding proteins, *etc.*, in various organisms (150, 151). Additional *cas* genes, loosely associated with the CRISPR loci, are also observed in genomes containing a given CRISPR/Cas system. These are called repeat associated mysterious proteins (RAMPs) and are only present in genomes with a CRISPR/Cas system (152). The different *cas* genes facilitate both the acquisition of new spacers (adaptation stage) and the neutralization of invading genetic elements (interference stage) (145).

CRISPR cassettes are frequently spread via horizontal gene transfer (HGT) (153, 154). A number of sequenced megaplasmids (> 40 kb) contain CRISPR loci and *cas* genes, and similarly several chromosomal CRISPR loci have been located near plasmid-associated genes (153). When published, Horvath *et al.* (2009) observed only one sequenced LAB CRISPR locus that was present on the *E. faecium* pHT beta plasmid (155, 156), although at least one other was discovered since then (see below).

CRISPR elements were not detected in the sequenced genomes of *Lactococcus*, *Leuconostoc*, or *Pediococcoccus* (156). It was highlighted that the absence of CRISPR in these bacteria may have been strain-specific and not

indicative of the species. This may indeed be the case, as recently a strain of *Lactococcus* was discovered to possess a plasmid-borne CRISPR system. However, CRISPRs are not common in *L. lactis* strains, as this plasmid-borne CRISPR system was identified as the only positive from a screen of 383 other industrial strains (157).

Barrangou *et al.* clearly showed that the adaptation of *S. thermophilus* to virulent phages is via the acquisition of new spacers at the leader end of the CRISPR locus (158). The phage genomic sequences which became spacers, called proto-spacers, were not randomly selected. Proto-spacer adjacent motifs (PAMs) were identified flanking the proto-spacer (159). In *S. thermophilus*, different CRISPR loci were clearly shown to have different PAM sequences: the PAM sequence of locus 1 (Sthe1) was shown to be AGAAW, while that of locus 3 (Sthe3) is GGNG (156, 160).

LAB possessing CRISPR elements may still become sensitive to phage infection. A single point mutation in the phage proto-spacer or neighbouring PAM sequence was shown to result in *S. thermophilus* reversion to a phage-sensitive phenotype (159, 161). This knowledge of CRISPR/Cas systems and phage resistance has allowed the rational design of BIMs of LAB for industrial application (159).

As previously mentioned, certain LAB strains do not possess CRISPR elements which would greatly improve their phage resistance. Recently, research has shown that the *S. thermophilus* Sthe3 CRISPR locus can be transferred to *E. coli* and retain its ability to provide interference against incoming plasmids and phages. The engineered plasmid could successfully interfere with plasmid transformation by more than five orders of magnitude and λ infection by three orders of magnitude. The authors noted how this potential allows for the development of phage-resistant

strains, and strains that are much less likely to incorporate and disseminate undesirable plasmids (162). In lactococcal dairy strains, however, many desirable technological properties (lactose utilization, peptidases, phage resistance mechanisms, *etc*.) are plasmid-encoded, and perhaps this is why lactococcal strains used in the dairy industry as a general rule do not possess CRISPR/Cas systems.

The CRISPR/Cas system, which is ironically spread via HGT, is a host defence mechanism against invading genetic material (such as phages, plasmids and transposons). As spacers in the CRISPR loci nearest the leader sequence are the most recently acquired, this provides a historical account of recent phage infections in the ecosystem. Loss of spacers, likely via homologous recombination, limits the size of the CRISPR loci (163). The CRISPR/Cas phage-defence mechanism reflects host-prey interactions that drive an evolutionary race between bacteria and phage, through the maintenance and renewal of CRISPR spacer sequences in the case of the bacterial host and the acquisition of mutations in the phage genome that allow escape from the CRISPR system. As mentioned, the CRISPR system acquisition of new spacers is believed to require an unsuccessful phage infection. Thus, the CRISPR/Cas system may work in tandem with other host defence mechanisms to limit phage infection.

Abi systems are found in a diverse range of organisms (164, 165); however, they have been extensively studied in *L. lactis*. The dominance of research pertaining to lactococcal Abi systems may have been influenced by the following factors: (i) *L. lactis* is an economically important bacterium for the production of various cheeses, (ii) phage infection of starter strains in dairy fermentation plants is an ever-persistent problem (57), (iii) significant data is available relating to the three main species of

lactococcal phages, c2, 936, and P335 (64), (iv) lactococcal strains often possess multiple plasmids that possess technologically desirable properties related to industrial production of cheeses (lactose utilization, proteases, bacteriocins, *etc.*), including Abi systems, which may readily transfer by conjugation, and (v) plasmid localization of many Abi systems has facilitated the relatively easy identification of the gene, or genes, involved in phage-resistance phenotype (120).

Abi systems, also called phage exclusion systems, refer to a broad collection of protein-mediated anti-phage mechanisms, which function to provide host protection by preventing normal phage multiplication. In several cases, Abi systems have been shown to inhibit phage proliferation by prematurely killing the infected cell. This altruistic nature of Abi systems sets them apart from other phage-defence mechanisms. Furthermore, unlike R/M systems, Abi systems are effective against phages regardless of prior phage-host interactions (i.e. phages become resistant to R/M systems after a single propagation due to methylation). Abi systems typically result in a drop in efficiency of plaque formation, burst size, and efficiency of formation of centres of infection (130).

If the action of a particular Abi system results in death of such an Abi-encompassing bacterial cell, it complicates the characterization of the mode of action of the system. Twenty three different Abi systems have thus far been identified against lactococcal phages (166). With just a few exceptions, they have been named alphabetically AbiA through to AbiZ (there is an AbiD but also an AbiD1; while there are no Abi systems that have (yet) been assigned the names AbiM, AbiW, AbiX, or AbiY). The majority of the characterized Abi systems operate through a single protein; however, Abi protection may be mediated by two proteins (in the case

of AbiE, AbiG, AbiL, AbiT, AbiU) or by multi-component systems (AbiR). The lactococcal Abi systems characterized to date highlight that almost all aspects of phage multiplication can be targeted; DNA replication, transcription, and synthesis of structural proteins (for a review, see (120)).

SECTION III Siphoviridae Tail Structure and Function

Bacteriophages are the most abundant biological entities in the biosphere, with 10^{31} prokaryotic viruses estimated on earth, outnumbering bacteria by an order of magnitude (53). *Siphoviridae*, referring to phages with a dsDNA genome and a long non-contractile tail, are the dominant viral morphotype (73). The abundance of *Siphoviridae* may be indicative of their efficiency at infecting bacteria.

Siphoviridae tails are multi-proteinaceous complex machines that act in a pre-programmed and coordinated manner to distinguish and adsorb to their host's receptor, penetrate the bacteria's cell envelope, and direct DNA ejection into their host's cytoplasm. Several *Siphoviridae* phages have been analysed in depth with respect to their tail structure. Enterobacterial phages λ and T5, lactococcal phages Tuc2009, TP901-1 and p2, and *Bacillus subtilis* phage SPP1 have all served as models to advance our understanding of *Siphoviridae* tail structure and function. Here we review *Siphoviridae* tail machines, with emphasis on these model phages, so that the structural and functional significance of their tail structures can be interpreted and applied to other *Siphoviridae* phages, despite divergent DNA sequences.

SECTION III.I Tail Length Determination

The length and flexibility of *Siphoviridae* tail structures may increase host recognition efficiency. In support of this hypothesis, a correlation between tail length and environment levels of biological activity was noted for various phages in aquatic ecosystems, with oligotrophic waters containing phages with longer tails than meso-

and eu-trophic waters (167). Ackermann (1998) reported that *Siphoviridae* tails range from 79-539 nm, although recently *Siphoviridae* with tails greater than 800 nm have been described (74, 168).

The length of a *Siphoviridae* phage tail is determined by the phage tape measure protein (TMP) (169-171), which acts as a molecular ruler, whereby specific repeating sequences are believed to act as a scaffold for the assembly of the final phage tail tube. Analysis by Belcaid and colleagues (2011) highlighted that many *Siphoviridae* TMP proteins possess specific repeat motifs, with the amino acids (aa) tryptophan or phenylalanine occurring either every 11 residues or following a pattern of 11-11-18 residues (172).

In-frame deletions and duplications of the TMP have been shown to proportionally decrease or increase the length of λ and TP901-1 tails (Fig. 13; (170, 171)). The TMP of phage λ measures 32 Major Tail Protein (MTP) disks. Katsura and Hendrix (1984) calculated that one amino acid of the TMP template protein roughly correlates to 0.15 nm length of the phage tail (169). Notably, ten 40 aa repeating motifs with regularly spaced aromatic residues within the TMP of lactococcal phage p2 are present and predicted to adopt a curved alpha-helix, which is assumed to determine the length of the lactococcal 936 phage tail (173).

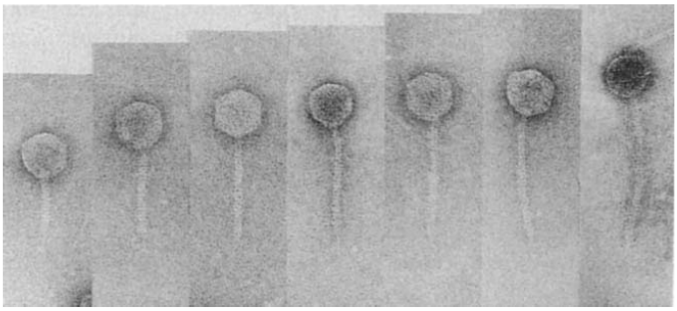


Figure 13. Tail length determination by phage λ TMP. The number of major tail protein rings, from left-to-right, are; 13, 21, 25, 26, 30 and wild type 32 (figure sourced from (174)).

Phages often possess a six-fold symmetry of their tail tube and also contain six copies of their TMP per mature phage particle (175, 176). Proteolytic processing of *E. coli* phage λ TMP occurs before the assembled head and tail structures are connected (177, 178). Maturational cleavage of the TMP also seems to take place in the *L. lactis* phage Tuc2009, although apparently processed and unprocessed proteins were detected in the mature virion (179).

SECTION III.II Initiating Tail Assembly

The tail tips of *Siphoviridae* phages frequently possess a long tail fibre, which plays an important role in tail morphogenesis. For example, the association of three copies of phage λ's gpJ tail fibre protein represents the first requisite step for the formation of a so-called initiator complex that acts as a hub for tail assembly (Fig. 14). The gpJ trimer is further complexed with proteins gpI, gpL and gpK (180), followed by the addition of six copies of TMP, or gpH, which are associated with

two chaperone proteins, gpG and gpGT (181, 182). The initiator complex is completed by gpM addition and stabilization (180).

Lactococcal phages Tuc2009 and TP901-1 share ~78-97 % identity in the proteins that make up their respective distal tail structure (176). These high levels of homology, both in terms of gene synteny and corresponding amino acid composition, have allowed extrapolation of the experimental data obtained for Tuc2009 and TP901-1, which in turn has resulted in a very comprehensive model regarding the tail structure and function of either of these phages (183). Tuc2009 and TP901-1 both possess a 135 nm tail (171), and morphogenesis of these respective tails, like phage λ, is believed to be dependent on the C-terminal portion of the TMP, the Distal Tail protein (Dit) and the N-terminal portion of the tail fibre or Tail-Associated Lysin (Tal) protein (184), as these three proteins are proposed to form a so-called initiation complex (171, 185). In agreement with this notion, it has been shown that phages deficient in any of these three proteins produce tail-less DNA-containing capsids (184). Recent work suggests that the N-terminal domain of Tal may not be necessary for assembling the baseplate of phage TP901-1 (see below), although it may be required for anchoring the assembled Dit-baseplate structure to the tail tip (186).

The TMP C-terminus, Dit and tail fibre N-terminus of TP901-1 and Tuc2009 share 24 %, 32 % and 26 % amino acid identity to its equivalents encoded by *Bacillus subtilis* phage SPP1, respectively, implying structural and functional parity between these phages. Veesler *et al.* (2010) have shown that the Dit protein is a central docking component in many Gram-positive phages (187). The SPP1 Dit equivalent is a donut-shaped hexamer, forming an internal central channel

approximately 40 nm wide. The internal diameter and negative surface charge of Dit are both congruent with DNA egression through Dit. The model for SPP1 initiation complex assembly proposes that several copies of the TMP bind at their extremity to the internal surface of the Dit hexamer, and protrude towards the capsid, while a dome-shaped trimer of gp21 N-terminus (the SPP1 tail fibre protein) associates with the Dit hexamer's tail tip exposed surface, capping the tail tube channel (187).

SECTION III.III Tail Tube Formation

Following the formation of a *Siphoviridae* initiator complex, the TMP is assumed to serve as a scaffold for tail tube formation. The proteins gpG and gpGT are not associated with mature *Siphoviridae*, but are proposed to function as chaperones for tail morphogenesis and major tail protein (MTP) polymerization. The ratio of gpG to gpGT is believed to be an important factor in controlling tail formation, with gpGT produced at an approximate ratio of 1:30 relative to gpG production (188). Controlling the relative abundance is achieved by means of a –1 ribosomal frameshift during mRNA translation at a specific slippery sequence located at the 3' end of the *gpG* gene, causing translational coupling to the *gpT* ORF to form protein gpGT. Production of gpGT analogues using sequence-specific translational frameshifts is conserved amongst many long-tailed bacteriophages (188, 189).

Recent publications on gpG and gpGT of phage λ have shown that both proteins are essential for correct assembly of its tails. Plasmid-induced complementation assays showed the abundance of gpG:gpGT was essential for

efficient tail production, which explains the important conservation of the frameshift mechanism for gpGT production in *Siphoviridae* (190). In a separate publication Xu *et al.* (2013) demonstrated that the "G" region of gpGT binds to the TMP of phage lambda, gpH, while the "T" domain of gpGT binds the MTP, gpV. Protein gpG binding to gpH (and also binding of the "G" domain of gpGT) is apparently evenly distributed across the whole length of λ TMP. However, the gpH-gpG and gpH-gpGT-gpV interactions remain poorly understood and structural analysis of the TMP protein is likely a key requirement to elucidating more of the tail tube assembly process (191).

The crystal structure of lactococcal phage p2 ORF12 shows that it is an orthologue of gpG. ORF12 crystallizes to form side-by-side spirals in solution (173). The spiral structure of tail-associated chaperone proteins was recently shown to be conserved in the enterobacterial phage HK97 (192). Regularly occurring hydrophobic pockets in the ORF12 oligomer of phage p2 correspond to the predicted positions of repeating aromatic residues along the curved α-helical region of TMP. It is predicted that the ORF12-TMP complex maintains the TMP hydrophobic helix in its correct shape and in solution until the MTP tail tube is formed (173).

For some phages it has been shown that MTP subunits form hexameric rings encircling the TMP, with such rings stacking along the length of TMP towards the capsid (for examples, see (175, 179, 193)). The polymerization process of phage λ MTP, gpV, possibly displaces the chaperones gpG and gpGT bound to the TMP; although there is currently no direct evidence for this. Current knowledge indicates that gpV polymerization on the gpH TMP scaffold pauses when the N-terminus of TMP is reached. The tail terminator protein of phage λ, gpU, which forms a

hexameric structure in phage λ and SPP1 (194, 195), is known to associate with the top ring, thereby preventing aberrant tail tube polymerization (196, 197). The tail tube of phage λ is completed by proteolysis of gpH to gpH*, changing the molecular weight of this protein from 90 to 78 kDa, and by the addition of protein gpZ whose function is currently unknown (182, 194, 196). The completed λ tail structure can spontaneously assemble with a completed λ capsid structure to form an infectious virus (Fig. 14; (114)).

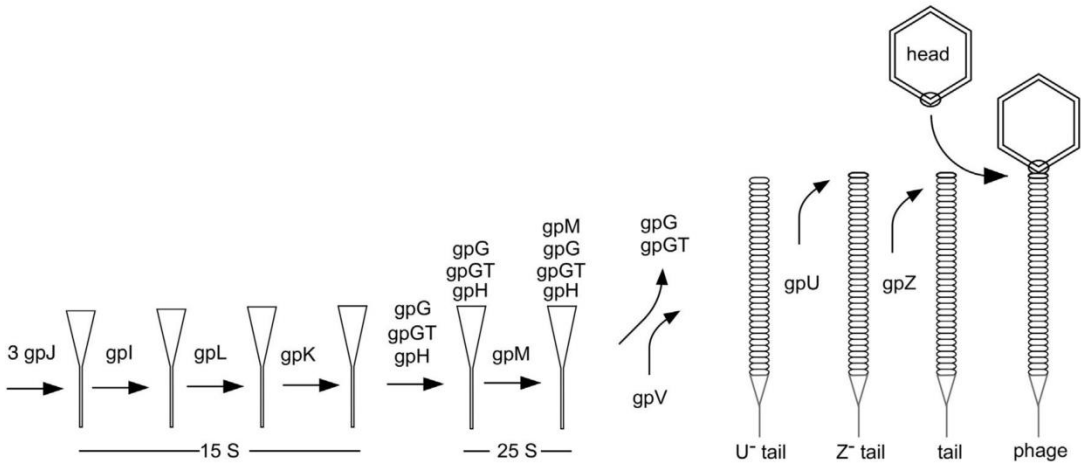


Figure 14. Cartoon representation of phage λ tail assembly (figure sourced from (190)).

The enterobacterial phage T5 possesses a 250 nm long flexible tail. Five or six copies of the 121 kDa protein, Pb2, form both the TMP and straight tail fibre (105). The 40 Pb6 MTP rings encircling T5 TMP are unusual amongst long tailed

phages as they are reported to have a trimeric and not a 6-fold symmetry, the significance of which is currently unknown (198).

The tail tube of phage SPP1 is the product of two MTPs, gp17.1 and gp17.1*. Both MTP proteins possess the same amino terminus, however, gp17.1* is the product of a +1 translational frameshift at a programmed sequence, adding approximately 9.1 kDa to the carboxyl terminus of gp17.1. This translational frameshift produces gp17.1 and gp17.1* at a ratio of approximately 3:1. Both gp17.1 and gp17.1* can solely be used in the production of tail tubes. However, tail tubes produced simply of gp17.1 are smooth and featureless, while gp17.1* only tail tubes are 'hairy' (199).

Plisson *et al.* (2007) reported on the structure of the MTP tail tube of SPP1 before and after phage receptor recognition. SPP1 MTP is composed of four domains, D1-D4. The signal generated by the phage anti-receptor causes MTP domains D1 and D4 to rotate clockwise by 14^{o} toward the tail axis, and D2 to bend slightly towards the tail axis. The alterations observed in D4 causes significant changes in D3 of the same MTP subunit, causing D3 to tilt to a vertical orientation. The movement of D3 reduces the tail internal diameter by 25 %, from 56 to 42 Å. Each ring of MTP is rotated approximately 21^{o} to the previous ring, and the movement of D3 induces an alteration in D4 of the next MTP ring. Signal transmission following receptor binding proceeds via each of the six MTP subunits per ring in a domino-cascade away from tail tip, opening the head-to-tail connector for DNA expulsion (193).

The tail tube of phage TP901-1 is composed of 33 stacked MTP rings. Like SPP1, each MTP ring of TP901-1 is composed of six subunits, however, each MTP

ring is rotated relative to the previous ring by 22.4^{o} (200). While the bacterial receptor recognized by TP901-1 has recently been elucidated (Ainsworth *et al.*, unpublished results), no comparison of the pre- and post-DNA ejection states for phage TP901-1 is available.

SECTION III.IV Phage Host-Recognition

The C-terminus of phage λ tail fibre, gpJ, functions as a receptor binding protein (RBP). For DNA injection by wild type phage, λ requires the bacterial outer membrane protein LamB, involved in maltose uptake to the periplasmic space (201), and the inner membrane maltose phosphoenolpyruvate phosphotransferase proteins IIC and IID (202). Homotrimers of LamB produce an outer membrane porin, although λ-ejected DNA does not pass through the pore (202). Phage λ gpJ binds in a two-step process; reversible adsorption is followed by an irreversible binding which commits the phage to DNA ejection and inactivation (203). A gap between the phage tail tube tip and LamB receptor, approximately equivalent to the length of λ tail fibre (21 nm), is observed in reversibly bound phage. Irreversibly adsorbed phages are seen to be in direct contact with the receptor and their capsids are largely emptied of DNA (204).

Phage side tail fibre (stf) proteins increase the efficiency of phage adsorption. The chaperone protein, Tfa, is required for phage λ stf assembly and is present in the mature virion. The common laboratory isolate of phage λ, designated λ-PaPa, and isolated in the 1950's, is missing its stf proteins due a frameshift mutation. The

original isolate of phage λ, Ur-λ, possesses stf proteins and has been shown to absorb faster and more efficiently (182, 205).

Binding of enterobacterial phage T5 to its proteinaceous receptor FhuA is mediated through Pb5, a protein forming the conical structure at the end of the tail tube (206, 207). The addition of FhuA to T5 phages *in vitro* causes DNA ejection and phage inactivation, and the addition of recombinant Pb5 to cells competes and blocks T5 adsorption. The interaction of Pb5 with FhuA results in a strong irreversibly bound complex, and separating Pb5 bound to FhuA required protein-denaturing conditions stronger than were needed to denature either Pb5 or FhuA in isolation (208, 209). Pb5 binding to FhuA is thus believed to cause conformational changes in Pb5, and can be observed to cause major conformational changes in Pb2 straight tail fibre, shrinking it from 50 to 23 nm and doubling its width from 2 to 4 nm (Fig. 11; (104)).

A baseplate-like structure, almost twice the diameter of the tail tube, is situated between the terminal MTP ring and conical tip of phage T5 (Fig. 15; (198)). Two putative proteins, Pb3 and Pb4, are believed to form this structure (105, 210). Three L-shaped tail fibre (LTF) proteins, encoded by Pb1, are attached to the tip of the T5 tail at this baseplate structure (211). These L-shaped tail fibres of phage T5 are responsible for the adsorption to *E. coli* LPS, and, although non-essential for infection, this interaction facilitates efficient host adsorption and thus infection (198, 212).

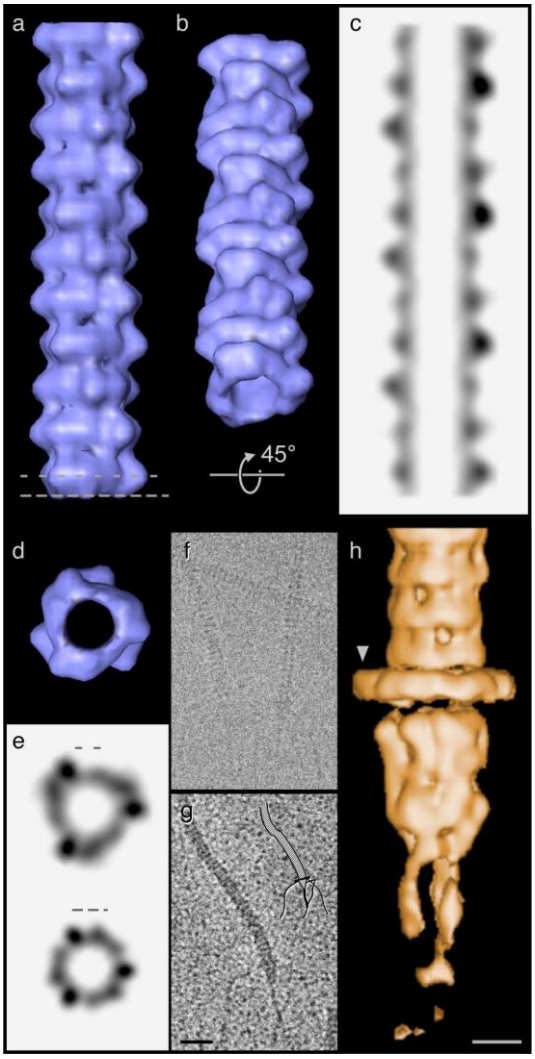


Figure 15. T5 tail and tail tip structure analysed by cryo-electron microscopy (EM). Images (a) and (b) show the surface of an 11-disc segment of the 40-disk tail with 3-fold symmetry, while (c) shows a longitudinal section. The image (d) shows the relative rotation of two major tail protein disks. Image (e) represents a cross-sectional view of the density map corresponding to the centre and edge of a disk, apparent in the banded appearance of the tail tube in (f) T5 micrographs. Picture (g) is a negatively stained EM image where the tail tip and fibres are evident. Image (h) shows a single particle analysis reconstruction of the tail tip with baseplate. Scale bars represent 50 Å (figure sourced from (198)).

The RBP of lactococcal phage p2 (belonging to the 936 species), represented by ORF18, was crystallized and structurally examined. The RBP was, similar to TP901-1, dissected into three domains: the N-terminal shoulder domain involved in anchoring the RBP to the phage tail tip, the linker or neck region, and the C-terminal host-recognizing head domain. The p2 RBP heads were, like TP901-1, found to be associated with a glycerol molecule in their crystal structure (107). The RBPs of p2 were also characterized in detail by examining antibodies which neutralized phage infection. These antibodies were effective in neutralizing 12 out of 50 tested 936 species lactococcal phages, and bound to head domains of phage p2 ORF18 at the cleft between the RBP trimer subunits (107, 213, 214).

Interestingly, addition of the cation calcium, known to be essential for many infecting lactococcal phages (215), was observed in crystal structures and EM density maps of phage p2 baseplate to cause conformational changes at the phage tail tip by rotating the RBPs by approximately 200^{o} to a 'heads-down' active conformation (Fig. 16; (216)). The baseplate is composed of two back-to-back ORF15 hexamers of the Dit protein, a trimeric assembly of ORF16 (orthologous to the N-terminus of TP901-1 Tal protein and SPP1 gp21), and six trimers of ORF18 RBPs. Phage p2 requires ORF16 for *in vitro* oligomerization of ORF15. The C-terminus of p2 ORF15 has an extended galectin-fold, creating 6 protrusions around the periphery that do not contact each other. Each extension, termed an 'arm', projects a three-digit 'hand' clear of the ORF15 hub. The arms of the upper ORF15 ring differ as they are in contact with each other, forming a continuous volume. In the inactive model of p2, the galectin hand of the lower ORF15 ring associates with the shoulder domain of a RBP homotrimer, and the upper ORF15 ring arms and

hands are in lateral contact with the head domains of RBP heads. Interpreting this inactive model, the RBPs are in a ‘heads-up’ orientation.

The reorientation of p2’s RBP direction during calcium activation causes the dome-shaped tail tube cap, composed of the N-terminus of three ORF16, to outwardly rotate and create an approximately 32 Å channel for DNA ejection. The N-terminus of the ORF16 proteins, previously not in any contact with RBPs, now shares a large contact area, keeping the RBPs in a heads-down conformation. A model for initiation of infection by the lactococcal phage p2 speculates that the primary interaction with the host cell surface LTA and/or cell wall polysaccharide (see (50) for information related to *L. lactis* cell wall polysaccharides), both of which are known to contain phosphate groups (and thus function as cation ‘sponges’), release calcium that triggers conformational changes in the baseplate to an activated form (Fig. 16; (216)).

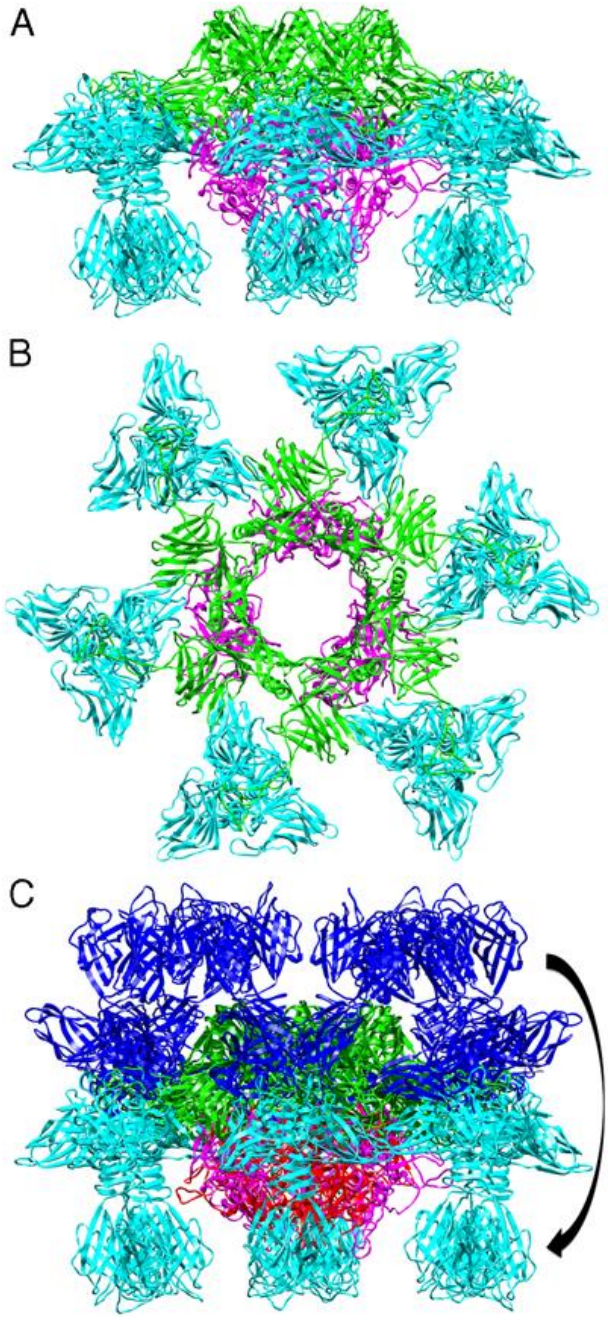


Figure 16. Crystal structure of lactococcal phage p2 baseplate and activation. ORF15 is green, ORF16 is pink, and ORF18 is blue. (A) Side view of p2 baseplate in a 'heads-down' active conformation. (B) Top-down view of p2 baseplate in a 'heads-down' active conformation. (C) Side view highlighting the rotation of p2 RBPs, ORF18, in their inactive 'heads-up' conformation (dark blue) and active 'heads-down' conformation (light blue) (figure sourced from (216)).

Phages such as Tuc2009 and TP901-1 recognize their bacterial hosts by a large proteinaceous complex, termed a baseplate, located at their tail tip (Fig. 17). The baseplate of TP901-1, with a calculated molecular weight of 1.8 MDa (217), constitutes a large part of the overall tail structure (estimated at 6.4 MDa by modifying Sciara *et al.*'s 2006 calculation for Tuc2009 tail structure, taking into account recent advances in our understanding on the baseplate (176)). The baseplates of Tuc2009 and TP901-1 are composed of six tripod or baseplate protein (Bpp) structures, associated around the periphery of a Dit hexamer structure formed during the initiator complex morphogenesis (184). The baseplate tripod assemblages of Tuc2009 and TP901-1 are both composed of upper baseplate proteins (BppU), forming the apex of the tripods, and the lower baseplate proteins (BppL), forming the three 'legs' of the tripods. The quaternary structure of each TP901-1 tripod is composed of a trimer of BppU and three trimers (one trimer for each leg of the tripod) of BppL (176, 213). The tripods of Tuc2009 differ slightly, as a BppU subunit is substituted by two copies of an associated baseplate protein, BppA (218).

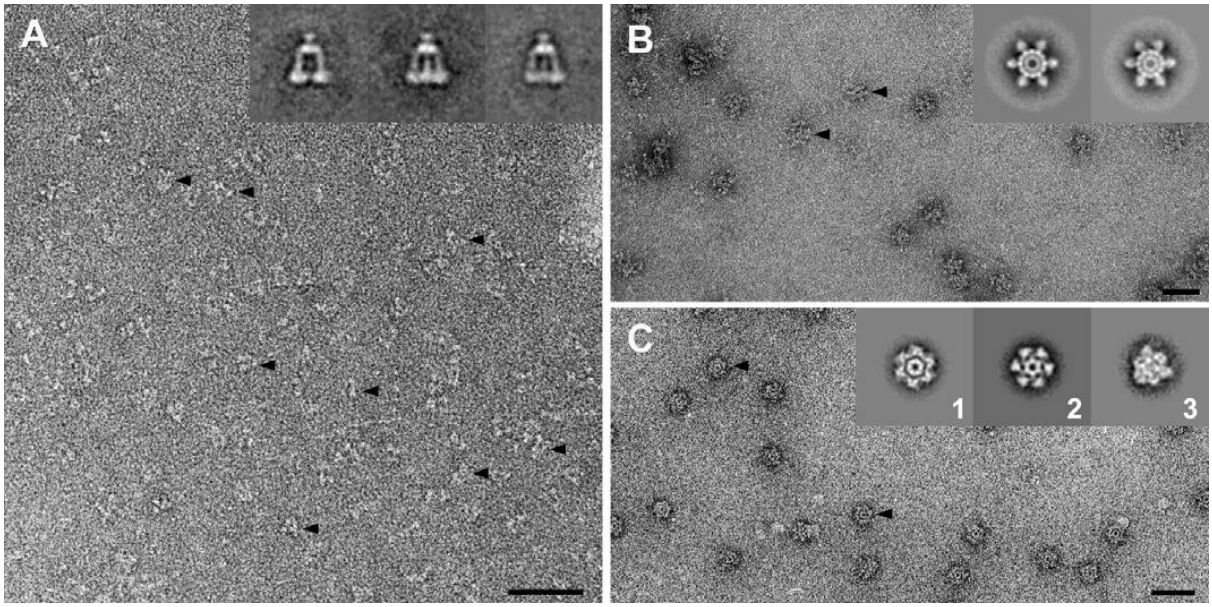


Figure 17. EM images of (A) TP901-1 BppU-BppL tripods, (B) TP901-1 Dit-BppU-BppL baseplates, and (C) p2 ORF15-ORF16-ORF18 baseplates. Scale bar is 500 Å. Inset images are computer-generated class averages of several observed structures (figure sourced from (186)).

The lower baseplate protein (BppL) of Tuc2009 and TP901-1 functions as the RBP. Both phages possess 18 homotrimers of BppL, representing the 18 tripod legs (three for each tripod, with six tripods present on each baseplate), providing each phage with a total of 54 RBP molecules. All characterized RBPs of lactococcal phages are modular in nature (219, 220). The N-terminal 'shoulder' domain of BppL connects the RBP to the upper baseplate protein BppU. A 'neck' domain links the shoulder to the 'head' domain (176, 220). The C-terminal head domain of lactococcal phage RBPs is responsible for recognizing saccharidic receptors on the bacterial surface (50, 213). The unique host ranges of Tuc2009 and TP901-1 could be experimentally swapped by exchanging $BppL_{901\text{-}1}$ and $BppL_{2009}$ (179). In addition

to BppL, it was recently demonstrated that the accessory baseplate protein BppA of phage Tuc2009 increases the host-affinity of the phage RBP tripods, facilitating efficient host recognition (183, 218).

Crystallization of the TP901-1 RBP, BppL, revealed that a glycerol molecule was co-crystallized at the presumed substrate binding site of BppL, while tryptophan fluorescence quenching experiments indicated that BppL also was able to bind to specific saccharides in solution with high affinity (213). The region of the TP901-1 RBP actively involved in host recognition was further characterized by specific designed ankyrin repeat protein (DARPin) binding-proteins and camelid antibodies which both bound to the tip of the RBP head domain. The interactions between camelid antibodies and DARPin molecules with lactococcal RBPs are specific and strong enough to prevent phage infection (214, 221, 222).

The topology of Tuc2009 and TP901-1 RBPs in the baseplate structure suggests they are in an infection-ready orientation to recognise their host bacteria. Unlike activation of phage p2 by calcium (or other divalent cations), which results in dramatic morphological alterations, calcium is not required for infection by TP901-1 and only partially required for infection by Tuc2009. Thus, the phage activation mechanism for these P335 species phages likely requires more subtle morphological rearrangements (215). Currently, the best candidate for generating the signal for phage genome ejection upon correct host recognition was the observation of subtle morphological alterations in the tripods of TP901-1. Electron microscopy of purified TP901-1 tripods showed two class-averages, with the distance between tripod legs differing by 20 Å in ‘packed’ and ‘open’ conformations (185).

The tail fibre of phage SPP1, encoded by gp21, is bi-functional. The N-terminus of gp21 functions as a tail tube cap capable of opening for DNA ejection, while the C-terminus acts as a RBP. A hinge region in gp21, located downstream of the N-terminal cap domain allows for a high degree of flexibility, measuring angles as great as 50^{o} (193). SPP1 reversibly adsorbs to teichoic acids on the surface of *B. subtilis* (223), yet gp21 recognizes and irreversibly binds to the exposed ecto-domain of the membrane protein YueB (224, 225). The C-terminus of gp21 is not predicted to contain a channel for DNA egression. Vinga and colleagues (2012) have indicated that following gp21 C-terminus interacting with YueB, the tail fibre's C-terminus is still present and not cleaved from the virion structure (226). Therefore, the gp21-YueB interaction may result in a conformational change in gp21 such that it does not obstruct DNA egression following the opening of the tail tube cap represented by the N-terminus of gp21. At the initiation of infection, the TMP of SPP1 is expelled from the tail tube volume, presumably through the opened tail tube cap, before the capsid-released DNA is directed along the tail tube into the host bacterium (193).

In several instances, phage structural proteins such as the major capsid protein or major tail protein are reported as incorporating domains to enhance host adsorption (227). The C-terminus of phage lambda MTP, gpV, has an Ig-like fold believed to facilitate adsorption to enterobacterial cell-surface carbohydrates (228). As mentioned, the tail tube of SPP1 is predominantly composed of gp17.1; however, a +1 frameshift can also result in a C-terminal extension, producing protein gp17.1*. The C-terminal extension of gp17.1* is proposed by Auzat and colleagues (2008) as facilitating the phage infection cycle through weak interactions with the cell surface of bacteria and contributing to the phage scanning for a specific receptor on a potential host (199). The tail tube of lactococcal phage p2 is composed of 31 MTP

rings and a ring formed by the tail terminator protein. The MTP of phage p2 also possesses a C-terminal domain described as ‘decorating’ the tail tube, which may similarly fulfil an adhesion-like role (229).

SECTION III.V Host Penetration and DNA Injection

Phage λ gpK, a protein required for initiator formation, possesses a putative PGH domain. However, the cell wall-hydrolysing role of gpK was initially indeterminate as detecting gpK in mature virions proved difficult (182). Recent work has shown that gpK possesses two domains, and N-terminal Mov34 domain and a C-terminal NlpC/P60 domain. The N-terminal Mov34 domain is essential in tail morphogenesis. Mov34 domains are often involved in proteolysis, and this domain may be involved in the proteolysis of gpH and tail activation. The C-terminal NlpC/P60 domain of gpK is a PGH essential for phage activity, as mutagenesis of the catalytic Cys121 of gpK was shown to result in inactive phage particles (230).

Protease sensitivity of gpH and gpJ changes during λ’s transition from reversible to irreversible adsorption, which is concomitant with the transition from type 1 to type 2 complexes, respectively. Protein gpJ is protease resistant in dispersed λ particles and type 1 complexes. However, gpJ becomes sensitive to proteolytic degradation following formation of type 2 complexes. It was proposed that a conformational change in gpJ tail fibre occurs, such that it does not enter the cell but is in an exposed configuration. Protein gpH in type 2 complexes becomes markedly more proteinase resistant. Uncharged amino acid residues at both ends of gpH, and the change to a protease resistant conformation, led to speculation that the

TMP gpH is involved in membrane penetration and anchoring, or in creating a transmembrane channel (231).

Phages penetrating their host can proceed via pore-forming or fusogenic proteins (29). As discussed above, the tip of enterobacterial phage T5 TMP protein, Pb2, possesses pore forming activity *in vitro* (102). This is due to two fusogenic alpha-helical transmembrane segments present in the area between the TMP region and its straight tail fibre region of phage T5 Pb2 protein (105). These regions allow T5 tail fibre to cross two lipid bilayers and inject DNA into proteoliposomes containing T5's protein receptor FhuA (104). Penetration and DNA injection by T5 is facilitated by Pb2 possessing a C-terminally located muramidase domain which degrades the host PG layer, enabling the TMP to traverse this obstacle so that it can form a channel in the bacterial cell envelope (105).

The tail fibres of phages Tuc2009 and TP901-1, in addition to the role of their N-termini in phage morphogenesis, possess a C-terminal cell wall-degrading or lytic domain, and were thus named Tal_{2009} and $Tal_{901\text{-}1}$, respectively (for Tail-Associated Lysin of Tuc2009 and TP901-1). The virion-associated lytic domain was predicted by Kenny *et al.* (2006) as an M23 peptidase, involved in degrading the lactococcal cell for initiation of phage infection (232). Endopeptidases of the M23 peptidase family were characterized in *Staphylococcus* as glycyl-glycine hydrolases, cleaving the cross-linkage of the bacterial cell wall (233). In *Bacillus* prophage SP-β the M23 peptidase possesses $_{DD}$-endopeptidase activity, degrading the peptide-bond between $_{D}$-Ala and diaminopimelic acid in the peptidoglycan cross-linkage (234). While the lytic activity of Tal_{2009} was not observed against staphylococcal cell walls,

it was proposed to degrade the cross-linkage of lactococcal cell walls like other members of this protein family (232).

An interesting observed phenomenon of Tal_{2009} and $Tal_{901\text{-}1}$ is the incorporation of two forms of their tail fibres in the mature virions, both the full length ~100 kDa Tal protein and a truncated derivative lacking the C-terminal lytic domain. The stoichiometry and mechanism by which the full length and truncated Tal proteins are incorporated is still unknown, but Kenny and colleagues identified the proteolysis site in Tal_{2009}. Since the amino acid region around the proteolytic site of Tal_{2009} and $Tal_{901\text{-}1}$ is glycine-rich (GGSSG*GG for Tuc2009 and GGNSG*GG for TP901-1; position of the proteolytic cleavage is marked by *), and several M23 peptidase enzymes cleave glycyl-glycine bonds, it was speculated that the Tal proteins may undergo self-mediated proteolysis (232). However, although the Tal proteins are proteolytically processed *in vitro*, the proteolysis of the Tal proteins is now known not to be autoproteolytic (see Chapter 2). The proteolysis of the Tal proteins of Tuc2009 and TP901-1 is now determined to be important in the efficient adsorption of the phages to their hosts and infection of cells with increased cell wall cross-linkage (235).

SECTION IV Concluding Remarks

Phages are remarkable biological entities which have been a part of many important milestones in microbiology and molecular biology. Phages continue to provide, and aid our understanding of, many complex molecular interactions. The structures of phage virions represent extremely efficient nano-machines, and due to the conserved nature of their component proteins, a 'blue skies science' approach to understanding lactococcal phages is applicable across multidisciplinary phage-related topics.

Chapter 2 of this thesis is dedicated to examining the lytic activities associated with phage virions and understanding their role during infection. This chapter demonstrates a remarkable strategy utilized by certain tailed phages to ensure successful infection under different circumstances.

Chapter 3 presented in this thesis is a broad assessment of the structural module of a *Siphoviridae* phage, and sought to specifically determine the role of proteins in the assembly and infection processes of phages. Through a variety of molecular techniques, it is apparent that phage virions have evolved from a conserved ancestral origin, as many phage proteins, despite divergent sequences, fulfil identical roles. Lactococcal *Siphoviridae* phages characterized here demonstrate how viruses are extremely efficient in a number of aspects, including the utilization of a compact genome and the ability to assemble into complex

quaternary structures both spontaneously and through specific chaperone interactions.

Chapter 4 of this thesis describes the investigation of the most characteristic feature, and a fundamental question, of *Siphoviridae* – why do they possess long non-contractile tails? Through specific deletions, regions of the phage tape measure protein were scrutinized to determine the role of this protein in *Siphoviridae* structure assembly and its function during the infection process.

SECTION V Bibliography

1. **Hankin EH.** 1896. The bacterial action of water of the Ganges and Jumna rivers on cholera microbes, p. 511-5517, Annales de I'Institut Pasteur.
2. **Sulakvelidze A, Alavidze Z, Morris JG, Jr.** 2001. Bacteriophage therapy. Antimicrob Agents Chemother **45:**649-659.
3. **Twort FW.** 1915. An investigation of the nature on the nature of ultra-microscopic viruses. Lancet **186:**1241-1243.
4. **d'Herelle F.** 1917. Sur un microbe invisible antagoniste des bacilles dysentériques. CR Acad. Sci. Paris **165:**373-375.
5. **Duckworth DH.** 1976. "Who discovered bacteriophage?". Bacteriol Rev **40:**793-802.
6. **Luria SE, Delbruck M.** 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics **28:**491-511.
7. **Hershey AD, Chase M.** 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol **36:**39-56.
8. **Zinder ND, Lederberg J.** 1952. Genetic exchange in *Salmonella*. J Bacteriol **64:**679-699.
9. **Volkin E, Astrachan L.** 1956. Phosphorus incorporation in *Escherichia coli* ribo-nucleic acid after infection with bacteriophage T2. Virology **2:**149-161.
10. **Kloc M, Foreman V, Reddy SA.** 2011. Binary function of mRNA. Biochimie **93:**1955-1961.
11. **Crick FH, Barnett L, Brenner S, Watts-Tobin RJ.** 1961. General nature of the genetic code for proteins. Nature **192:**1227-1232.

12. **Nirenberg M, Matthaei JH.** 1961. Dependence of Cell-Free Protein Synthesis in *Escherichia coli* Upon Naturally Occurring or Synthetic Polyribonucleotides. Proc Natl Acad Sci U S A **47:**1588-&.

13. **Jacob F, Monod J.** 1961. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol **3:**318-356.

14. **Ptashne M.** 1967. Isolation of the lambda phage repressor. Proc Natl Acad Sci U S A **57:**306-313.

15. **Campbell AM.** 1992. Chromosomal insertion sites for phages and plasmids. J Bacteriol **174:**7495-7499.

16. **Ptashne M.** 2011. Principles of a switch. Nat Chem Biol **7:**484-487.

17. **Dussoix D, Arber W.** 1962. Host specificity of DNA produced by *Escherichia coli*. II. Control over acceptance of DNA from infecting phage lambda. J Mol Biol **5:**37-49.

18. **Gellert M.** 1967. Formation of covalent circles of lambda DNA by *Escherichia coli* extracts. Proc Natl Acad Sci U S A **57:**148-155.

19. **Georgopoulos CP, Hendrix RW, Casjens SR, Kaiser AD.** 1973. Host participation in bacteriophage lambda head assembly. J Mol Biol **76:**45-60.

20. **Morange M.** 2005. What history tells us II. The discovery of chaperone function. J Biosci **30:**461-464.

21. **Maxam AM, Gilbert W.** 1977. A new method for sequencing DNA. Proc Natl Acad Sci U S A **74:**560-564.

22. **Sanger F, Nicklen S, Coulson AR.** 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A **74:**5463-5467.

23. **Ptashne M.** 2004. A Genetic Switch – Phage Lambda Revisited. 3rd. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

24. **Stiles ME.** 1996. Biopreservation by lactic acid bacteria. Antonie Van Leeuwenhoek **70:**331-345.

25. **Mc Grath S, van Sinderen D.** 2007. Bacteriophage: genetics and molecular biology. Caister Academic Press, Norfolk, UK.

26. **Dept. of Agriculture, Food and the Marine.** 2013. Annual Review & Outlook 2012/2013, p. 1-141. Dept. Agriculture, Food and the Marine (ed.).

27. **Lahtinen S, Salminen S, Von Wright A, Ouwehand AC.** 2011. Lactic acid bacteria: microbiological and functional aspects. CRC Press.

28. **Lebeer S, Vanderleyden J, De Keersmaecker SC.** 2010. Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. Nature Rev Microbiol **8:**171-184.

29. **Poranen MM, Daugelavicius R, Bamford DH.** 2002. Common principles in viral entry. Ann Rev Microbiol **56:**521-538.

30. **Weidel W, Pelzer H.** 1964. Bagshaped Macromolecules--a New Outlook on Bacterial Cell Walls. Adv Enzymol Relat Areas Mol Biol **26:**193-232.

31. **Schleifer KH, Kandler O.** 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev **36:**407-477.

32. **Vollmer W, Blanot D, de Pedro MA.** 2008. Peptidoglycan structure and architecture. FEMS Microbiol Rev **32:**149-167.

33. **Vollmer W, Joris B, Charlier P, Foster S.** 2008. Bacterial peptidoglycan (murein) hydrolases. FEMS Microbiol Rev **32:**259-286.

34. **Crow VL, Coolbear T, Gopal PK, Martley FG, McKay LL, Riepe H.** 1995. The role of autolysis of lactic acid bacteria in the ripening of cheese. Int Dairy J **5:**855-875.

35. **Lortal S, Chapot-Chartier MP.** 2005. Role, mechanisms and control of lactic acid bacteria lysis in cheese. Int Dairy J **15:**857-871.

36. **Vollmer W, Seligman SJ.** 2010. Architecture of peptidoglycan: more data and more models. Trends Microbiol **18:**59-66.

37. **Gan L, Chen S, Jensen GJ.** 2008. Molecular organization of Gram-negative peptidoglycan. Proc Natl Acad Sci U S A **105:**18953-18957.

38. **Hayhurst EJ, Kailas L, Hobbs JK, Foster SJ.** 2008. Cell wall peptidoglycan architecture in *Bacillus subtilis*. Proc Natl Acad Sci U S A **105:**14603-14608.

39. **Andre G, Kulakauskas S, Chapot-Chartier MP, Navet B, Deghorain M, Bernard E, Hols P, Dufrene YF.** 2010. Imaging the nanoscale organization of peptidoglycan in living *Lactococcus lactis* cells. Nat Commun **1:**27.

40. **Cross ML.** 2002. Microbes versus microbes: immune signals generated by probiotic lactobacilli and their role in protection against microbial pathogens. FEMS Immunol Med Mic **34:**245-253.

41. **Servin AL, Coconnier MH.** 2003. Adhesion of probiotic strains to the intestinal mucosa and interaction with pathogens. Best practice & research. Clinical Gastroenterology **17:**741-754.

42. **Tripathi P, Beaussart A, Andre G, Rolain T, Lebeer S, Vanderleyden J, Hols P, Dufrene YF.** 2012. Towards a nanoscale view of lactic acid bacteria. Micron **43:**1323-1330.

43. **Sengupta R, Altermann E, Anderson RC, McNabb WC, Moughan PJ, Roy NC.** 2013. The role of cell surface architecture of lactobacilli in host-microbe interactions in the gastrointestinal tract. Mediators Inflamm 2013:237921.

44. **Ventura M, Callegari ML, Morelli L.** 1999. Surface layer variations affecting phage adsorption on seven *Lactobacillus helveticus* strains. Annali di microbiologia ed enzimologia **49:**45-53.

45. **Monteville MR, Ardestani B, Geller BL.** 1994. Lactococcal bacteriophages require a host cell wall carbohydrate and a plasma membrane protein for adsorption and ejection of DNA. Appl Environ Microbiol **60:**3204-3211.

46. **Silhavy TJ, Kahne D, Walker S.** 2010. The bacterial cell envelope. Cold Spring Harbor Perspectives in Biology **2:**a000414.

47. **Douglas LJ, Wolin MJ.** 1971. Cell wall polymers and phage lysis of *Lactobacillus plantarum*. Biochemistry **10:**1551-1555.

48. **Raisanen L, Schubert K, Jaakonsaari T, Alatossava T.** 2004. Characterization of lipoteichoic acids as *Lactobacillus delbrueckii* phage receptor components. J Bacteriol **186:**5529-5532.

49. **Quiberoni A, Stiefel JI, Reinheimer JA.** 2000. Characterization of phage receptors in *Streptococcus thermophilus* using purified cell walls obtained by a simple protocol. J Appl Microbiol **89:**1059-1065.

50. **Chapot-Chartier MP, Vinogradov E, Sadovskaya I, Andre G, Mistou MY, Trieu-Cuot P, Furlan S, Bidnenko E, Courtin P, Pechoux C, Hols P, Dufrene YF, Kulakauskas S.** 2010. Cell surface of *Lactococcus lactis* is covered by a protective polysaccharide pellicle. J Biol Chem **285:**10464-10471.

51. **Mahony J, Kot W, Murphy J, Ainsworth S, Neve H, Hansen LH, Heller KJ, Sorensen SJ, Hammer K, Cambillau C, Vogensen FK, van Sinderen D.** 2013. Investigation of the relationship between lactococcal host cell wall

polysaccharide genotype and 936 phage receptor binding protein phylogeny. Appl Environ Microbiol **79:**4385-4392.

52. **Binetti AG, Quiberoni A, Reinheimer JA.** 2002. Phage adsorption to *Streptococcus thermophilus*. Influence of environmental factors and characterization of cell receptors. Food Res Int **35:**73-83.

53. **Breitbart M, Rohwer F.** 2005. Here a virus, there a virus, everywhere the same virus? Trends Microbiol **13:**278-284.

54. **Whitehead HR, Cox GA.** 1936. The Occurrence of Bacteriophage in Cultures of Lactic Streptococci: J Dairy Res **7**:55-62.

55. **González AR, García P, Raya RR.** 2010. Bacteriophages of lactic acid bacteria. Biotechnology of Lactic Acid Bacteria: Novel Applications**:**111.

56. **Garneau JE, Moineau S.** 2011. Bacteriophages of lactic acid bacteria and their impact on milk fermentations. Microb Cell Fact **10 Suppl 1:**S20.

57. **Mahony J, Murphy J, van Sinderen D.** 2012. Lactococcal 936-type phages and dairy fermentation problems: from detection to evolution and prevention. Front Microbiol **3:**335.

58. **O'Sullivan D, Coffey A, Fitzgerald GF, Hill C, Ross RP.** 1998. Design of a phage-insensitive lactococcal dairy starter via sequential transfer of naturally occurring conjugative plasmids. Appl Environ Microbiol **64:**4618-4622.

59. **Coffey A, Ross RP.** 2002. Bacteriophage-resistance systems in dairy starter strains: molecular analysis to application. Antonie Van Leeuwenhoek **82:**303-321.

60. **Murphy J, Mahony J, Bonestroo M, Nauta A, van Sinderen D.** 2013. Impact of thermal and biocidal treatments on lactococcal 936-type phages. Int Dairy J.

61. **Verreault D, Gendron L, Rousseau GM, Veillette M, Masse D, Lindsley WG, Moineau S, Duchaine C.** 2011. Detection of airborne lactococcal bacteriophages in cheese manufacturing plants. Appl Environ Microbiol **77:**491-497.

62. **Garbutt KC, Kraus J, Geller BL.** 1997. Bacteriophage Resistance in *Lactococcus lactis* Engineered by Replacement of a Gene for a Bacteriophage Receptor. J Dairy Sci **80:**1512-1519.

63. **Dupont K, Vogensen FK, Neve H, Bresciani J, Josephsen J.** 2004. Identification of the receptor-binding protein in 936-species lactococcal bacteriophages. Appl Environ Microbiol **70:**5818-5824.

64. **Mahony J, van Sinderen D.** 2012. Structural aspects of the interaction of dairy phages with their host bacteria. Viruses **4:**1410-1424.

65. **Forde A, Fitzgerald GF.** 1999. Bacteriophage defence systems in lactic acid bacteria. Antonie Van Leeuwenhoek **76:**89-113.

66. **Brussow H.** 2001. Phages of dairy bacteria. Ann Rev Microbiol **55:**283-303.

67. **Brussow H, Desiere F.** 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol Microbiol **39:**213-222.

68. **Kutter E, Sulakvelidze A.** 2004. Bacteriophages: biology and applications. CRC Press.

69. **Mahony J, Ainsworth S, Stockdale S, van Sinderen D.** 2012. Phages of lactic acid bacteria: the role of genetics in understanding phage-host interactions and their co-evolutionary processes. Virology **434:**143-150.

70. **Ackermann HW.** 2003. Bacteriophage observations and evolution. Res Microbiol **154:**245-251.

71. **Bradley DE.** 1967. Ultrastructure of bacteriophage and bacteriocins. Bacteriol Rev **31:**230-314.

72. **Maniloff J, Ackermann HW.** 1998. Taxonomy of bacterial viruses: establishment of tailed virus genera and the order *Caudovirales*. Arch Virol **143:**2051-2063.

73. **Ackermann HW.** 2007. 5500 Phages examined in the electron microscope. Arch Virol **152:**227-243.

74. **Ackermann HW.** 1998. Tailed bacteriophages: the order *Caudovirales*. Adv Virus Res **51:**135-201.

75. **Deveau H, Labrie SJ, Chopin MC, Moineau S.** 2006. Biodiversity and classification of lactococcal phages. Appl Environ Microbiol **72:**4338-4346.

76. **Rohwer F, Edwards R.** 2002. The Phage Proteomic Tree: a genome-based taxonomy for phage. J Bacteriol **184:**4529-4535.

77. **Moldovan R, Chapman-McQuiston E, Wu XL.** 2007. On kinetics of phage adsorption. Biophys J **93:**303-315.

78. **Sturino JM, Klaenhammer TR.** 2006. Engineered bacteriophage-defence systems in bioprocessing. Nature reviews. Microbiology **4:**395-404.

79. **Rao GR, Burma DP.** 1971. Purification and properties of phage P22-induced lysozyme. J Biol Chem **246:**6474-6479.

80. **Mindich L, Lehman J.** 1979. Cell wall lysin as a component of the bacteriophage phi6 virion. J Virol **30:**489-496.

81. **Kao SH, McClain WH.** 1980. Baseplate protein of bacteriophage T4 with both structural and lytic functions. J Virol **34:**95-103.

82. **Rydman PS, Bamford DH.** 2002. The lytic enzyme of bacteriophage PRD1 is associated with the viral membrane. J Bacteriol **184:**104-110.

83. **Rashel M, Uchiyama J, Takemura I, Hoshiba H, Ujihara T, Takatsuji H, Honke K, Matsuzaki S.** 2008. Tail-associated structural protein gp61 of *Staphylococcus aureus* phage phi MR11 has bifunctional lytic activity. FEMS Microbiol Lett **284:**9-16.

84. **Moak M, Molineux IJ.** 2004. Peptidoglycan hydrolytic activities associated with bacteriophage virions. Mol Microbiol **51:**1169-1183.

85. **Rydman PS, Bamford DH.** 2000. Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity. Mol Microbiol **37:**356-363.

86. **Molineux IJ.** 2001. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. Mol Microbiol **40:**1-8.

87. **Piuri M, Hatfull GF.** 2006. A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. Mol Microbiol **62:**1569-1585.

88. **Cuervo A, Carrascosa JL.** 2012. Bacteriophages: structure. eLS.

89. **Cotter P.** 2011. Microbiology: Molecular syringes scratch the surface. Nature **475:**301-303.

90. **Kanamaru S, Leiman PG, Kostyuchenko VA, Chipman PR, Mesyanzhinov VV, Arisaka F, Rossmann MG.** 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature **415:**553-557.

91. **Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC, van Raaij MJ.** 2010. Structure of the bacteriophage T4 long tail fiber receptor-binding tip. Proc Natl Acad Sci U S A **107:**20287-20292.

92. **Leiman PG, Arisaka F, van Raaij MJ, Kostyuchenko VA, Aksyuk AA, Kanamaru S, Rossmann MG.** 2010. Morphogenesis of the T4 tail and tail fibers. Virology J **7:**355.

93. **Yu F, Mizushima S.** 1982. Roles of lipopolysaccharide and outer membrane protein OmpC of *Escherichia coli* K-12 in the receptor function for bacteriophage T4. J Bacteriol **151:**718-722.

94. **Berlyn MKB.** 1998. Linkage map of *Escherichia coli* K-10: the traditional map. Microbiol Mol Biol Rev **62:**814-984.

95. **Rakhuba DV, Kolomiets EI, Dey ES, Novik GI.** 2010. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. Pol J Microbiol **59:**145-155.

96. **Wood WB, Henninger M.** 1969. Attachment of tail fibers in bacteriophage T4 assembly: some properties of the reaction *in vitro* and its genetic control. J Mol Biol **39:**603 - 618.

97. **Nishima W, Kanamaru S, Arisaka F, Kitao A.** 2011. Screw motion regulates multiple functions of T4 phage protein gene product 5 during cell puncturing. J Am Chem Soc **133:**13571-13576.

98. **Leiman PG, Shneider MM.** 2012. Contractile tail machines of bacteriophages. Adv Exp Med Biol **726:**93-114.

99. **Hu B, Margolin W, Molineux IJ, Liu J.** 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science **339:**576-579.

100. **Kemp P, Gupta M, Molineux IJ.** 2004. Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. Mol Microbiol **53:**1251-1265.

101. **Roessner CA, Ihler GM.** 1986. Formation of transmembrane channels in liposomes during injection of lambda DNA. J Biol Chem **261:**386-390.

102. **Feucht A, Schmid A, Benz R, Schwarz H, Heller KJ.** 1990. Pore formation associated with the tail-tip protein pb2 of bacteriophage T5. J Biol Chem **265:**18561-18567.

103. **Marinelli LJ.** 2008. Characterization of the role of the Rpf motif in mycobacteriophage tape measure proteins. Ph.D. Thesis. University of Pittsburgh.

104. **Bohm J, Lambert O, Frangakis AS, Letellier L, Baumeister W, Rigaud JL.** 2001. FhuA-mediated phage genome transfer into liposomes: a cryo-electron tomography study. Curr Biol **11:**1168-1175.

105. **Boulanger P, Jacquot P, Plancon L, Chami M, Engel A, Parquet C, Herbeuval C, Letellier L.** 2008. Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. J Biol Chem **283:**13556-13564.

106. **Gueguen E, Rousseau P, Duval-Valentin G, Chandler M.** 2005. The transpososome: control of transposition at the level of catalysis. Trends Microbiol **13:**543-549.

107. **Tremblay DM, Tegoni M, Spinelli S, Campanacci V, Blangy S, Huyghe C, Desmyter A, Labrie S, Moineau S, Cambillau C.** 2006. Receptor-binding protein of *Lactococcus lactis* phages: identification and characterization of the saccharide receptor-binding site. J Bacteriol **188:**2400-2410.

108. **Groth AC, Calos MP.** 2004. Phage integrases: biology and applications. J Mol Biol **335:**667-678.

109. **Friedman DI, Mozola CC, Beeri K, Ko CC, Reynolds JL.** 2011. Activation of a prophage-encoded tyrosine kinase by a heterologous infecting phage results in a self-inflicted abortive infection. Mol Microbiol **82:**567-577.

110. **Tinsley CR, Bille E, Nassif X.** 2006. Bacteriophages and pathogenicity: more than just providing a toxin? Microbes and infection / Institut Pasteur **8:**1365-1371.

111. **Galkin VE, Yu X, Bielnicki J, Ndjonka D, Bell CE, Egelman EH.** 2009. Cleavage of bacteriophage lambda cI repressor involves the RecA C-terminal domain. J Mol Biol **385:**779-787.

112. **Hendrix RW, Lawrence JG, Hatfull GF, Casjens S.** 2000. The origins and ongoing evolution of viruses. Trends Microbiol **8:**504-508.

113. **Friedman DI, Olson ER, Georgopoulos C, Tilly K, Herskowitz I, Banuett F.** 1984. Interactions of bacteriophage and host macromolecules in the growth of bacteriophage lambda. Microbiol Rev **48:**299-325.

114. **Weigle J.** 1966. Assembly of phage lambda *in vitro*. Proc Natl Acad Sci U S A **55:**1462-1466.

115. **Wang IN, Smith DL, Young R.** 2000. Holins: the protein clocks of bacteriophage infections. Ann Rev Microbiol **54:**799-825.

116. **Stern A, Sorek R.** 2011. The phage-host arms race: shaping the evolution of microbes. Bioessays **33:**43-51.

117. **Mc Grath S, Fitzgerald GF, van Sinderen D.** 2002. Identification and characterization of phage-resistance genes in temperate lactococcal bacteriophages. Mol Microbiol **43:**509-520.

118. **Mahony J, McGrath S, Fitzgerald GF, van Sinderen D.** 2008. Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. Appl Environ Microbiol **74:**6206-6215.

119. **Labrie SJ, Samson JE, Moineau S.** 2010. Bacteriophage resistance mechanisms. Nature Rev Microbiol **8:**317-327.

120. **Chopin MC, Chopin A, Bidnenko E.** 2005. Phage abortive infection in lactococci: variations on a theme. Curr Opin Microbiol **8:**473-479.

121. **Sanders ME, Klaenhammer TR.** 1983. Characterization of Phage-Sensitive Mutants from a Phage-Insensitive Strain of *Streptococcus lactis* - Evidence for a Plasmid Determinant That Prevents Phage Adsorption. Appl Environ Microbiol **46:**1125-1133.

122. **Sijtsma L, Sterkenburg A, Wouters JTM.** 1988. Properties of the Cell-Walls of *Lactococcus lactis* subsp. *cremoris* Sk110 and Sk112 and Their Relation to Bacteriophage Resistance. Appl Environ Microbiol **54:**2808-2811.

123. **Sijtsma L, Jansen N, Hazeleger WC, Wouters JT, Hellingwerf KJ.** 1990. Cell Surface Characteristics of Bacteriophage-Resistant *Lactococcus lactis* subsp. *cremoris* SK110 and Its Bacteriophage-Sensitive Variant SK112. Appl Environ Microbiol **56:**3230-3233.

124. **Lucey M, Daly C, Fitzgerald GF.** 1992. Cell-Surface Characteristics of *Lactococcus lactis* Harboring pCI528, a 46 Kb Plasmid Encoding Inhibition of Bacteriophage Adsorption. J Gen Microbiol **138:**2137-2143.

125. **Forde A, Fitzgerald GF.** 2003. Molecular organization of exopolysaccharide (EPS) encoding genes on the lactococcal bacteriophage adsorption blocking plasmid, pCI658. Plasmid **49:**130-142.

126. **Miklič A, Rogelj I.** 2007. Screening for natural defence mechanisms of *Lactococcus lactis* strains isolated from traditional starter cultures. Int J Food Sci Tech **42:**777-782.

127. **Sun X, Göhler A, Heller KJ, Neve H.** 2006. The *ltp* gene of temperate *Streptococcus thermophilus* phage TP-J34 confers superinfection exclusion to *Streptococcus thermophilus* and *Lactococcus lactis*. Virology **350:**146-157.

128. **Bebeacua C, Fajardo L, Carlos J, Blangy S, Spinelli S, Bollmann S, Neve H, Cambillau C, Heller KJ.** 2013. X-ray structure of a superinfection exclusion lipoprotein from phage TP-J34 and identification of the tape measure protein as its target. Mol Microbiol **89**:152-165.

129. **Raleigh EA, Brooks JE.** 1998. Restriction modification systems: where they are and what they do, p. 78-92, Bacterial Genomes. Springer.

130. **Allison GE, Klaenhammer TR.** 1998. Phage resistance mechanisms in lactic acid bacteria. Int Dairy J **8:**207-226.

131. **Krüger DH, Bickle TA.** 1983. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiol Rev **47:**345.

132. **Sharp PM.** 1986. Molecular evolution of bacteriophages: evidence of selection against the recognition sites of host restriction enzymes. Mol Biol Evol **3:**75-83.

133. **Villion M, Chopin MC, Deveau H, Ehrlich SD, Moineau S, Chopin A.** 2009. P087, a lactococcal phage with a morphogenesis module similar to an *Enterococcus faecalis* prophage. Virology **388:**49-56.

134. **Lu Z, Altermann E, Breidt F, Kozyavkin S.** 2010. Sequence analysis of *Leuconostoc mesenteroides* bacteriophage phi1-A4 isolated from an industrial vegetable fermentation. Appl Environ Microbiol **76:**1955-1966.

135. **Drake JW, Charlesworth B, Charlesworth D, Crow JF.** 1998. Rates of spontaneous mutation. Genetics **148:**1667-1686.

136. **Drake JW.** 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci U S A **88:**7160-7164.

137. **Geis A, El Demerdash HA, Heller KJ.** 2003. Sequence analysis and characterization of plasmids from *Streptococcus thermophilus*. Plasmid **50:**53-69.

138. **Smith RM, Diffin FM, Savery NJ, Josephsen J, Szczelkun MD.** 2009. DNA cleavage and methylation specificity of the single polypeptide restriction-modification enzyme LlaGI. Nucleic Acids Res **37:**7206-7218.

139. **Teresa Alegre M, Rodriguez MC, Mesas JM.** 2009. Characterization of pRS5: a theta-type plasmid found in a strain of *Pediococcus pentosaceus* isolated from wine that can be used to generate cloning vectors for lactic acid bacteria. Plasmid **61:**130-134.

140. **Fallico V, McAuliffe O, Fitzgerald GF, Ross RP.** 2011. Plasmids of raw milk cheese isolate *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis* DPC3901 suggest a plant-based origin for the strain. Appl Environ Microbiol **77:**6451-6462.

141. **Suarez V, Zago M, Giraffa G, Reinheimer J, Quiberoni A.** 2009. Evidence for the presence of restriction/modification systems in *Lactobacillus delbrueckii*. J Dairy Res **76:**433-440.

142. **Guinane CM, Kent RM, Norberg S, Hill C, Fitzgerald GF, Stanton C, Ross RP.** 2011. Host specific diversity in *Lactobacillus johnsonii* as evidenced by a major chromosomal inversion and phage resistance mechanisms. PloS One **6:**e18740.

143. **Lee SH, Jung JY, Jeon CO.** 2011. Complete genome sequence of *Leuconostoc kimchii* strain C2, isolated from Kimchi. J Bacteriol **193:**5548.

144. **Horvath P, Barrangou R.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327:**167-170.

145. **Deveau H, Garneau JE, Moineau S.** 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Ann Rev Microbiol **64:**475-493.

146. **Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S.** 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468:**67-71.

147. **Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C.** 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology **155:**733-740.

148. **Marraffini LA, Sontheimer EJ.** 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science **322:**1843-1845.

149. **Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP.** 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell **139:**945-956.

150. **Haft DH, Selengut J, Mongodin EF, Nelson KE.** 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. Plos Comput Biol **1:**e60.

151. **Terns MP, Terns RM.** 2011. CRISPR-based adaptive immune systems. Curr Opin Microbiol **14:**321-327.

152. **Sorek R, Kunin V, Hugenholtz P.** 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol **6:**181-186.

153. **Godde JS, Bickerton A.** 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J Mol Evol **62:**718-729.

154. **Portillo MC, Gonzalez JM.** 2009. CRISPR elements in the Thermococcales: evidence for associated horizontal gene transfer in *Pyrococcus furiosus*. J Appl Genet **50:**421-430.

155. **Tomita H, Ike Y.** 2005. Genetic analysis of transfer-related regions of the vancomycin resistance *Enterococcus* conjugative plasmid pHTbeta: identification of *oriT* and a putative relaxase gene. J Bacteriol **187:**7727-7737.

156. **Horvath P, Coute-Monvoisin AC, Romero DA, Boyaval P, Fremaux C, Barrangou R.** 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. Int J Food Microbiol **131:**62-70.

157. **Millen AM, Horvath P, Boyaval P, Romero DA.** 2012. Mobile CRISPR/Cas-mediated bacteriophage resistance in *Lactococcus lactis*. PloS One **7:**e51663.

158. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.

159. **Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R.** 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J Bacteriol **190:**1401-1412.

160. **van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ.** 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci **34:**401-407.

161. **Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S.** 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol **190:**1390-1400.

162. **Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V.** 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. Nucleic Acids Res **39:**9275-9282.

163. **Aranaz A, Romero B, Montero N, Alvarez J, Bezos J, de Juan L, Mateos A, Dominguez L.** 2004. Spoligotyping profile change caused by deletion of a direct variable repeat in a *Mycobacterium tuberculosis* isogenic laboratory strain. J Clin Microbiol **42:**5388-5391.

164. **Molineux IJ.** 1991. Host-parasite interactions: recent developments in the genetics of abortive phage infections. New Biol **3:**230-236.

165. **Uzan M, Miller E.** 2010. Post-transcriptional control by bacteriophage T4: mRNA decay and inhibition of translation initiation. Virology J **7:**360.

166. **Mills S, Ross RP, Neve H, Coffey A.** 2011. Bacteriophage and Anti-Phage Mechanisms in Lactic Acid Bacteria. Lactic Acid Bacteria: Microbiological and Functional Aspects**:**165.

167. **Drucker VV, Dutova NV.** 2006. Study of the morphological diversity of bacteriophages in Lake Baikal. Dokl Biol Sci **410:**421-423.

168. **Lin L, Hong W, Ji X, Han J, Huang L, Wei Y.** 2010. Isolation and characterization of an extremely long tail *Thermus* bacteriophage from Tengchong hot springs in China. J Basic Microbiol **50:**452-456.

169. **Katsura I, Hendrix RW.** 1984. Length determination in bacteriophage lambda tails. Cell **39:**691-698.

170. **Katsura I.** 1987. Determination of bacteriophage lambda tail length by a protein ruler. Nature **327:**73-75.

171. **Pedersen M, Ostergaard S, Bresciani J, Vogensen FK.** 2000. Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. Virology **276:**315-328.

172. **Belcaid M, Bergeron A, Poisson G.** 2011. The evolution of the tape measure protein: units, duplications and losses. BMC Bioinformatics **12 Suppl 9:**S10.

173. **Siponen M, Sciara G, Villion M, Spinelli S, Lichiere J, Cambillau C, Moineau S, Campanacci V.** 2009. Crystal structure of ORF12 from *Lactococcus lactis* phage p2 identifies a tape measure protein chaperone. J Bacteriol **191:**728-734.

174. **Katsura I.** 1987. Determination of bacteriophage λ tail length by a protein ruler. Nature **327:**73-75.

175. **Casjens SR, Hendrix RW.** 1974. Locations and amounts of major structural proteins in bacteriophage lambda. J Mol Biol **88:**535-545.

176. **Sciara G, Blangy S, Siponen M, Mc Grath S, van Sinderen D, Tegoni M, Cambillau C, Campanacci V.** 2008. A topological model of the baseplate of lactococcal phage Tuc2009. J Biol Chem **283:**2716-2723.

177. **Hendrix RW, Casjens SR.** 1974. Protein cleavage in bacteriophage lambda tail assembly. Virology **61:**156-159.

178. **Tsui LC, Hendrix RW.** 1983. Proteolytic processing of phage lambda tail protein gpH: timing of the cleavage. Virology **125:**257-264.

179. **Vegge CS, Vogensen FK, Mc Grath S, Neve H, van Sinderen D, Brondsted L.** 2006. Identification of the lower baseplate protein as the antireceptor of the temperate lactococcal bacteriophages TP901-1 and Tuc2009. J Bacteriol **188:**55-63.

180. **Katsura I, Kuhl PW.** 1975. Morphogenesis of the tail of bacteriophage lambda. III. Morphogenetic pathway. J Mol Biol **91:**257-273.

181. **Xu J.** 2001. A conserved frameshift strategy in dsDNA long tailed bacteriophages. Ph.D. Thesis. University of Pittsburgh.

182. **Dai X.** 2003. Expression, purification, and characterisation of bacteriophage lambda tail tip proteins. Ph.D. Thesis. Zhejiang University.

183. **Mc Grath S, Neve H, Seegers JF, Eijlander R, Vegge CS, Brondsted L, Heller KJ, Fitzgerald GF, Vogensen FK, van Sinderen D.** 2006. Anatomy of a lactococcal phage tail. J Bacteriol **188:**3972-3982.

184. **Vegge CS, Brondsted L, Neve H, Mc Grath S, van Sinderen D, Vogensen FK.** 2005. Structural characterization and assembly of the distal tail structure

of the temperate lactococcal bacteriophage TP901-1. J Bacteriol **187:**4187-4197.

185. **Bebeacua C, Bron P, Lai L, Vegge CS, Brondsted L, Spinelli S, Campanacci V, Veesler D, van Heel M, Cambillau C.** 2010. Structure and molecular assignment of lactococcal phage TP901-1 baseplate. J Biol Chem **285:**39079-39086.

186. **Campanacci V, Veesler D, Lichiere J, Blangy S, Sciara G, Moineau S, van Sinderen D, Bron P, Cambillau C.** 2010. Solution and electron microscopy characterization of lactococcal phage baseplates expressed in *Escherichia coli*. J Struct Biol **172:**75-84.

187. **Veesler D, Robin G, Lichiere J, Auzat I, Tavares P, Bron P, Campanacci V, Cambillau C.** 2010. Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): a baseplate hub paradigm in Gram-positive infecting phages. J Biol Chem **285:**36666-36673.

188. **Levin ME, Hendrix RW, Casjens SR.** 1993. A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. J Mol Biol **234:**124-139.

189. **Xu J, Hendrix RW, Duda RL.** 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol Cell **16:**11-21.

190. **Xu J, Hendrix RW, Duda RL.** 2013. A Balanced Ratio of Proteins from Gene G and Frameshift-Extended Gene GT Is Required for Phage Lambda Tail Assembly. J Mol Biol **425:**3476-3487.

191. **Xu J, Hendrix RW, Duda RL.** 2013. Chaperone-Protein Interactions That Mediate Assembly of the Bacteriophage Lambda Tail to the Correct Length. J Mol Biol **426**:1004-1018.

192. **Pell LG, Cumby N, Clark TE, Tuite A, Battaile KP, Edwards AM, Chirgadze NY, Davidson AR, Maxwell KL.** 2013. A conserved spiral structure for highly diverged phage tail assembly chaperones. J Mol Biol **425:**2436-2449.

193. **Plisson C, White HE, Auzat I, Zafarani A, Sao-Jose C, Lhuillier S, Tavares P, Orlova EV.** 2007. Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. EMBO J **26:**3720-3728.

194. **Edmonds L, Liu A, Kwan JJ, Avanessy A, Caracoglia M, Yang I, Maxwell KL, Rubenstein J, Davidson AR, Donaldson LW.** 2007. The NMR structure of the gpU tail-terminator protein from bacteriophage lambda: identification of sites contributing to Mg(II)-mediated oligomerization and biological function. J Mol Biol **365:**175-186.

195. **Chagot B, Auzat I, Gallopin M, Petitpas I, Gilquin B, Tavares P, Zinn-Justin S.** 2012. Solution structure of gp17 from the *Siphoviridae* bacteriophage SPP1: Insights into its role in virion assembly. Proteins **80:**319-326.

196. **Katsura I.** 1976. Morphogenesis of bacteriophage lambda tail. Polymorphism in the assembly of the major tail protein. J Mol Biol **107:**307-326.

197. **Katsura I, Tsugita A.** 1977. Purification and characterization of the major protein and the terminator protein of the bacteriophage lambda tail. Virology **76:**129-145.

198. **Effantin G, Boulanger P, Neumann E, Letellier L, Conway JF.** 2006. Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. J Mol Biol **361:**993-1002.

199. **Auzat I, Droge A, Weise F, Lurz R, Tavares P.** 2008. Origin and function of the two major tail proteins of bacteriophage SPP1. Mol Microbiol **70:**557-569.

200. **Bebeacua C, Lai L, Vegge CS, Brondsted L, van Heel M, Veesler D, Cambillau C.** 2013. Visualizing a complete *Siphoviridae* member by single-particle electron microscopy: the structure of lactococcal phage TP901-1. J Virol **87:**1061-1068.

201. **Randall-Hazelbauer L, Schwartz M.** 1973. Isolation of the bacteriophage lambda receptor from *Escherichia coli*. J Bacteriol **116:**1436-1446.

202. **Esquinas-Rychen M, Erni B.** 2001. Facilitation of bacteriophage lambda DNA injection by inner membrane proteins of the bacterial phosphoenol-pyruvate: carbohydrate phosphotransferase system (PTS). J Mol Microbiol Biotechnol **3:**361-370.

203. **Braun-Breton C, Hofnung M.** 1981. *In vivo* and *in vitro* functional alterations of the bacteriophage lambda receptor in *lamB* missense mutants of *Escherichia coli* K-12. J Bacteriol **148:**845-852.

204. **Roessner CA, Struck DK, Ihler GM.** 1983. Morphology of complexes formed between bacteriophage lambda and structures containing the lambda receptor. J Bacteriol **153:**1528-1534.

205. **Hendrix RW, Duda RL.** 1992. Bacteriophage lambda PaPa: not the mother of all lambda phages. Science **258:**1145-1148.

206. **Heller KJ, Schwarz H.** 1985. Irreversible binding to the receptor of bacteriophages T5 and BF23 does not occur with the tip of the tail. J Bacteriol **162:**621-625.

207. **Mondigler M, Vogele RT, Heller KJ.** 1995. Overproduced and purified receptor binding protein pb5 of bacteriophage T5 binds to the T5 receptor protein FhuA. FEMS Microbiol Lett **130:**293-300.

208. **Mondigler M, Holz T, Heller KJ.** 1996. Identification of the receptor-binding regions of pb5 proteins of bacteriophages T5 and BF23. Virology **219:**19-28.

209. **Plancon L, Janmot C, le Maire M, Desmadril M, Bonhivers M, Letellier L, Boulanger P.** 2002. Characterization of a high-affinity complex between the bacterial outer membrane protein FhuA and the phage T5 protein pb5. J Mol Biol **318:**557-569.

210. **Feucht A, Heinzelmann G, Heller KJ.** 1989. Irreversible binding of bacteriophage T5 to its FhuA receptor protein is associated with covalent cross-linking of 3 copies of tail protein pb4. FEBS Lett **255:**435-440.

211. **Guihard G, Boulanger P, Letellier L.** 1992. Involvement of phage T5 tail proteins and contact sites between the outer and inner membrane of *Escherichia coli* in phage T5 DNA injection. J Biol Chem **267:**3173-3178.

212. **Heller K, Braun V.** 1979. Accelerated adsorption of bacteriophage T5 to *Escherichia coli* F, resulting from reversible tail fiber-lipopolysaccharide binding. J Bacteriol **139:**32-38.

213. **Spinelli S, Campanacci V, Blangy S, Moineau S, Tegoni M, Cambillau C.** 2006. Modular structure of the receptor binding proteins of *Lactococcus lactis* phages. The RBP structure of the temperate phage TP901-1. J Biol Chem **281:**14256-14262.

214. **Desmyter A, Farenc C, Mahony J, Spinelli S, Bebeacua C, Blangy S, Veesler D, van Sinderen D, Cambillau C.** 2013. Viral infection modulation

and neutralization by camelid nanobodies. Proc Natl Acad Sci U S A **110:**E1371-1379.

215. **Veesler D, Spinelli S, Mahony J, Lichiere J, Blangy S, Bricogne G, Legrand P, Ortiz-Lombardia M, Campanacci V, van Sinderen D, Cambillau C.** 2012. Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. Proc Natl Acad Sci U S A **109:**8954-8958.

216. **Sciara G, Bebeacua C, Bron P, Tremblay D, Ortiz-Lombardia M, Lichiere J, van Heel M, Campanacci V, Moineau S, Cambillau C.** 2010. Structure of lactococcal phage p2 baseplate and its mechanism of activation. Proc Natl Acad Sci U S A **107:**6852-6857.

217. **Shepherd DA, Veesler D, Lichiere J, Ashcroft AE, Cambillau C.** 2011. Unraveling lactococcal phage baseplate assembly by mass spectrometry. Mol Cell Proteomics **10:**M111 009787.

218. **Collins B, Bebeacua C, Mahony J, Blangy S, Douillard FP, Veesler D, Cambillau C, van Sinderen D.** 2013. Structure and functional analysis of the host recognition device of lactococcal phage Tuc2009. J Virol **87:**8429-8440.

219. **Ricagno S, Campanacci V, Blangy S, Spinelli S, Tremblay D, Moineau S, Tegoni M, Cambillau C.** 2006. Crystal structure of the receptor-binding protein head domain from *Lactococcus lactis* phage bIL170. J Virol **80:**9331-9335.

220. **Siponen M, Spinelli S, Blangy S, Moineau S, Cambillau C, Campanacci V.** 2009. Crystal structure of a chimeric receptor binding protein constructed from two lactococcal phages. J Bacteriol **191:**3220-3225.

221. **De Haard HJW, Bezemer S, Ledeboer AM, Müller WH, Boender PJ, Moineau S, Coppelmans MC, Verkleij AJ, Frenken LGJ, Verrips CT.** 2005. Llama antibodies against a lactococcal protein located at the tip of the phage tail prevent phage infection. J Bacteriol **187:**4531-4541.

222. **Veesler D, Dreier B, Blangy S, Lichiere J, Tremblay D, Moineau S, Spinelli S, Tegoni M, Pluckthun A, Campanacci V, Cambillau C.** 2009. Crystal structure and function of a DARPin neutralizing inhibitor of lactococcal phage TP901-1: comparison of DARPin and camelid VHH binding mode. J Biol Chem **284:**30718-30726.

223. **Baptista C, Santos MA, Sao-Jose C.** 2008. Phage SPP1 reversible adsorption to *Bacillus subtilis* cell wall teichoic acids accelerates virus recognition of membrane receptor YueB. J Bacteriol **190:**4989-4996.

224. **Sao-Jose C, Baptista C, Santos MA.** 2004. *Bacillus subtilis* operon encoding a membrane receptor for bacteriophage SPP1. J Bacteriol **186:**8337-8346.

225. **Sao-Jose C, Lhuillier S, Lurz R, Melki R, Lepault J, Santos MA, Tavares P.** 2006. The ectodomain of the viral receptor YueB forms a fiber that triggers ejection of bacteriophage SPP1 DNA. J Biol Chem **281:**11464-11470.

226. **Vinga I, Baptista C, Auzat I, Petipas I, Lurz R, Tavares P, Santos MA, Sao-Jose C.** 2012. Role of bacteriophage SPP1 tail spike protein gp21 on host cell receptor binding and trigger of phage DNA ejection. Mol Microbiol **83:**289-303.

227. **Fraser JS, Maxwell KL, Davidson AR.** 2007. Immunoglobulin-like domains on bacteriophage: weapons of modest damage? Curr Opin Microbiol **10:**382-387.

228. **Pell LG, Gasmi-Seabrook GM, Morais M, Neudecker P, Kanelis V, Bona D, Donaldson LW, Edwards AM, Howell PL, Davidson AR, Maxwell KL.** 2010. The solution structure of the C-terminal Ig-like domain of the bacteriophage lambda tail tube protein. J Mol Biol **403:**468-479.

229. **Bebeacua C, Tremblay D, Farenc C, Chapot-Chartier MP, Sadovskaya I, van Heel M, Veesler D, Moineau S, Cambillau C.** 2013. Structure, Adsorption to Host, and Infection Mechanism of Virulent Lactococcal Phage p2. J Virol **87**:12302-12312.

230. **Coburn DL.** 2011. The Role of Bacteriophage Lambda gpK in Tail Assembly and Host Cell Entry. Ph.D. Thesis. University of Toronto.

231. **Roessner CA, Ihler GM.** 1984. Proteinase sensitivity of bacteriophage lambda tail proteins gpJ and pH in complexes with the lambda receptor. J Bacteriol **157:**165-170.

232. **Kenny JG, McGrath S, Fitzgerald GF, van Sinderen D.** 2004. Bacteriophage Tuc2009 encodes a tail-associated cell wall-degrading activity. J Bacteriol **186:**3480-3491.

233. **Odintsov SG, Sabala I, Marcyjaniak M, Bochtler M.** 2004. Latent LytM at 1.3 A resolution. J Mol Biol **335:**775-785.

234. **Sudiarta IP, Fukushima T, Sekiguchi J.** 2010. *Bacillus subtilis* CwlP of the SP-{beta} prophage has two novel peptidoglycan hydrolase domains, muramidase and cross-linkage digesting $_{DD}$-endopeptidase. J Biol Chem **285:**41232-41243.

235. **Stockdale SR, Mahony J, Courtin P, Chapot-Chartier MP, van Pijkeren JP, Britton RA, Neve H, Heller KJ, Aideh B, Vogensen FK, van Sinderen D.** 2013. The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for infection specialization. J Biol Chem **288:**5581-5590.

CHAPTER II

The Lactococcal Phages Tuc2009 and TP901-1 Incorporate Two Alternate Forms of their Tail Fibre into their Virions for Infection Specialization

Notes:

Purified NZ9000 peptidoglycan was supplied, and analysed by RP-HPLC and MALDI-TOF MS following enzymatic digestion, by Dr. Pascal Courtin and Dr. Marie-Pierre Chapot-Chartier.

This chapter has been peer reviewed and published as a research article:

Stockdale, S. R., Mahony, J., Courtin, P., Chapot-Chartier, M. P., van Pijkeren, J. P., Britton, R. A., Neve, H., Heller, K. J., Aideh, B., Vogensen, F. K., & van Sinderen, D. (2013). The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fibre into their virions for infection specialization. *Journal of Biological Chemistry*, **288** (8), 5581-5590.

ABSTRACT

Lactococcal phages Tuc2009 and TP901-1 possess a conserved tail fibre, called a Tail-Associated Lysin (referred to as Tal_{2009} for Tuc2009, and $Tal_{901\text{-}1}$ for TP901-1), suspended from their tail tips that projects a Peptidoglycan-Hydrolase (PGH) domain towards a potential host bacterium. Tal_{2009} and $Tal_{901\text{-}1}$ can undergo proteolytic processing mid-protein at the glycine-rich sequence GG(S/N)SGGG, removing their C-terminal structural lysin. In this study, we show that the PGH of these Tal proteins is an M23 peptidase which exhibits $_{D}$-Ala-$_{D}$-Asp endopeptidase activity, and that this activity is required for efficient infection of stationary phase cells. Interestingly, the observed proteolytic processing of Tal_{2009} and $Tal_{901\text{-}1}$ facilitates increased host adsorption efficiencies of the resulting phages. This represents, to the best of our knowledge, the first example of tail fibre proteolytic processing that results in a heterogeneous population of two phage types. Phages which possess a full length tail fibre, or a truncated derivative, are better adapted to efficiently infect cells with extensively cross-linked cell wall, or infect with increased host-adsorption efficiencies, respectively.

INTRODUCTION

It is predicted that 60% of the estimated 10^{31} phages in the biosphere belong to the *Siphoviridae* (1, 2). Yet despite their prevalence, the molecular details surrounding the initiation of infection by such Siphophages are not fully resolved. The *Siphoviridae* tail tip is a multi-proteinaceous structure responsible for host recognition and the initiation of phage infection. *Siphoviridae* often possess a central tail fibre suspending from their tail tip known to be essential for various aspects of phage infection, such as tail morphogenesis, bacterial adsorption, cell envelope penetration, and directing DNA ejection (3-11).

Coliphages lambda and T5, and *Bacillus* phage SPP1, are arguably the best studied models for understanding the various roles that central tail fibre proteins fulfil during *Siphoviridae* infection. Lambda possesses a central tail fibre, composed of three copies of gpJ protein, which forms a hub around which the tail organelle assembles (3). The C-terminus of gpJ functions as a receptor-binding protein, attaching to LamB on the surface of *E. coli* (5, 12). The tail-tube tips of reversibly bound lambda particles are approximately 17 nm from the cell surface, while the tail-tube tips are in direct contact with the bacterial envelope when this phage is irreversibly bound to its host (13). The transition from reversibly to irreversibly bound phage lambda coincides with conformational changes in gpJ (detected through proteinase-sensitivity) and the appearance of a transmembrane channel proposed to be formed by the Tape Measure Protein (TMP) gpH (10, 14).

The tail protein Pb2 of *Siphoviridae* phage T5 serves both as the TMP and straight tail fibre. Insertion of purified Pb2 into black lipid bilayer membranes results in the formation of pores of ca. 2 nm diameter (15). The C-terminus of Pb2 possesses

a PGH domain which promotes the penetration of the *E. coli* cell envelope (9). Located between the TMP and tail fibre region domains of Pb2 are two transmembrane domains. These have been shown to insert T5's tail fibre into a lipid layer containing T5's proteinaceous receptor FhuA. Insertion of Pb2 into FhuA-containing liposomes causes conformational changes in Pb2, shortening its length while simultaneously doubling its width from 2 to 4 nm (9, 17).

The central tail fibre of *Bacillus subtilis* phage SPP1 is composed of three copies of gp21 (11). The N-terminus of gp21 forms a cap at the distal end of the tail-tube, which has been shown to open at the onset of infection (18). The C-terminus of gp21 acts as a receptor binding protein attaching to the membrane-spanning protein YueB (19-21). Irreversible adsorption of SPP1 to YueB initiates infection, and is accompanied by conformational changes transmitted across the full length of the tail-tube to the portal protein by means of a cascading rotational-movement of the Siphophage's major tail protein (11).

Siphoviridae phages infecting Lactic Acid Bacteria (LAB), and in particular *Lactococcus lactis*, are the single largest cause of industrial milk fermentation problems and result in significant economic losses (22). The P335 phage species is one of the three major lactococcal-infecting phage species which cause regular problems for the dairy industry (the other species being c2 and 936) (23, 24). Scientific efforts to understand and control this particular problem in the dairy industry have allowed lactococcal phages to become one of the most intensely studied phage groups, and one of the model systems for understanding how *Siphoviridae* phages infect their host (25).

The tail fibre of lactococcal phages Tuc2009 and TP901-1 is composed of a trimer of the so-called Tal protein, which between them share 95% amino acid identity (26). Tal_{2009} and Tal_{901-1} are predicted to possess a C-terminally located M23 peptidase PGH domain, which distally protrudes from the large host-recognizing baseplate structure of each of these phages (27-29). The N-terminus of Tal_{2009} and that of Tal_{901-1} both share 26% amino acid identity with the N-terminus of SPP1 tail fibre protein gp21 (30). The N-terminal portion of Tal_{2009} and Tal_{901-1} is therefore expected to fulfil an analogous role at the initiation of infection, opening to facilitate ejection of the TMP and DNA. Located between the predicted C-terminal M23 peptidase domain and the N-terminal tail-tube cap of Tal_{2009} and Tal_{901-1} is a glycine-rich sequence, GG(S/N)SGG. Interestingly, GGSSG*GG has been shown to undergo proteolysis *in vitro* and *in vivo* (cleavage location indicated by *), which results in mature Tuc2009 virions that have incorporated both full-length and truncated Tal proteins (31).

Here we present the biochemical and mutational analysis of Tal to further understand the role of tail fibre proteins during infection. The importance of the cell wall-degrading activity and proteolytic processing of Tal_{2009} and Tal_{901-1} for host penetration and adsorption is discussed, and placed within the context of current knowledge on host penetration by *Siphoviridae* phages.

MATERIALS AND METHODS

Bioinformatic analysis. Relevant DNA sequences were downloaded from NCBI and manipulated using the DNAstar software package (DNASTAR Inc., Madison, WI). Protein domains were identified using EMBL-EBI's InterProScan (http://wwwdev.ebi.ac.uk/interpro/). Multiple sequence alignments were generated online using EMBL's ClustalW program (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The amino acid percentage identity between protein sequences was estimated by BLAST2p (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Bacterial strains, culture conditions, phage preparations and plasmids. Bacteria, phages and plasmids used in this study are listed in Table 1. *L. lactis* strains were grown at 30^{o}C in M17 (Oxoid) broth or agar supplemented with 0.5% (w/v) glucose (GM17). Erythromycin and chloramphenicol were used as necessary at concentrations of 5 and 10 µg/ml, respectively. *L. lactis* strain NZ9000 carrying the TP901-1*erm* prophage was isolated by simultaneously inoculating 2% (v/v) NZ9000 from a fresh overnight and 2% (v/v) TP901-1*erm* from a 10^{10} PEG-precipitated lysate, followed by 6 hours incubation, before selecting for erythromycin-resistant colonies. Large-scale phage preparations of TP901-1*erm* and mutant derivatives were performed as described previously (4) by inducing 1.6 L of the lysogenized NZ9000 host with mitomycin C. TBT buffer (100 mM NaCl, 50 mM $CaCl_2$, 10 mM $MgCl_2$, 20 mM Tris-HCl [pH 7.0]) was used as the phage resuspension buffer.

Recombineering and oligonucleotides. All oligonucleotides used in this study are shown in Supplementary Table 1. Recombineering was performed as previously described (32). However, specific modifications optimized for *L. lactis*, such as

phosphorothioate linkages and transforming 500 µg of oligos, were also employed (33). Recombineering oligonucleotides were ordered from IDT (Leuven, Belgium), while all other oligonucleotides were ordered from Eurofins MWG (Ebersberg, Germany).

Protein expression and purification. The terminal 501 bp of the 3'-end of tal_{2009}, was overproduced in *E. coli* using the pQE60 vector construct named $pTal_{2009}$-5ΔN as previously described (31). The 501 bp 3'-end of tal_{2009} was also used to fuse it in-frame with the DNA encoding two LysM Cell-wall Binding Domains (CBDs) of Tuc2009's endolysin by SOEing PCR (34, 35). The resulting amplicon, which specified a Tal_{2009}M23peptidase-LysM fusion protein, and designated TalM23-LysM, was cloned into pNZ8048 using standard techniques, to generate plasmid pTalM23-LysM. The TalM23-LysM protein was overproduced in NZ9000 using the NICE expression system (36), and required purification using a denaturing-renaturation procedure as described in the QIAexpressionist handbook (QIAGEN).

The terminal 501 bp regions of TP901-1's tail fibre, amplified using TP901-1*erm* and mutant TP901-1$erm_{His892Ala}$ lysates as DNA templates, were cloned into pQE60 as described for $pTal_{2009}$-5ΔN, and the corresponding proteins, named Tal_{901-1}-5ΔN and Tal_{901-1}-5Δ$N_{His892Ala}$, respectively, were overproduced and purified as described above.

Assay for lytic activity. L. lactis UC509.9 cells were autoclaved and the insoluble material was washed twice with water, once with 10 mM sodium phosphate buffer (pH 6.5), and finally resuspended in the latter buffer to produce a suspension with an OD_{600} of 0.450. All lysis experiments were performed in

triplicate at 30°C. While only water was added to the negative control, mutanolysin (Sigma), at a concentration of 0.1 U/ml, was used in all samples to hydrolyse the glycan strand linkages of the lactococcal cell wall. Tal_{2009}-5ΔN and TalM23-LysM were added at 1.0 and 1.7 µg/ml (equimolar concentrations), respectively, to achieve further degradation of the cell wall. This combined degradation of the cell wall results in lysis and in a reduction in OD_{600}. Cell wall hydrolysis using $Tal_{901\text{-}1}$-5ΔN and $Tal_{901\text{-}1}$-$5\Delta N_{His892Ala}$ was performed in a similar manner, except that autoclaved cells were prepared from the TP901-1 host *L. lactis* 3107. $Tal_{901\text{-}1}$-5ΔN and $Tal_{901\text{-}1}$-$5\Delta N_{His892Ala}$ were assayed at equal concentrations of 1.75 µg/ml.

Determination of hydrolytic specificity on pure peptidoglycan (PG). L. lactis MG1363 PG was prepared and hydrolysed by mutanolysin as outlined previously (37). The resulting soluble muropeptides were reduced with sodium borohydride. Following this pre-digestion step and prior to the addition of TalM23-LysM, the samples were adjusted to pH 6.5 by the addition of NaOH. TalM23-LysM-mediated digestion of MG1363 PG-derived muropeptides was carried out for 16 hours at 30 °C with gentle agitation, using 0.04 mg/ml of enzyme. Samples were then analysed by RP-HPLC with an Agilent UHPLC1290 system using ammonium phosphate buffer and methanol linear gradient as described previously (34). All the peaks of the mutanolysin digest and most of the peaks (peaks 4 to 12) of the mutanolysin plus TalM23-LysM digest were analysed without desalting by MALDI-TOF mass spectrometry (MS) using a Voyager-DE STR mass spectrometer (Applied Biosystems) as reported previously (37). Structural identification of the muropeptides was deduced from the previously published reference profile (37).

Phage assays. Phages used in this study were induced from their lysogenic host by the addition of 3 µg/ml mitomycin C (Sigma) to a liquid culture of the relevant strain when it had reached an OD_{600} of 0.1. Phage lysates were sterilized by filtration (0.45 µm pore filter) and stored at 4^{o}C. Phage titres were determined in triplicate by plaque assay employing three independent lysates using the standard double agar method (38) against the TP901-1 sensitive host *L. lactis* 3107. All GM17 media contained 5 g/l glycine and 10 mM $CaCl_2$.

Adsorption assays were performed against host 3107 at an OD_{600} of 0.3 with a phage-host Multiplicity of Infection (MOI) of 0.01. Phages were allowed to adsorb for 10 minutes at 30 ^{o}C in the presence of 10 mM $CaCl_2$. The adsorption assay control without bacterial cells (i.e. media only) allowed enumeration of the initial phage titre. The percentage of phage which adsorbed was calculated as [(initial phage titre-final phage titre)/initial phage titre]*100 (39).

To compare the ability of TP901-1 and its Tal protein mutant derivatives to infect exponential phase or stationary phase cultures, Centre Of Infection (COI) assays were performed. Phage infection of *L. lactis* 3107 was performed using an MOI of 0.01 at 30 ^{o}C in the presence of 10 mM $CaCl_2$. Phages were added to exponentially growing 3107 strain at OD_{600} ~0.3 (host 1), or to stationary phase cells of 3107 following culture dilution to OD_{600} ~0.3 using cell-free 'spent' GM17 medium (host 2). After 10 minutes, cells were washed twice in GM17 to remove free phage, and subsequent dilutions were plaque assayed against an uninfected 3107 culture to determine the COI values of host 1 and host 2, after which the Efficiency of COI (ECOI) formation was calculated as [($COI_{host\ 1}$)/($COI_{host\ 2}$)] (40).

Competition assays between $Tal_{901\text{-}1}$ mutants TP901-1*erm*$_{Gly>Arg}$, which possesses a full length tail fibre, and TP901-1*erm*$_{Gly603*}$, which has a truncated Tal protein, were performed to directly compare the biological significance of the tail fibre and associated lytic domain. An equal volume of freshly prepared phage lysates of TP901-1*erm*$_{Gly>Arg}$ and TP901-1*erm*$_{Gly603*}$ were mixed together, and the phage cocktail was diluted 1:10 into both an exponential and stationary phase culture of *L. lactis* 3107 as described above, except the phages were allowed only 5 minutes to establish infection.

Western blots. Phage proteins, from CsCl-purified phage particles, were separated on a pre-cast 12 % SDS-PAGE gel (Thermo Scientific) using a 0.1 M Tris-HCl, 0.1 M HEPES, 3.0 mM SDS (pH 8.0) running buffer. Proteins were transferred to a nitrocellulose membrane (Hybond-C Extra, Amersham) in a electrotransfer unit (BioRad) for 2 hours at 100 V, in a 10 % methanol, 10 mM CAPS (pH 11.0) transfer buffer, which was kept cool by placing the unit on ice. The two primary antibodies for Western blots, recognizing the N-terminus of the Tal protein or the baseplate protein BppU, had previously been described (41), and shown to interact with the relevant protein of both Tuc2009 and TP901-1 (4). Western blot secondary antibodies, IRDye 680RD goat anti-rabbit antibodies (Li-Cor Biosciences), and pre-stained protein molecular weight ladder (11-250 kDa; Li-Cor Biosciences), were detected in the 700 nm channel using an Odyssey CLx infrared imager (Li-Cor Biosciences).

RESULTS

Analyses of P335-species structural proteins to identify virion-associated PGH domains. Kenny and colleagues previously demonstrated that the P335-species phage Tuc2009 possesses a tail-associated lytic activity (31). Although analysis of the phage genomes of two c2-type and forty 936-type phages did not reveal PGH domains associated with tail proteins (data not shown). The genomes of the eleven publicly available P335-species phages (23) all contain a gene that encodes an easily recognizable (due to its large size) TMP, followed by two genes that specify the Distal Tail protein (Dit) and tail fibre protein (Fig. 1) (25). ClustalW alignments of the presumed TMP, Dit and tail fibre protein sequences resulted in four closely related P335-species phage with a lysozyme-like domain associated with their TMP, five closely related P335-species phages with M23 peptidase domains associated with their tail fibre proteins, and two P335-species phage with no recognizable PGH domain (Fig. 1).

Tal_{2009} PGH activity against lactococcal cells. The constructs Tal_{2009}-5ΔN and TalM23-LysM both generate proteins with the same the C-terminal M23 peptidase of Tal_{2009} tail fibre, however, the latter contains two additional LysM cell wall binding domains (Fig. 2A). In order to verify the cell wall-degrading activity of Tal_{2009}-5ΔN and TalM23-LysM proteins, the lytic activity of these purified recombinant enzymes was assayed on autoclaved bacterial cells as outlined in the Material and Methods. Neither the addition of water (negative control) nor the addition of mutanolysin at low concentrations (0.1 U/ml) resulted in significant lysis of autoclaved cells, as measured through OD_{600} reduction of the PG-containing suspension (Fig. 2B). In contrast, a combination of mutanolysin plus Tal_{2009}-5ΔN or

TalM23-LysM resulted in a clear reduction in OD_{600}. TalM23-LysM was significantly more effective at hydrolysing PG relative to Tal_{2009}-5ΔN over 20 hours (Fig. 2B), which is consistent with other reports where the addition of cell wall binding motifs were used to increase the lytic activity of PGH catalytic domains (42-44). TalM23-LysM was therefore used in subsequent analyses.

Characterization of Tal_{2009} PGH hydrolytic specificity. The PGH catalytic domain of Tal_{2009} is located at its C-terminal extremity, as previously reported and confirmed here, and contains an M23 peptidase domain (31). The M23 peptidase family includes endopeptidases that target the interpeptide bridge of PG. In order to confirm this prediction, it was necessary to identify the specific chemical bond targeted by the purified TalM23-LysM. PG was extracted from *L. lactis* MG1363 and first degraded by mutanolysin, a muramidase, as previously described by Courtin *et al.* (37). The resulting soluble muropeptides were further incubated with TalM23-LysM recombinant protein. Muropeptides obtained after mutanolysin digestion (Fig. 3A) and after mutanolysin plus TalM23-LysM digestion (Fig. 3B) were separated by RP-HPLC. Comparison of the two obtained profiles revealed that the higher molecular weight muropeptides obtained by mutanolysin digestion of PG, such as dimers (peaks 16-30), trimers (peaks 31-36) and tetramers (peaks 37-39), were degraded by TalM23-LysM to its constituent monomers (peaks 1-15; Supplementary Table 2). This pattern of degradation shows that the PGH catalytic domain of Tal_{2009} acts as an endopeptidase that hydrolyses the lactococcal peptidoglycan within the interpeptide bridge at $_{D}$-Ala-$_{D}$-Asp/Asn bond. Thus, like other M23 peptidases (45-47), Tal_{2009} is a $_{DD}$-endopeptidase able to hydrolyse PG cross-bridges.

Construction of NZ9000-TP901-1erm for Tal_{901-1} mutagenesis. In order to investigate the precise role of lactococcal Tal proteins during infection, mutagenesis was attempted on the gene encoding the Tal_{2009} protein. However, no mutagenesis system was found to be compatible with Tuc2009's host strain UC509.9 (results not shown). As Tal_{2009} and Tal_{901-1} share 95% amino acid identity, and since mutagenesis of the TP901-1 genome has previously been successfully employed (4, 48-50), mutagenesis was instead performed on the Tal_{901-1} genome. However, rather than using the rather laborious pGhost mutagenesis method (51), we decided to employ recombineering, a mutagenesis system that has recently been adopted for *L. lactis* (32, 33).

For the purpose of achieving TP901-1 mutagenesis by recombineering, we constructed a derivative of *L. lactis* strain NZ9000 that harboured the TP901-1*erm* prophage (this phage is a derivative of TP901-1 carrying an erythromycin marker; (50)), which was designated NZ9000-TP901-1*erm* (see Materials and Methods). The integration of TP901-1*erm*, via *attB*-*attP* recombination (52), was confirmed by (a) the lack of a PCR product with primers which flanked NZ9000s *attB* sequence, and (b) by generating PCR products from the primers flanking NZ9000s *attB* sequence to TP901-1 chromosomal sequences (results not shown). TP901-1*erm* phages induced from NZ9000 with mitomycin C were still capable of infecting *L. lactis* 3107 at levels of approximately 10^7 pfu/ml.

Employing the NZ9000-TP901-1*erm* strain, we introduced three different mutations in the tal_{901-1} gene, which resulted in three mutant TP901-1*erm* phages, designated TP901-1$erm_{Gly>Arg}$, TP901-1$erm_{His892Ala}$ and TP901-1$erm_{Gly603*}$, and discussed in detail below.

Characterization of $Tal_{901\text{-}1}$ mutants. Specific mutations were made in TP901-1*erm* Tal protein in order to investigate the role(s) of the various conserved features observed in Tal_{2009} and $Tal_{901\text{-}1}$ during infection. Mutant TP901-1$erm_{Gly>Arg}$ was based on previous work (31), which had shown that altering the DNA sequence of 3 glycine codons, corresponding to Tal_{2009}'s proteolytic processing site GGSSGGG, to arginine codons, GRSSRRG, prevented *in vitro* proteolysis. The desired mutation was introduced into the TP901-1*erm* genome, and one such mutant prophage was selected and its genome sequenced to verify the genotype of the generated mutant phage. Western blots detecting the N-terminus of $Tal_{901\text{-}1}$ protein verified that the mutated Tal protein of mutant TP901-1$erm_{Gly>Arg}$ was unable to undergo proteolysis as only a single band corresponding to the full length $Tal_{901\text{-}1}$ protein (102.1 kDa) was detected, in contrast to the Tal protein of TP901-1*erm* (Fig. 4).

Western blots using primary antibodies directed against the Upper Baseplate Protein, BppU, which immediately succeeds $tal_{901\text{-}1}$, was performed on all mutants to verify that there were no downstream effects caused by mutating $tal_{901\text{-}1}$ (results not shown).

The second mutation in $tal_{901\text{-}1}$, generating mutant prophage TP901-1$erm_{His892Ala}$, was designed to inactivate the M23 peptidase domain, as demonstrated for the *Bacillus subtilis* prophage SP-β (45). The histidine at position 892 (His892) of $Tal_{901\text{-}1}$, underlined in TP901-1*erm*'s M23 peptidase sequence P<u>H</u>LHF, was targeted for mutation to an alanine residue as this is the first histidine in the HxH motif characteristic of zinc metallopeptidases (53). In order to verify the M23 peptidase of mutant TP901-1$erm_{His892Ala}$ was inactive, recombinant proteins of

TP901-1's C-terminal M23 peptidase domain were created. Constructs $Tal_{901\text{-}1}$-5ΔN and $Tal_{901\text{-}1}$-5$\Delta N_{His892Ala}$, representing the same C-terminal regions of their respective tail fibres as construct Tal_{2009}-5ΔN, were assayed against autoclaved cells of *L. lactis* 3107 (Fig. 2C). As expected, recombinant enzyme $Tal_{901\text{-}1}$-5ΔN derived from TP901-1*erm* showed lytic activity, while $Tal_{901\text{-}1}$-5$\Delta N_{His892Ala}$, constructed from mutant TP901-1$erm_{His892Ala}$, showed no activity (Fig. 2D). Western blots determined that the Tal protein of mutant TP901-$1_{His892Ala}$ still undergoes proteolytic processing. This is evident as the full length and truncated $Tal_{901\text{-}1}$ proteins (102.1 and 67.4 kDa, respectively) are detected (Fig. 4). These *in vivo* results show that the M23 peptidase domain $Tal_{901\text{-}1}$ is not involved in the proteolytic separation of the Tal protein N- and C-terminus.

The third mutant TP901-1 derivative, TP901-1$erm_{Gly603*}$, was designed to introduce a stop codon into the coding sequence of $tal_{901\text{-}1}$ at the position where it is known to undergo proteolysis. This mutation would be expected to result in a homogeneous population of phages, the reverse of TP901-1$erm_{Gly>Arg}$, with all the $Tal_{901\text{-}1}$ proteins of the TP901-1*erm* phage population truncated at the sequence GGNSG*GG (position of stop codon highlighted by *). TP901-1$erm_{Gly603*}$ also allows a comparison with mutant TP901-1$erm_{His892Ala}$, whose Tal protein is also expected to be incapable of degrading peptidoglycan. Western blots verified the truncated TP901-1$erm_{Gly603*}$ protein (67.4 kDa), which runs slightly below its expected position on the SDS-PAGE gel (Fig. 4). As expected, no full length $Tal_{901\text{-}1}$ protein (102.1 kDa) was detected in the Western blots of TP901-1$erm_{Gly603*}$ in contrast TP901-1*erm*.

Efficiency of plaque formation of TP901-1erm and its Tal_{901-1} mutant derivatives. In order to determine the effect of the various mutations of tal_{901-1} on the efficiency of plaque formation of its corresponding TP901-1*erm*-derived phages, the ability of these mutant phages to infect lactococcal host 3107 was tested by plaque assay analysis. Plaque assays of TP901-1*erm* and its mutant derivatives were performed in triplicate on three independently produced lysates. The EOPs of TP901-1$erm_{Gly>Arg}$, TP901-1$erm_{His892Ala}$ and TP901-1$erm_{Gly603*}$, relative to TP901-1*erm* phage, were thus determined to be 0.408, 0.050 and 0.062, respectively (Fig. 5A). The plaque morphologies of the Tal_{901-1} mutant phages were also noticeably different from those produced by wild type TP901-1*erm*: TP901-1$erm_{Gly>Arg}$ produced plaques that were slightly smaller and hazier compared to TP901-1*erm*, while TP901-1$erm_{His892Ala}$ and TP901-1$erm_{Gly603*}$ both formed pinprick-sized plaques (results not shown).

Adsorption assays of TP901-1 and Tal_{901-1} mutant derivates. Due to the location of Tal_{901-1} in the TP901-1 virion, i.e. suspending directly below the host-recognizing baseplate organelle (26), we were curious about the effects of the Tal_{901-1} mutations on adsorption. We therefore performed adsorption assays on the TP901-1*erm* and its derived mutant phages, and expressed the obtained data as the percentage of phages which bound to the bacterial host cells in 10 minutes (Fig. 5B). The MOI for the adsorption assays, 0.01 phage per bacterial cell, was optimized based on conditions that yielded >95 % of unmutated TP901-1*erm* adsorbing to its host *L. lactis* 3107 within 10 minutes (data not shown). The obtained phage adsorption efficiencies for the various phages were as follows: TP901-1*erm*, 98.01 $^{+}$/- 2.04 %; TP901-1$erm_{Gly>Arg}$, 87.24 $^{+}$/- 1.36 %; TP901-1$erm_{His892Ala}$, 93.28 $^{+}$/- 2.74 %; and TP901-1$erm_{Gly603*}$, 99.56, $^{+}$/- 0.44 %. These results demonstrate that mutants

TP901-1*erm*$_{Gly>Arg}$ and TP901-1*erm*$_{His892Ala}$ significantly differ in their host adsorption efficiencies compared to TP901-1 (p-value < 0.05), while mutant TP901-1$_{Gly603*}$ does not (Fig. 5B).

Efficiency of Formation of Centres of Infection. The M23 peptidase domain associated with Tal$_{2009}$ and Tal$_{901-1}$ is expected to facilitate cell wall penetration and genome delivery, particularly during conditions where the cell wall is extensively cross-linked, as would be expected for stationary phase cells (8, 54). To compare the ability of TP901-1*erm* and mutant derivatives to infect stationary phase cells compared to exponentially growing cells, the phage ECOIs were calculated (see Materials and Methods). The ECOI for TP901-1*erm* was calculated as 0.41 $^{+}$/- 0.02, TP901-1*erm*$_{Gly>Arg}$ as 0.56 $^{+}$/- 0.05, TP901-1*erm*$_{His892Ala}$ as 0.22 $^{+}$/- 0.005 and TP901-1*erm*$_{Gly603*}$ as 0.27 $^{+}$/- 0.03. The difference in the mutant phage's abilities to infect stationary phase cells, relative to TP901-1*erm*, is considered to be statistically significant (p-value < 0.05; Fig. 5C).

Competition assays between TP901-1 mutants with and without tail fibre. In order to compare and contrast the biological significance of TP901-1 proteolytic processing of Tal$_{901-1}$, the TP901-1 mutants permanently possessing or lacking the tail fibre were mixed together and simultaneously allowed to infect. The number of COIs established could be calculated due to distinguishing plaque morphologies formed by these phages. TP901-1*erm*$_{Gly>Arg}$, which possesses the C-terminus of Tal$_{901-1}$ and thus the M23 peptidase domain, forms more COIs in stationary phase cells than TP901-1*erm*$_{Gly603*}$. In contrast, TP901-1*erm*$_{Gly603*}$, which lacks the C-terminus of the tail fibre and the PGH motif, is more efficient at establishing COIs in exponentially growing cells (Fig. 5D).

DISCUSSION

The proteolysis of Tuc2009 and TP901-1 Tal protein represents a molecular 'bet-hedging' mechanism to produce a naturally mixed population of phages, which are simultaneously adapted for efficient host adsorption and host penetration. Many phages have evolved sophisticated strategies to efficiently infect their hosts, however, to our knowledge, the proteolytic processing step of Tal_{2009} and $Tal_{901\text{-}1}$ represents a novel phage mechanism.

Not all phages appear to require PGH domains to facilitate host penetration and infection, as many phages lack any recognizable PGH domain associated with their structural proteins. However, there is growing evidence to support that virion-associated PGH domains facilitate infection under conditions where the cell wall is extensively cross-linked (for a recent review, see (55)). In the current study, we show that the M23 peptidase of Tal_{2009} possesses a $_{D}$-Ala-$_{D}$-Asp endopeptidase specificity. This tail-associated lytic activity facilitates phage infection by undoing the interpeptide cross-linkage of peptidoglycan glycan strands. As the majority of the sequenced P335-species lactococcal phages possess PGH domains associated with their tail proteins, the localized degradation of the bacterial cell wall is predicted to be an important step in the initial stages of infection by this species of phage.

In this study, we clearly show that TP901-1*erm*-derived mutants defective in their tail-associated PGH activity are not completely deficient in phage infection, but display decreased infection efficiencies. This is consistent with the study of Moak and colleagues, who demonstrated that mutagenesis of T7's virion-associated PGH domain did not inhibit, but merely delayed, infection (56). Furthermore, the reduced ability of TP901-1*erm* PGH mutants to infect stationary phase cells is in agreement

with the study of Puiri *et al.*, where the tail fibre lytic domain of *Mycobacterium* phage TM4 was necessary for efficient infection of stationary phase cells (8).

It was previously reported by Scholl and colleagues that phage K1-5 encodes and incorporates two different tail fibre proteins into its virion, adapting the phage to penetrating hosts with a different cell envelope composition (57). Phages Tuc2009 and TP901-1 (and perhaps many other phages) have similarly evolved a novel strategy to generate a heterogeneous population. Through Tal proteolysis, the lactococcal phages Tuc2009 and TP901-1 generate two phage types; incorporating either a truncated Tal protein or incorporating a full length tail fibre with an associated lytic domain. Phages which feature in their virion a PGH domain are better adapted to infecting bacteria with a higher degree of cross-linkage in their cell wall. However, this evolutionary advantage appears to be a double-edged sword, as the fitness cost is at the expense of phage adsorption.

The study of lactococcal phages pertains not only to their industrial significance, but their broader importance as model systems. The clear differences in the tail fibres and infection mechanisms of lactococcal phages show the multifaceted approach by phages to achieve infection. Studies characterizing these underlying molecular processes serve to better understand *Siphoviridae*, and their host interactions, in the biosphere.

REFERENCES

1. **Ackermann HW.** 2007. 5500 Phages examined in the electron microscope. Arch Virol **152:**227-243.
2. **Wommack KE, Colwell RR.** 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev **64:**69-114.
3. **Katsura I, Kuhl PW.** 1975. Morphogenesis of the tail of bacteriophage lambda. III. Morphogenetic pathway. J Mol Biol **91:**257-273.
4. **Vegge CS, Brondsted L, Neve H, Mc Grath S, van Sinderen D, Vogensen FK.** 2005. Structural characterization and assembly of the distal tail structure of the temperate lactococcal bacteriophage TP901-1. J Bacteriol **187:**4187-4197.
5. **Berkane E, Orlik F, Stegmeier JF, Charbit A, Winterhalter M, Benz R.** 2006. Interaction of bacteriophage lambda with its cell surface receptor: an *in vitro* study of binding of the viral tail protein gpJ to LamB (Maltoporin). Biochemistry **45:**2708-2720.
6. **Duplessis M, Moineau S.** 2001. Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. Mol Microbiol **41:**325-336.
7. **Jakutyte L, Lurz R, Baptista C, Carballido-Lopez R, Sao-Jose C, Tavares P, Daugelavicius R.** 2012. First steps of bacteriophage SPP1 entry into *Bacillus subtilis*. Virology **422:**425-434.
8. **Piuri M, Hatfull GF.** 2006. A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. Mol Microbiol **62:**1569-1585.

9. **Boulanger P, Jacquot P, Plancon L, Chami M, Engel A, Parquet C, Herbeuval C, Letellier L.** 2008. Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. J Biol Chem **283:**13556-13564.

10. **Roessner CA, Ihler GM.** 1986. Formation of transmembrane channels in liposomes during injection of lambda DNA. J Biol Chem **261:**386-390.

11. **Plisson C, White HE, Auzat I, Zafarani A, Sao-Jose C, Lhuillier S, Tavares P, Orlova EV.** 2007. Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. EMBO J **26:**3720-3728.

12. **Gurnev PA, Oppenheim AB, Winterhalter M, Bezrukov SM.** 2006. Docking of a single phage lambda to its membrane receptor maltoporin as a time-resolved event. J Mol Biol **359:**1447-1455.

13. **Roessner CA, Struck DK, Ihler GM.** 1983. Morphology of complexes formed between bacteriophage lambda and structures containing the lambda receptor. J Bacteriol **153:**1528-1534.

14. **Roessner CA, Ihler GM.** 1984. Proteinase sensitivity of bacteriophage lambda tail proteins gpJ and pH in complexes with the lambda receptor. J Bacteriol **157:**165-170.

15. **Feucht A, Schmid A, Benz R, Schwarz H, Heller KJ.** 1990. Pore formation associated with the tail-tip protein pb2 of bacteriophage T5. J Biol Chem **265:**18561-18567.

16. **Boulanger P, Jacquot P, Plancon L, Chami M, Engel A, Parquet C, Herbeuval C, Letellier L.** 2008. Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. J Biol Chem **283:**13556-13564.

17. **Bohm J, Lambert O, Frangakis AS, Letellier L, Baumeister W, Rigaud JL.** 2001. FhuA-mediated phage genome transfer into liposomes: a cryo-electron tomography study. Curr Biol **11:**1168-1175.

18. **Goulet A, Lai-Kee-Him J, Veesler D, Auzat I, Robin G, Shepherd DA, Ashcroft AE, Richard E, Lichiere J, Tavares P, Cambillau C, Bron P.** 2011. The opening of the SPP1 bacteriophage tail, a prevalent mechanism in Gram-positive-infecting siphophages. J Biol Chem **286:**25397-25405.

19. **Sao-Jose C, Baptista C, Santos MA.** 2004. *Bacillus subtilis* operon encoding a membrane receptor for bacteriophage SPP1. J Bacteriol **186:**8337-8346.

20. **Sao-Jose C, Lhuillier S, Lurz R, Melki R, Lepault J, Santos MA, Tavares P.** 2006. The ectodomain of the viral receptor YueB forms a fiber that triggers ejection of bacteriophage SPP1 DNA. J Biol Chem **281:**11464-11470.

21. **Baptista C, Santos MA, Sao-Jose C.** 2008. Phage SPP1 reversible adsorption to *Bacillus subtilis* cell wall teichoic acids accelerates virus recognition of membrane receptor YueB. J Bacteriol **190:**4989-4996.

22. **Garneau JE, Moineau S.** 2011. Bacteriophages of lactic acid bacteria and their impact on milk fermentations. Microbial Cell Factories **10 Suppl 1:**S20.

23. **Deveau H, Labrie SJ, Chopin MC, Moineau S.** 2006. Biodiversity and classification of lactococcal phages. Appl Environ Microbiol **72:**4338-4346.

24. **Mahony J, Murphy J, van Sinderen D.** 2012. Lactococcal 936-type phages and dairy fermentation problems: from detection to evolution and prevention. Front Microbiol **3:**335.

25. **Veesler D, Cambillau C.** 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol Mol Biol Rev **75:**423-433.

26. **Bebeacua C, Bron P, Lai L, Vegge CS, Brondsted L, Spinelli S, Campanacci V, Veesler D, van Heel M, Cambillau C.** 2010. Structure and molecular assignment of lactococcal phage TP901-1 baseplate. J Biol Chem **285:**39079-39086.

27. **Johnsen MG, Neve H, Vogensen FK, Hammer K.** 1995. Virion positions and relationships of lactococcal temperate bacteriophage TP901-1 proteins. Virology **212:**595-606.

28. **Campanacci V, Veesler D, Lichiere J, Blangy S, Sciara G, Moineau S, van Sinderen D, Bron P, Cambillau C.** 2010. Solution and electron microscopy characterization of lactococcal phage baseplates expressed in *Escherichia coli*. J Struct Biol **172:**75-84.

29. **Bebeacua C, Lai L, Vegge CS, Brondsted L, van Heel M, Veesler D, Cambillau C.** 2012. Visualizing a Complete *Siphoviridae* by Single-particle Electron Microscopy: The Structure of Lactococcal Phage TP901-1. J Virol **87**:1061-1068.

30. **Veesler D, Robin G, Lichiere J, Auzat I, Tavares P, Bron P, Campanacci V, Cambillau C.** 2010. Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): a baseplate hub paradigm in Gram-positive infecting phages. J Biol Chem **285:**36666-36673.

31. **Kenny JG, McGrath S, Fitzgerald GF, van Sinderen D.** 2004. Bacteriophage Tuc2009 encodes a tail-associated cell wall-degrading activity. J Bacteriol **186:**3480-3491.

32. **van Pijkeren JP, Britton RA.** 2012. High efficiency recombineering in lactic acid bacteria. Nucleic Acids Res **40:**e76.

33. **van Pijkeren JP, Neoh KM, Sirias D, Findley AS, Britton RA.** 2012. Exploring optimization parameters to increase ssDNA recombineering in *Lactococcus lactis* and *Lactobacillus reuteri*. Bioengineered **3**.

34. **Seegers JFML, Mc Grath S, O'Connell-Motherway M, Arendt EK, van de Guchte M, Creaven M, Fitzgerald GF, van Sinderen D.** 2004. Molecular and transcriptional analysis of the temperate lactococcal bacteriophage Tuc2009. Virology **329:**40-52.

35. **Arendt EK, Daly C, Fitzgerald GF, Vandeguchte M.** 1994. Molecular Characterization of Lactococcal Bacteriophage-Tuc2009 and Identification and Analysis of Genes Encoding Lysin, a Putative Holin, and Two Structural Proteins. Appl Environ Microbiol **60:**1875-1883.

36. **Kuipers OP, de Ruyter PGGA, Kleerebezem M, de Vos WM.** 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. J Biotechnol **64**:15-21.

37. **Courtin P, Miranda G, Guillot A, Wessner F, Mezange C, Domakova E, Kulakauskas S, Chapot-Chartier MP.** 2006. Peptidoglycan structure analysis of *Lactococcus lactis* reveals the presence of an $_{L,D}$-carboxypeptidase involved in peptidoglycan maturation. J Bacteriol **188:**5293-5298.

38. **Lillehaug D.** 1997. An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. J Appl Microbiol **83:**85-90.

39. **Sanders ME, Klaenhammer TR.** 1980. Restriction and modification in group N streptococci: effect of heat on development of modified lytic bacteriophage. Appl Environ Microbiol **40:**500-506.

40. **Moineau S, Durmaz E, Pandian S, Klaenhammer TR.** 1993. Differentiation of Two Abortive Mechanisms by Using Monoclonal-Antibodies Directed toward Lactococcal Bacteriophage Capsid Proteins. Appl Environ Microbiol **59:**208-212.

41. **Mc Grath S, Neve H, Seegers JFML, Eijlander R, Vegge CS, Brondsted L, Heller KJ, Fitzgerald GF, Vogensen FK, van Sinderen D.** 2006. Anatomy of a lactococcal phage tail. J Bacteriol **188:**3972-3982.

42. **Donovan DM, Dong S, Garrett W, Rousseau GM, Moineau S, Pritchard DG.** 2006. Peptidoglycan hydrolase fusions maintain their parental specificities. Appl Environ Microbiol **72:**2988-2996.

43. **Loessner MJ, Kramer K, Ebel F, Scherer S.** 2002. C-terminal domains of *Listeria monocytogenes* bacteriophage murein hydrolases determine specific recognition and high-affinity binding to bacterial cell wall carbohydrates. Mol Microbiol **44:**335-349.

44. **Sass P, Bierbaum G.** 2007. Lytic activity of recombinant bacteriophage phi11 and phi12 endolysins on whole cells and biofilms of *Staphylococcus aureus*. Appl Environ Microbiol **73:**347-352.

45. **Sudiarta IP, Fukushima T, Sekiguchi J.** 2010. *Bacillus subtilis* CwlP of the SP-beta Prophage Has Two Novel Peptidoglycan Hydrolase Domains, Muramidase and Cross-linkage Digesting $_{DD}$-Endopeptidase. J Biol Chem **285:**41232-41243.

46. **Odintsov SG, Sabala I, Marcyjaniak M, Bochtler M.** 2004. Latent LytM at 1.3 A resolution. J Mol Biol **335:**775-785.

47. **Spencer J, Murphy LM, Conners R, Sessions RB, Gamblin SJ.** 2010. Crystal Structure of the LasA Virulence Factor from *Pseudomonas aeruginosa*: Substrate Specificity and Mechanism of M23 Metallopeptidases. J Mol Biol **396:**908-923.

48. **Pedersen M, Ostergaard S, Bresciani J, Vogensen FK.** 2000. Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. Virology **276:**315-328.

49. **Vegge CS, Neve H, Brondsted L, Heller KJ, Vogensen FK.** 2006. Analysis of the collar-whisker structure of temperate lactococcal bacteriophage TP901-1. Appl Environ Microbiol **72:**6815-6818.

50. **Vegge CS, Vogensen FK, Mc Grath S, Neve H, van Sinderen D, Brondsted L.** 2006. Identification of the lower baseplate protein as the antireceptor of the temperate lactococcal bacteriophages TP901-1 and Tuc2009. J Bacteriol **188:**55-63.

51. **Law J, Buist G, Haandrikman A, Kok J, Venema G, Leenhouts K.** 1995. A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. J Bacteriol **177:**7011-7018.

52. **Christiansen B, Johnsen MG, Stenby E, Vogensen FK, Hammer K.** 1994. Characterization of the lactococcal temperate phage TP901-1 and its site-specific integration. J Bacteriol **176:**1069-1076.

53. **Hooper NM.** 1994. Families of zinc metalloproteases. FEBS Lett **354:**1-6.

54. **Vollmer W, Blanot D, de Pedro MA.** 2008. Peptidoglycan structure and architecture. FEMS Microbiol Rev **32:**149-167.

55. **Rodriguez-Rubio L, Martinez B, Donovan DM, Rodriguez A, Garcia P.** 2012. Bacteriophage virion-associated peptidoglycan hydrolases: potential new enzybiotics. Crit Rev Microbiol **39**:427-434.

56. **Moak M, Molineux IJ.** 2000. Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. Mol Microbiol **37:**345-355.

57. **Scholl D, Rogers S, Adhya S, Merril CR.** 2001. Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of *Escherichia coli*. J Virol **75:**2509-2515.

58. **Wegmann U, O'Connell-Motherway M, Zomer A, Buist G, Shearman C, Canchaya C, Ventura M, Goesmann A, Gasson MJ, Kuipers OP, van Sinderen D, Kok J.** 2007. Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. J Bacteriol **189:**3256-3270.

59. **Koch B, Christiansen B, Evison T, Vogensen FK, Hammer K.** 1997. Construction of specific erythromycin resistance mutations in the temperate lactococcal bacteriophage TP901-1 and their use in studies of phage biology. Appl Environ Microbiol **63:**2439-2441.

60. **Braun V, Hertwig S, Neve H, Geis A, Teuber T.** 1989. Taxonomic Differentiation of Bacteriophages of *Lactococcus lactis* by Electron Microscopy, DNA-DNA Hybridization, and Protein Profiles. Microbiology **135:**2551-2560.

61. **Arendt EK, Daly C, Fitzgerald GF, van de Guchte M.** 1994. Molecular characterization of lactococcal bacteriophage Tuc2009 and identification and analysis of genes encoding lysin, a putative holin, and two structural proteins. Appl Environ Microb **60:**1875-1883.

62. **Costello VA.** 1988. Characterization of bacteriophage-host interactions in *Streptococcus cremoris* UC503 and related lactic streptococci. Ph.D. Thesis. University College Cork.

63. **Kuipers OP, Beerthuyzen MM, Siezen RJ, De Vos WM.** 1993. Characterization of the nisin gene cluster nisABTCIPR of *Lactococcus lactis*. Requirement of expression of the *nisA* and *nisI* genes for development of immunity. Eur J Biochem **216:**281-291.

64. **de Ruyter PG, Kuipers OP, de Vos WM.** 1996. Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. Appl Environ Microbiol **62:**3662-3667.

65. **Kleerebezem M, Beerthuyzen MM, Vaughan EE, de Vos WM, Kuipers OP.** 1997. Controlled gene expression systems for lactic acid bacteria: transferable nisin-inducible expression cassettes for *Lactococcus*, *Leuconostoc*, and *Lactobacillus* spp. Appl Environ Microbiol **63:**4581-4584.

Table 1. Bacteria, phages and plasmids used in this study.

Strains, phages and plasmids	Relevant features	Source
L. lactis		
MG1363	Prototypical *Lactococcus lactis* strain	(58)
NZ9000	MG1363 *pepN::nisRK*	(36)
901-1*erm*	Lysogenic host and source of TP901-1*erm* prophage	(59, 60)
3107	Lytic host for phage TP901-1*erm*	(52)
UC509	Lysogenic host and source of Tuc2009 prophage	(61)
UC509.9	Lytic host for phage Tuc2009	(62)
NZ9700	Nisin over-producing strain	(63)
NZ9000-TP901-1*erm*	NZ9000 lysogenized with TP901-1*erm* prophage	This study
NZ9000-TP901-1*erm*$_{Gly>Arg}$	Glycine to arginine mutations in the proteolysis site of TP901-1*erm* prophage Tal$_{901-1}$ protein	This study
NZ9000-TP901-1*erm*$_{His892Ala}$	Histidine to alanine mutation at amino acid position 892 of TP901-1*erm* prophage Tal$_{901-1}$ protein	This study
NZ9000-TP901-1*erm*$_{Gly603*}$	Stop codon inserted in TP901-1 prophage Tal$_{901-1}$ proteolysis site sequence mimicking natural truncation	This study
Phage		
TP901-1*erm*	Temperate P335-species phage infecting 3107; contains erythromycin marker	(59, 60)
Tuc2009	Temperate P335-species phage infecting UC509.9	(62)
TP901-1*erm*$_{Gly>Arg}$	TP901-1 phage with glycine to arginine mutations in the proteolysis site of Tal$_{901-1}$ protein	This study
TP901-1*erm*$_{His892Ala}$	TP901-1 phage with Tal$_{901-1}$ histidine at position 892 mutated to alanine	This study
TP901-1*erm*$_{Gly603*}$	TP901-1 phage with stop codon inserted at the site of Tal$_{901-1}$ proteolysis	This study
Plasmid		
pNZ8048	Nisin-inducible protein expression vector	(64)
pQE60	QIAGEN expression vector, N-terminal 6xHis tag	QIAGEN
pREP4	QIAGEN vector supplying LacI	QIAGEN
pJP005	pNZ8048 derivative expressing RecT protein	(32)
pNZ9530	Nisin helper plasmid supplying NisRK genes for protein expression; erythromycin marker	(65)
pTalM23-LysM	pNZ8048 construct expressing TalM23-LysM recombinant protein	This study
pTal$_{2009}$-5ΔN	pQE60 construct expressing terminal 501 bp of Tal$_{2009}$	(31)
pTal$_{901-1}$-5ΔN	pQE60 construct expressing terminal 501 bp of Tal$_{901-1}$	This study
pTal$_{901-1}$-5ΔN$_{His892Ala}$	pQE60 construct expressing terminal 501 bp of Tal$_{901-1}$ from mutant TP901-1*erm*$_{His892Ala}$	This study

Supplementary Table 1. Oligonucleotides used in this study.

Primer name	Primer sequence (5' → 3')	Notes
LysM-FP	ACGACGCCATGGAACATCACCATCACCATCACGATTCTACAAGCCAGCCAACG	NcoI
LysM-RP	GTCCGTTTAAATAATTTAGTGTTTGACCAGCGTA	SOEing PCR overlapping oligo
Tal_{2009}M23-FP	ACACTAAATTATTTAAACGGACACCCTGAACGTA	SOEing PCR overlapping oligo
Tal_{2009}M23-RP	ACGACGTCTAGATTAAAATTTGATATAATCCCTTGGATTCTTAAAGTG	XbaI
$Tal_{901\text{-}1}$-5ΔN-FP	ACGACGCCATGGGATTAAACGGACACCCTGAACGTAGT	NcoI
$Tal_{901\text{-}1}$-5ΔN-RP	ACGACGAGATCTAAATTTGATATAATCCCTTGGATTCTTAAAGTGAG	BglII
NZ-attB-Fw	TGTGAAGGCTCTAATCATCG	Forms PCR product with NZ-attB-Rv primer using NZ9000 DNA as template
NZ-attB-Rv	ATGCTGATGAATGTTGGGAAC	
TP-attP-Fw	GAGATAACTGATGAACTATGG	Forms PCR product with NZ-attB-Fw primer using NZ9000-TP901-1*erm* DNA as template
TP-attB-Rv	CCTATTCTTGATACTTCGTCA	
oJP1161Ψ	C*T*T*C*T*GGTGGGTATTGACCATTCCAACCTGTATCACCTCGGCGTCTATTTCGGCCACCATTAGTATCAATTTTTGTTCCGTTGACATAG	Gly>Arg-creating recombineering oligo
oJP1162	ATTGAAGAATCTGATCATTTTCTTGTAGC	Oligo screening Gly>Arg
oJP1163¥	AACCTGTATCACCTCGGCGTCTA	Oligo screening Gly>Arg
oJP1164	TTGGAGCAATATAACCTCCGCC	Oligo screening Gly>Arg
oJP1168Ψ	G*A*T*G*A*TGGCCAATATTGGTCCATGAATTGGAAGTGCAGTGCCGGTCCAGTGACCGGACCAGTCGCTCCCATAAGTCCAAT	His892Ala-creating recombineering oligo
oJP1169	ACAAAACGGCCAATGGATTGCG	Oligo screening His892Ala
oJP1170¥	ATGAATTGGAAGTGCAGTGCC	Oligo screening His892Ala
oJP1171	TCTTCAATATACTTATTAAAGATTCG	Oligo screening His892Ala
oJP1106Ψ, §	T*G*A*C*T*TCTGGTGGGTATTGACCATTCCAACCTGTAT*AGATCTA*GCCAGAGTTTCCGCCACCATTAGTATCAATTTTTGTT	Gly603*-creating recombineering oligo
oJP1110	ATCATTTTCTTGTAGCTAACTCAGAAGG	Oligo screening Gly603*
oJP1111	AACAAATTTGTTATACCATTCTTGGGCC	Oligo screening Gly603*

Ψ Nucleotides of recombineering oligos which are designed to alter the base sequence of $Tal_{901\text{-}1}$ are underlined.
¥ Nucleotides of screening oligos which bind to mutating bases are similarly underlined.
* Phosphorothioate linkages of recombineering oligos
§ Novel BglII restriction site indicated in italics.

Supplementary Table 2. Proposed structures of muropeptides separated by RP-HPLC.

Peak[a]	Proposed Structure[b]	Peak	Proposed Structure	Peak	Proposed Structure
Monomers		Dimers		Trimers	
1	Tri	16	Tri-D-Tetra-D	31a	Tri-N-Tetra-N-Tetra-N (deAc)
2	Tri-D missing GlcNAc	17	Tri-N-Tetra	31b	Tri-N-Tetra-N-Tetra-D[c]
3	Tetra	18	Tri-N-Tetra-D[c] missing GlcNAc	32	Tri-N-Tetra-N-Tetra-N
A	Tri-N missing GlcNAc	19	Tri-N-Tetra-D[c]	33a	Tetra-N-Tetra-N-Tetra-N (deAc)
4	Tri-D	20a	Tri-N-Tetra-D[c] (deAc)	33b	Tetra-N-Tetra-N-Tetra-D[c]
5	Tri-D with (NH_2)[d]	20b	Tri-N-Tetra-N missing GlcNAc	34	Tetra-N-Tetra-N-Tetra-N missing GlcNAc
6	Penta	21	Tri-N-Tetra-N (deAc)	35a	Tetra-N-Tetra-N-Tetra-N
7	Tri-N (deAc)	22	Tetra-N-Tetra	35b	Penta-N-Tetra-N-Tetra-D[c]
8	Tri-N	23	Tri-N-Tetra-N	36	Penta-N-Tetra-N-Tetra-N
B	Tetra-N missing GlcNAc	24	Tetra-N-Tetra-D[c]		
9	Tetra-D	25	Penta-N-Tetra-D[c]	Tetramers	
10	Penta-D	26	Tetra-N-Tetra-N	37	Tri-N-Tetra-N-Tetra-N-Tetra-N
11	Tetra-N	27	Penta-N-Tetra-N	38	Tetra-N-Tetra-N-Tetra-N-Tetra-N
12	Penta-N	28	Tri-N-Tetra-D[c] (Ac)	39	Penta-N-Tetra-N-Tetra-N-Tetra-N
13	Tri-N (Ac)	29	Tetra-N-Tetra-D[c] (Ac)		
14	Tetra-D (Ac)	30	Tetra-N-Tetra-N (Ac)		
15	Tetra-N (Ac)				

[a]Peak numbers refer to Fig. 3.
[b]Structures were deduced from MALDI-TOF MS analysis and according to the reference profile established by Courtin *et al.* (2006). Tri, disaccharide tripeptide ($_L$-Ala-$_D$-iGln-$_L$-Lys); Tetra, disaccharide tetrapeptide ($_L$-Ala-$_D$-iGln-$_L$-Lys-$_D$-Ala); Penta, disaccharide pentapeptide ($_L$-Ala-$_D$-iGln-$_L$-Lys-$_D$-Ala-$_D$-Ala); disaccharide, GlcNAc-MurNAc; D, $_D$-Asp; N, $_D$-Asn; (NH_2), amidation; deAc, deacetylation; Ac, acetylation; iGln, α-amidated isoGlu (isoGln).
[c]Positions of N and D in either peptide chain are arbitrary.
[d]Position of amidation was not identified.

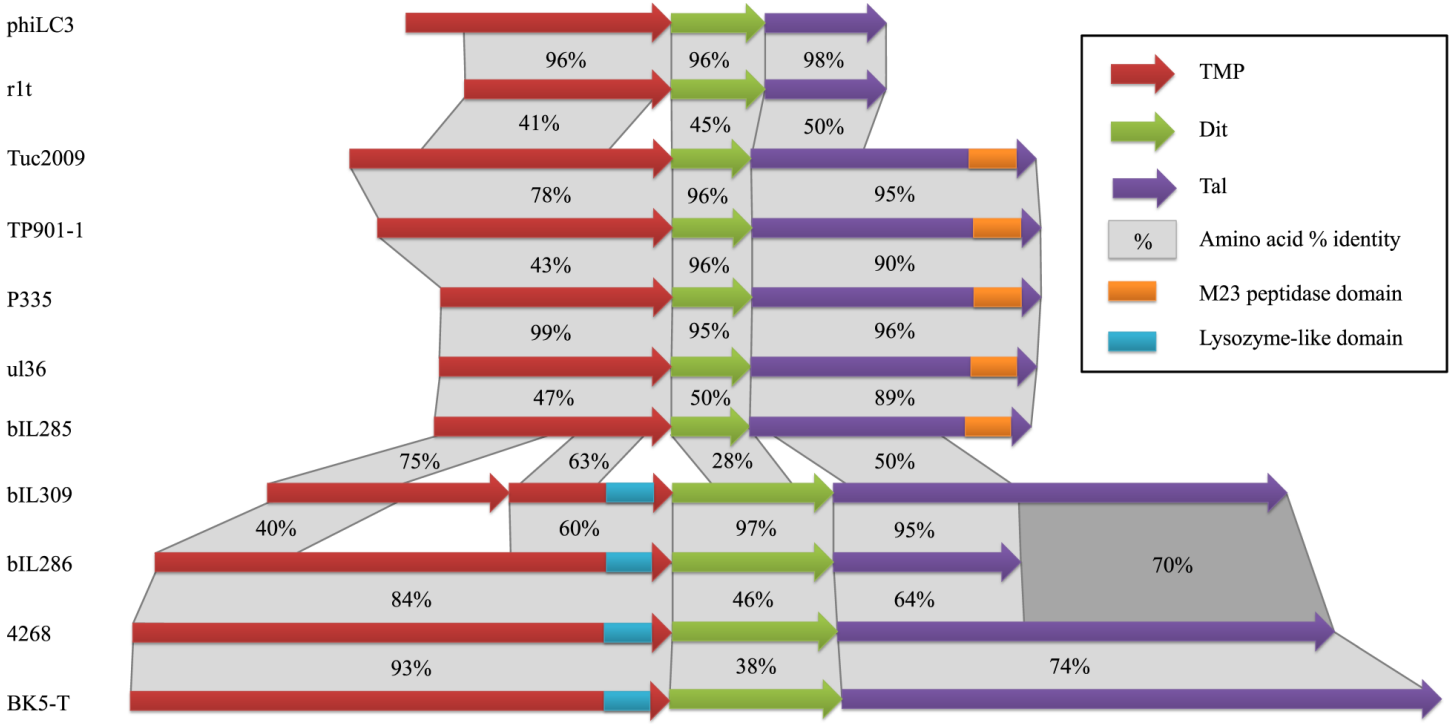
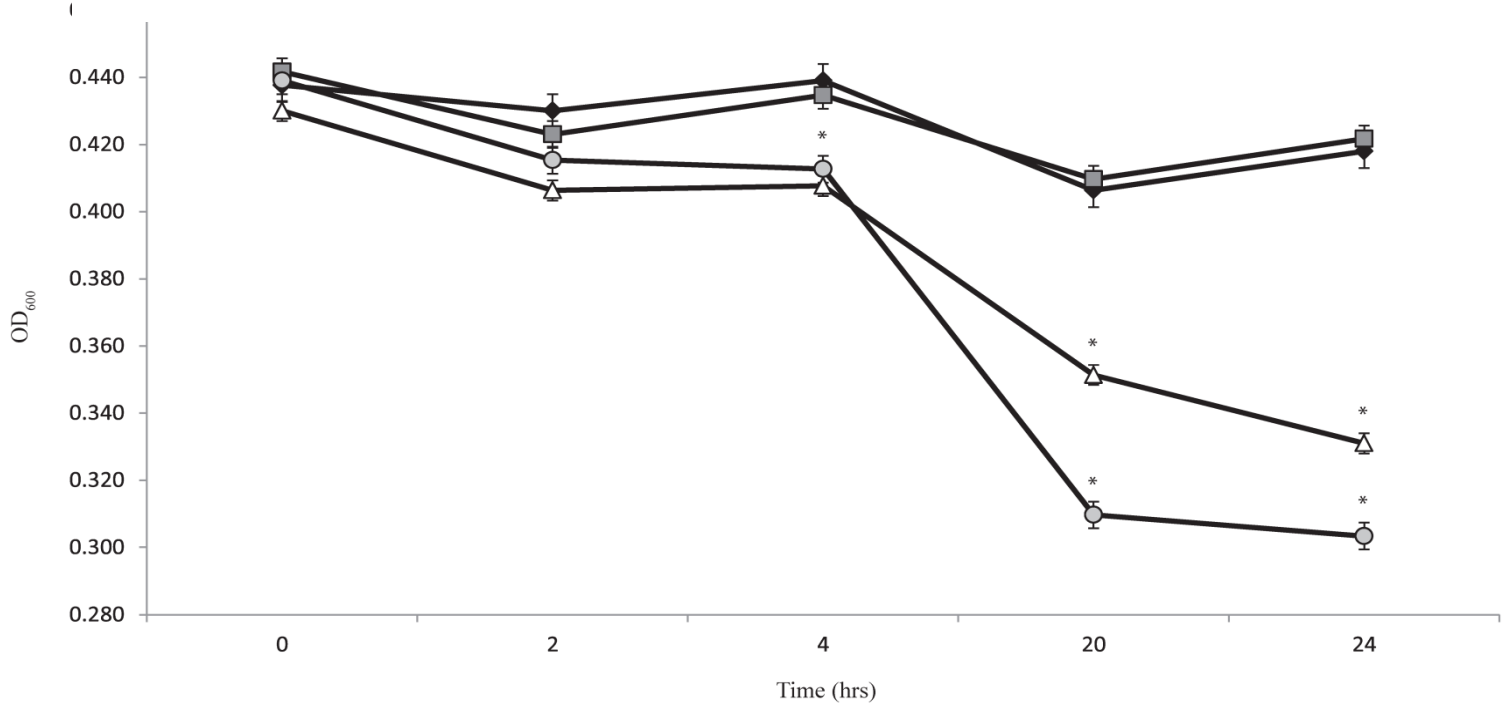


Figure 1. Comparative analysis of the TMP, Dit, and Tal proteins from P335-species phages. The locations of predicted peptidoglycan hydrolase domains are highlighted by coloured boxes, as indicated in the legend. The percentage amino acid identity between protein sequences was estimated by BLAST2p and is indicated in the shaded grey boxes.

A


1
740
907
M23 peptidase
Tal2009
M23 peptidase
Tal2009-5ΔN
LysM
LysM
M23 peptidase
TalM23-LysM

B


OD600
0.440
0.420
0.400
0.380
0.360
0.340
0.320
0.300
0.280
0
2
4
20
24
Time (hrs)
*
*
*
*
*

C


1
752
His892
919
M23 peptidase
Tal901-1
His892
M23 peptidase
Tal901-5ΔN
Ala892
M23 peptidase
Tal901-5ΔNHis892Ala

D

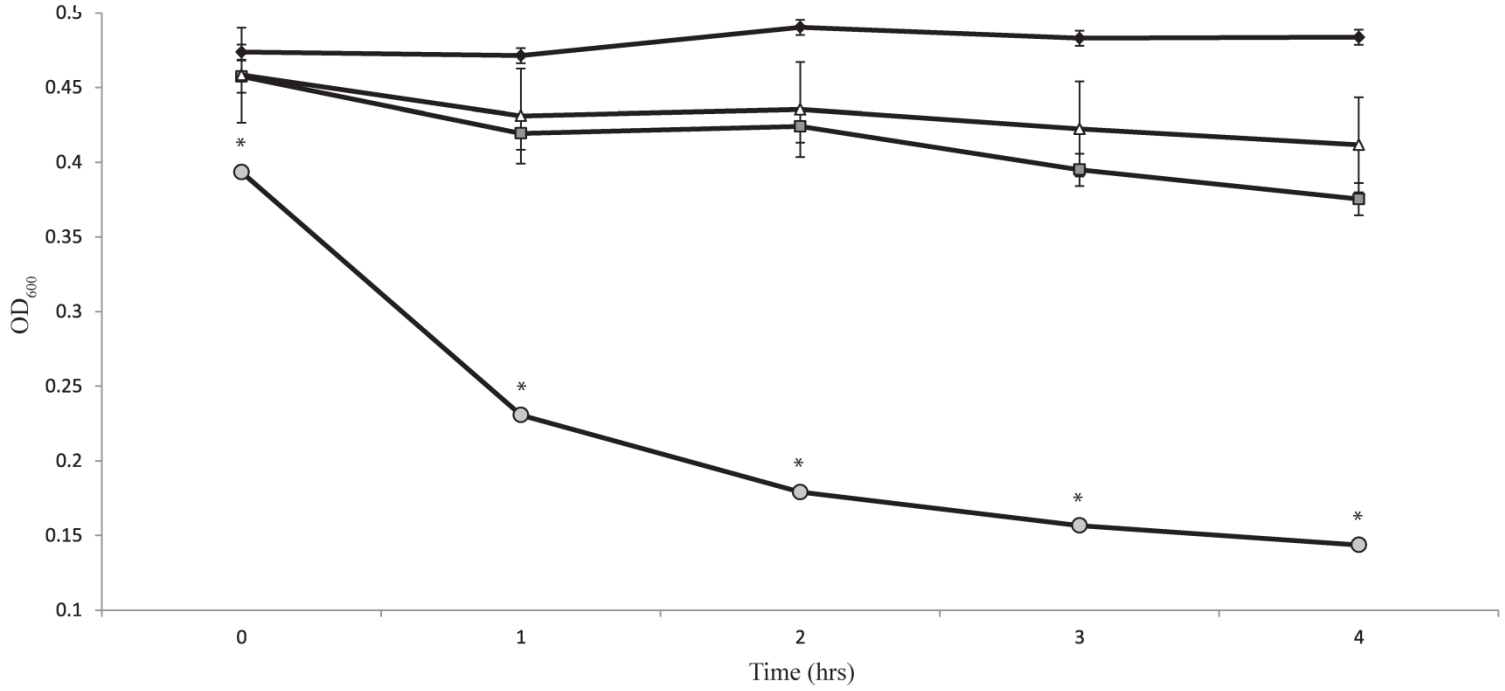

0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
OD600
0
1
2
3
4
Time (hrs)
*
*
*
*
*

Figure 2. Analysis of peptidoglycan hydrolase activity associated with M23 peptidase of Tal protein constructs. **(A)** Diagrammatic representation of Tal_{2009} tail fibre, and recombinant proteins Tal_{2009}-5ΔN and TalM23-LysM. **(B)** Hydrolysis of *L. lactis* UC509.9 cell wall. The negative control (water) is marked as black diamonds, mutanolysin-only is marked with dark grey squares, mutanolysin and Tal_{2009}-5ΔN is marked with white triangles, and mutanolysin and TalM23-LysM is marked with light grey circles. Error bars indicate standard deviation, while asterisks are located above statistically significant values (*p*-value < 0.001). **(C)** Diagram displaying $Tal_{901\text{-}1}$ tail fibre and recombinant proteins $Tal_{901\text{-}1}$-5ΔN and $Tal_{901\text{-}1}$-5Δ$N_{His892Ala}$. The histidine or alanine amino acid at position 892 in the M23 peptidase catalytic site of constructs $Tal_{901\text{-}1}$-5ΔN and $Tal_{901\text{-}1}$-5Δ$N_{His892Ala}$, respectively, is highlighted. **(D)** Hydrolysis of *L. lactis* 3107 cell wall. The negative control (water) is marked as black diamonds, mutanolysin-only is marked with dark grey squares, mutanolysin and $Tal_{901\text{-}1}$-5Δ$N_{His892Ala}$ is marked with white triangles, and mutanolysin and $Tal_{901\text{-}1}$-5ΔN is marked with light grey circles. Error bars indicate standard deviation, while asterisks are located above statistically significant values (*p*-value < 0.001).

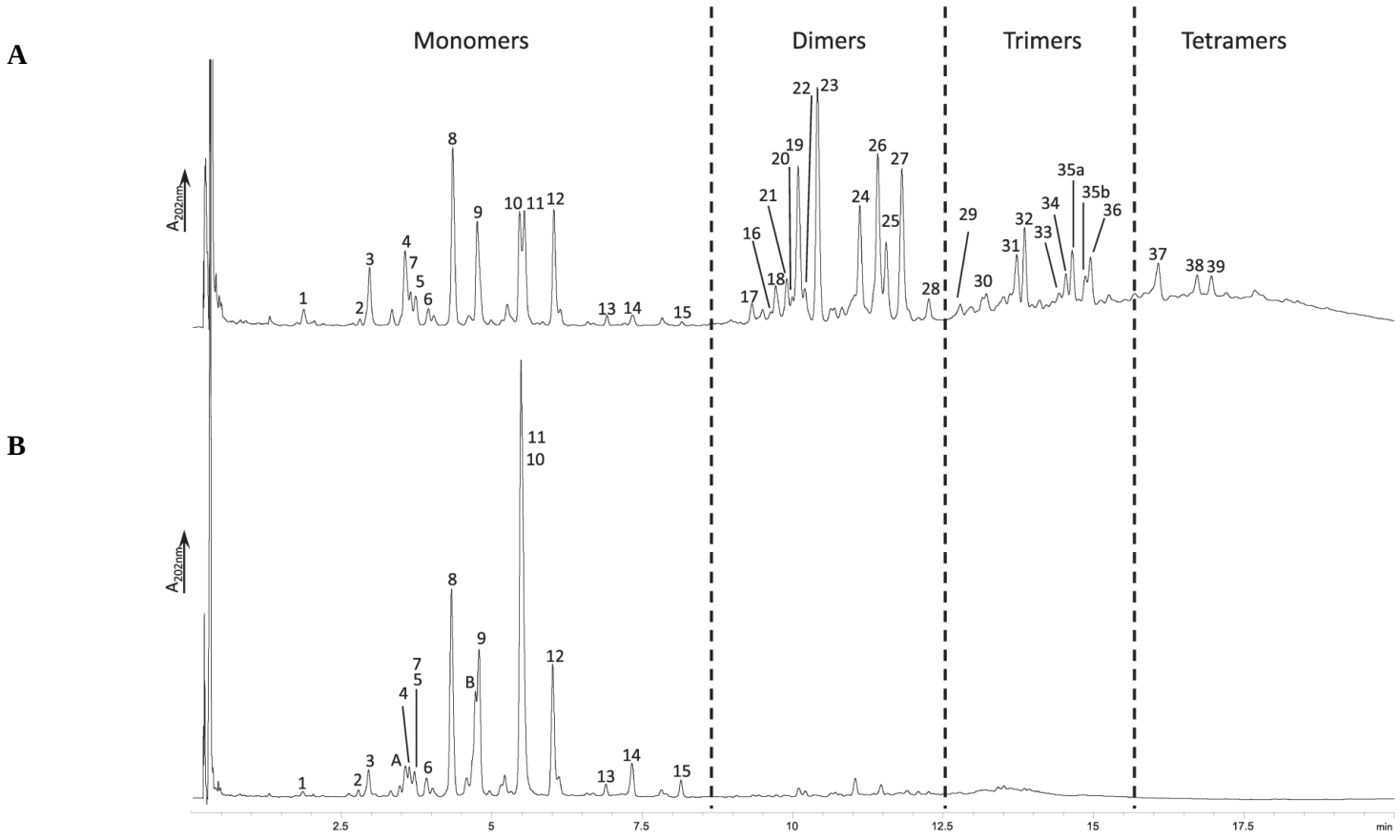


Figure 3. RP-HPLC separation of muropeptides obtained by digestion of *L. lactis* PG by **(A)** mutanolysin and **(B)** mutanolysin + TalM23-LysM. Peak numbers refer to structures identified by MALDI-TOF (Supplementary Table 2).

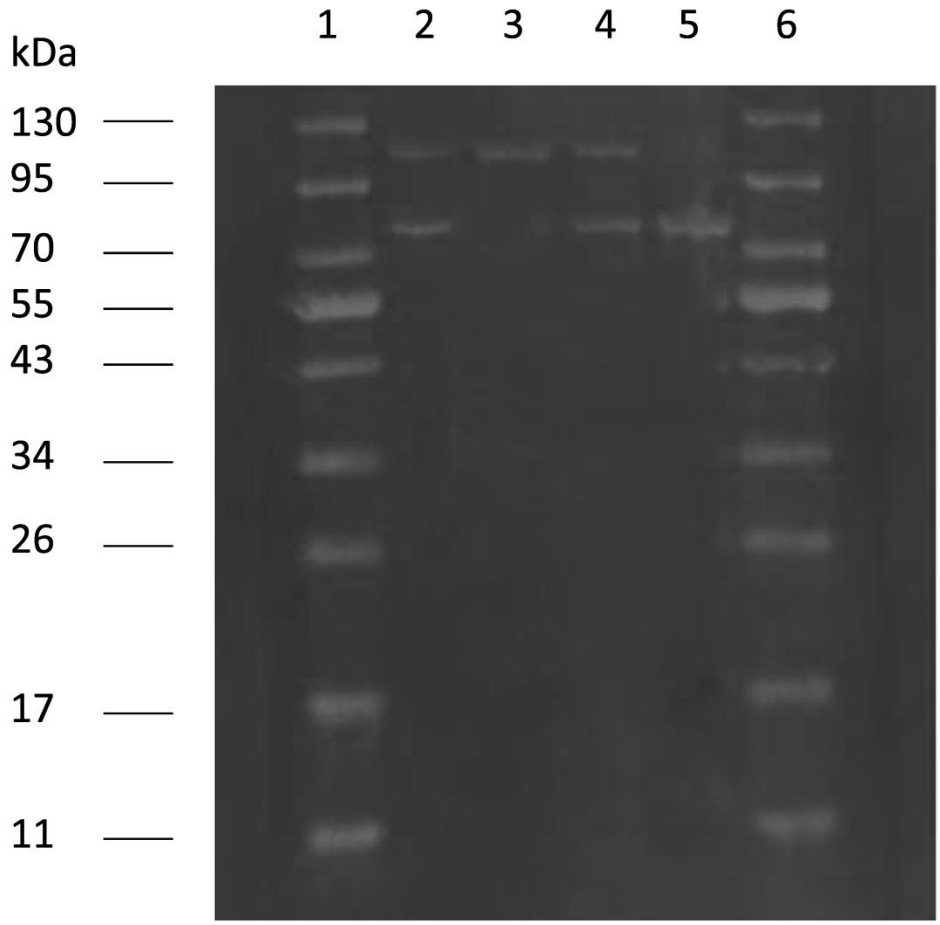
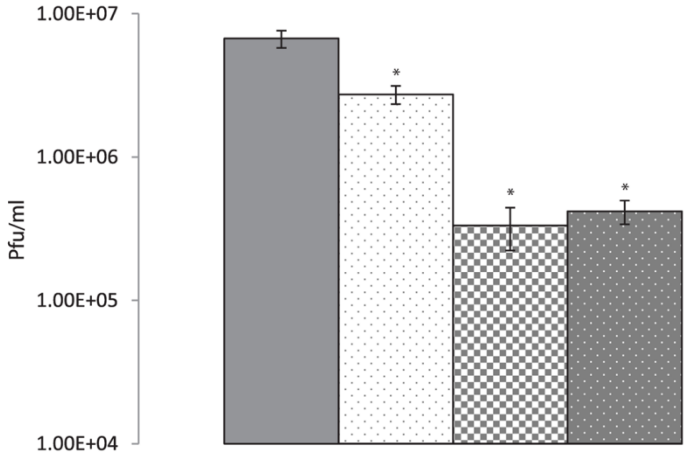
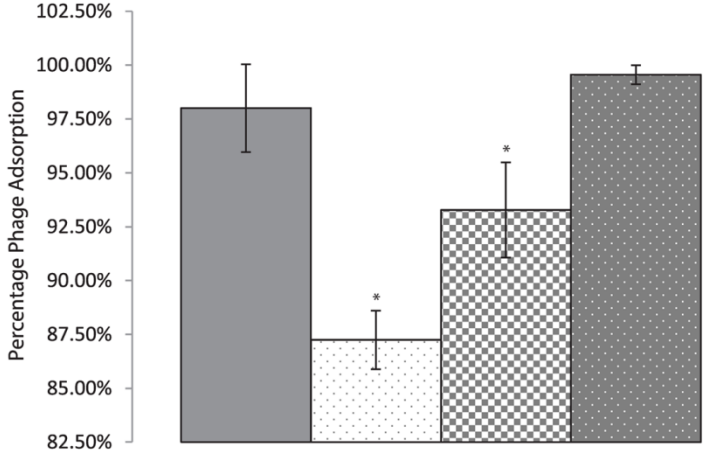


Figure 4. Western blots of TP901-1*erm* and mutant derivatives using anti-N-terminal Tal_{2009} antibodies. Lanes 1 & 6, molecular weight ladder; lanes 2-5, TP901-1*erm*, TP901-$1_{Gly>Arg}$, TP901-$1_{His892Ala}$ and TP901-$1_{Gly603*}$, respectively.

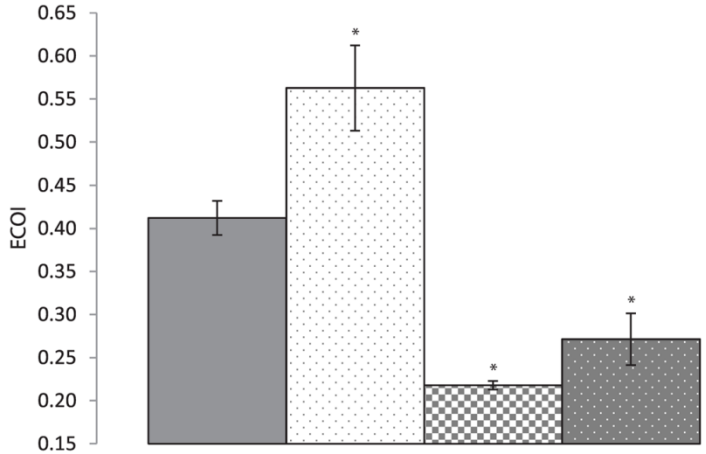
A


1.00E+07
1.00E+06
1.00E+05
1.00E+04
Pfu/ml

B


102.50%
100.00%
97.50%
95.00%
92.50%
90.00%
87.50%
85.00%
82.50%
Percentage Phage Adsorption

C


0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
ECOI

D

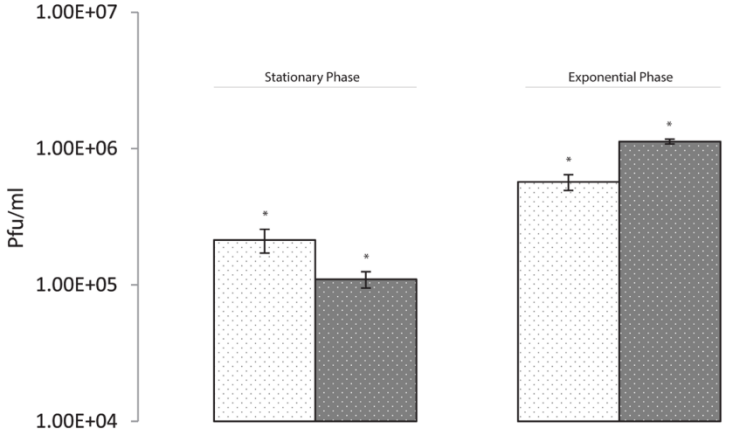

1.00E+07
1.00E+06
1.00E+05
1.00E+04
Pfu/ml
Stationary Phase
Exponential Phase

Figure 5. Bar charts showing phage assays of TP901-1*erm* and mutant derivatives. TP901-1*erm* results are in solid grey, TP901-1$erm_{Gly>Arg}$ are white with grey dots, TP901-1$erm_{His892Ala}$ are grey and white checkers, and TP901-1$erm_{Gly603*}$ are grey with white dots. Error bars indicate standard deviation, while asterisks indicate mutants that are statistical significance (p-value < 0.05). **(A)** Plaque assays of TP901-1*erm* and mutant derivatives. **(B)** Adsorption assays of TP901-1*erm* and mutant derivatives, measuring the percentage of bacterial-bound phage in 10 minutes. **(C)** Efficiency of Centre of Infection (ECOI) formation by TP901-1*erm* and mutant derivatives. Each ECOI was calculated as the number of Centres of Infection (COIs) formed on stationary phase cells relative to the number of COIs formed on exponentially growing cells. (**D)** Competition assays between TP901-1$erm_{Gly>Arg}$ and TP901-1$erm_{Gly603*}$, comparing their ability to form Centres of Infection (COIs) within 5 minutes. TP901-1$erm_{Gly>Arg}$ are white with grey dots, TP901-1$erm_{Gly603*}$ are grey with white dots. Mutant phage TP901-1$erm_{Gly>Arg}$ permanently retains the C-terminus of its tail fibre protein, which possesses the M23 peptidase domain, while mutant TP901-1$erm_{Gly603*}$ incorporates only a truncated derivative of Tal_{901-1}.

CHAPTER III

Mutational and Molecular Analysis of Lactococcal Phage TP901-1

Notes:

Recombinant protein-expression constructs, their overproduction, and subsequent antibody generation was conducted by Dr. Barry Collins and Dr. François Douillard.

Tuc2009 time-course infection assay and Western-blots were performed by Dr. Barry Collins.

Electron micrographs were taken by Dr. Silvia Spinelli.

This chapter is in preparations for publication as a research article:

Stockdale, S. R., Collins, B., Spinelli, S., Douillard, F. P., Mahony, J., Cambillau, C., van Sinderen, D. Mutational and Molecular Analysis of Lactococcal Phage TP901-1. *Manuscript in preparation.*

ABSTRACT

Members of the *Siphoviridae* family represent the most abundant viral morphology in the biosphere, yet many aspects of their virion structure and associated functions have yet to be determined. In this study, we present a comprehensive mutational and molecular analysis of lactococcal bacteriophage TP901-1. Inactivation of twelve genes located within the structural module of TP901-1 through the generation of non-sense mutations, and the introduction of two additional mutations causing aberrant protein production, allowed us to determine the essential nature of many of the corresponding gene products and to provide information on their functionality. Furthermore, *in vitro* and *in vivo* assembly of phage virions was assessed in this study. Based on the obtained information, we propose a functional model for the proteins that make up the TP901-1 virion.

INTRODUCTION

Siphoviridae phages represent the dominant viral morphology, frequently outnumbering other viral morphotypes in environments such as the human gut and oceanic waters (1-4). Despite the prevalence of phages with long non-contractile tails, only a handful of *Siphoviridae* have been extensively studied at a structural level (for a review, see (5)). Extensive research in the field of dairy lactococci and their infecting phages has been undertaken since their discovery (6-9). Consequently, lactococcal phages have become one of the best characterized group of bacterial viruses that occupy a defined ecological niche and infect a single bacterial species (10). The prototypical and related phages TP901-1 and Tuc2009 infect the economically important dairy-starter bacterium *Lactococcus lactis*, and have been important in understanding lactococcal and other Gram-positive *Siphoviridae* due to the genetic accessibility of their Gram-positive hosts, the various molecular tools available for manipulating their genomes and the conserved nature of phage structural proteins.

Recently, a composite model of the TP901-1 virion was constructed by single-particle electron microscopy (EM) analyses of the phage's capsid, head-tail connector, tail-tube and baseplate regions (11). Several studies have focused on characterizing the structure and function of lactococcal phage baseplates, as these tail-tip structures possess receptor-binding proteins that are involved in the specific recognition of and initial interactions with their hosts (12-22). While the capsid, head-tail connector, and tail-tube assemblages of TP901-1 and Tuc2009 phages are not as well described, much can be predicted regarding these regions due to the evolutionary conserved nature of tailed phages and their structural proteins (23).

Morphogenesis of a typical icosahedral phage capsid requires three essential factors: the portal protein, also known as the connector protein; the major head protein (MHP); and a scaffolding protein, or a scaffolding domain associated with the MHP (5). The assembly of functional capsids, through an intermediate DNA-free procapsid structure, necessitates MHP subunits to assemble into 12 pentamers and often a variable number of hexamers, and incorporate a dodecamer portal protein complex into a unique capsid vertex (24). While scaffolding proteins or domains are not believed to be part of mature capsid structures (25), portal protein complexes and scaffolding proteins have been demonstrated to interact and are often proposed to initiate capsid morphogenesis (26-29). In a model of phage P22 capsid assembly, Prevelige and colleagues (1993) visualized incomplete capsid structures which showed scaffolding proteins and viral coat proteins sequentially added to the edges of the assembling procapsid shell (30, 31), suggesting a mechanism by which a single portal vertex nucleates capsid morphogenesis. Head-tail connector proteins of *Siphoviridae* phages associate with the portal protein-containing vertex of DNA packaged capsids and, as their name suggests, serve as an attachment site for the tail organelle (32). Head-tail connector protein gp16 of phage SPP1 also fulfils a capsid DNA-plug role (33). Finally, mature phage capsid structures can be decorated with various accessory proteins; previous studies have demonstrated that TP901-1 and several other lactococcal phages possess neck passage structure (NPS) proteins which may bind to a protein, or proteins, associated with the portal vertex (34).

Siphoviridae tail assembly requires the tail tape measure protein or TMP, probably as a hexameric complex (35, 36), to dictate the length of the phage tail structure (37, 38). Two tail chaperone proteins, gpG and gpGT, are essential for the correct assembly of *Siphoviridae* tails. Chaperone gpGT is produced through a -1

ribosomal frame-shift during mRNA translation at a specific sequence corresponding to the 3' end of gene *gpG*, which represents a genetic peculiarity conserved in many dsDNA tailed phages (39). Protein gpGT binds the hydrophobic TMP through the G domain and recruits the phage major tail protein (MTP) via its T domain (40, 41). Polymerization of MTP around and along the TMP to form the phage tail tube pauses when the end of the TMP is reached (42). Tail terminator proteins associate with the terminal hexameric ring of *Siphoviridae* tail tubes to complete MTP polymerization (43-46). In phage λ, tails also require activation through an unknown mechanism by protein gpZ before they can associate with phage capsids to produce complete virions (42).

In the current study, we analysed the effect of fourteen mutations introduced into genes ORF32-44 of the structural module of TP901-1 on the efficiency of plaquing and on phage virion integrity, as determined by *in vitro* assembly, immunological detection and EM. This new data complements and significantly expands current knowledge on lactococcal *Siphoviridae* (34, 37, 47), allowing the formulation of a detailed molecular model describing the structure, function and assembly of the TP901-1 virion.

MATERIALS AND METHODS

Bioinformatic Analyses. DNA sequences were downloaded from NCBI GenBank (48). Inducible prophage t712 of NZ9000, described by Ventura *et al.* (2007) (49), was identified using the PHAge Search Tool (PHAST) (50). BLAST, Pfam and HHpred analyses were used for functional annotations of proteins (51-54). Putative promoter sequences of NZ9000 prophage t712 were identified using SoftBerry BPROM (http://www.softberry.com). Sequence percentage similarities between relevant TP901-1 and Tuc2009 proteins were calculated using LALIGN with default settings (55), and are listed in Table 1. 'Dot plot'-like graphs comparing protein sequences were performed using PLALIGN (http://fasta.bioch.virginia.edu). The results of significant HHpred hits of TP901-1 and Tuc2009 structural module proteins are provided in Table 1, alongside Protein Data Bank (PDB) identification codes. Alignment of (multiple) amino acid sequences was performed using ClustalW (56). Protein secondary structure predictions and intrinsically disordered regions were detected using Quick2D (http://toolkit.tuebingen.mpg.de/) (57-61).

Bacteria, Phages and Growth Conditions. Bacterial strains and phages used in this study are listed in Supplementary Table 1. Growth experiments of lactococcal strains and phage propagations were performed as described previously (62, 63). TP901-1*erm* and mutant derivatives were induced from corresponding lysogens of *L. lactis* NZ9000-Cro_{t712} under the following conditions: the relevant strains were grown at 30 ^{o}C to an A_{600} of 0.3, at which point mitomycin C (Sigma) was added to a final concentration of 3 µg/ml with subsequent incubation at 20 ^{o}C until lysis occurred. NaCl, 1 M (w/v) final concentration, was added to the resultant phage

lysates followed by centrifugation in a Thermo Scientific SL16R centrifuge at 5580 × *g* for 15 min and stored at 4 ^{o}C.

Mutant Generation. Recombineering mutagenesis was performed as described previously (63-65). A detailed description of all TP901-1*erm*-derived mutants created in this study, and their nomenclature, is provided in Supplementary Table 2. Point mutations, creating a BamHI restriction site, were introduced in the deduced promoter sequence of the predicted anti-repressor-encoding gene (cro_{t712}) of prophage t712 of NZ9000 in order to prevent its induction during mitomycin C addition (generating strain NZ9000-Cro_{t712}) (66). Mutations in targeted ORFs of the TP901-1 structural module were generated so as to insert an in-frame stop codon, thereby terminating translation (mutant nomenclature $ORF_{TP901-1}$::Ter), and where possible, introduce a restriction enzyme site. In order to prevent gpT production, an in-frame BamHI restriction site was introduced into the 'slippery' sequence (GGGAAAG) at the 3' end of the TP901-1 *gpG* gene which is required for the sequence-specific -1 ribosomal frame-shift required for gpT translation (mutant termed $gpT_{TP901-1}$::BamHI) (39). To further analyse the production and role of chaperone protein gpT, a single nucleotide insertion, coupled with several point mutations, was introduced into the *gpG* slippery sequence resulting in a direct translational fusion between the *gpG* and *gpT* gene sequences (the resulting mutant phage was designated gpG*fs*$T_{TP901-1}$). Mismatch Amplification Mutation Assay (MAMA) PCRs and restriction digests were performed to identify TP901-1*erm* prophages containing the desired mutations. Oligonucleotides used for recombineering and MAMA PCR screening reactions were purchased from Integrated DNA Technologies (IDT, Belgium), and are listed in Supplementary Table 3. Mutations were confirmed by Sanger sequencing relevant PCR-amplified

regions using chromosomal DNA from the mutated TP901-1*erm*-containing *L. lactis* NZ9000-Cro_{t712} lysogen as a template (sequencing performed by MWG, Germany).

Phage Purification and Sucrose Gradients. Phage lysates were purified by sucrose gradient ultracentrifugation as follows: polyethylene glycol 8000 (PEG 8000; 10 % w/v) was added to phage lysates and mixed gently at Room Temperature (RT) until completely dissolved. Phage samples were placed on ice for 3-4 hr before centrifugation in a 4 ^{o}C pre-chilled Thermo Scientific SL16R centrifuge at 5580 × *g* for 20 min. Pellets were air-dried, before resuspending in 1:50 initial volume of SM buffer (50 mM Tris-HCl, 100 mM NaCl, 10 mM $MgSO_4$; pH 7.5). PEG 8000 was removed from phage samples by chloroform-extraction with an equal volume of chloroform, followed by vortexing for 30 sec and centrifugation at 1660 × *g* for 15 min. The aqueous phase was extracted and applied to a sucrose gradient.

Linear sucrose gradients, 20-70 % (w/v), were prepared using a Masterflex peristaltic pump (Thermo Scientific) in Beckman Thinwall Ultra-Clear centrifuge tubes (14 x 89 mm). Sucrose was added to all phage samples, at a final concentration of 14 % (w/v), before layering them on top of the prepared linear sucrose gradients. Samples were centrifuged for 1.5 hr at 250,000 × *g* in a Beckman SW40Ti ultracentrifuge rotor at 4 ^{o}C. Bands were extracted from the sucrose gradient and immediately dialysed against SM buffer (12-14 kDa MW cut-off; Medicell International Ltd., London, UK.).

Phage Assays. Spot and plaque assays of TP901-1 and Tuc2009 were performed as described previously (67). *In vitro* assembly of TP901-1 structural module mutants was performed using sucrose gradient-purified bands of TP901-1*erm* and mutant derivatives of this phage. Sucrose gradient-purified fractions of

mutant $MHP_{TP901\text{-}1}$::Ter and mutant $MTP_{TP901\text{-}1}$::Ter were used as either a phage tail or capsid donor, respectively. Briefly, 10 µl of purified TP901-1*erm* capsids or tails were mixed with 10 µl of purified TP901-1*erm* structural module mutant derivative fractions, and incubated for 16 hr at 30 ^{o}C. The phage mix was then tested for its ability to infect *L. lactis* 3107 by assessing the frequency of lysogeny, which was determined through acquisition of the adenine methylase gene present on the TP901-1*erm* phage genome (or its mutant derivatives), conferring resistance to 5 µg/ml erythromycin (Erm5) (68, 69). This was performed by adding 50 µl of the phage mix to 450 µl of *L. lactis* 3107 at an A_{600} of ~0.3 and incubated at 30 ^{o}C for 1 hr before plating. Three technical replicates calculating the frequency of lysogeny were performed and results averaged.

Protein Expression, Purification and Antibody Generation. Recombinant protein-expression constructs created during this study, and the oligonucleotides used to create them, are outlined in Supplementary Tables 1 and 3, respectively. Primers used for cloning were purchased from Eurofins MWG (Germany). KOD high-fidelity DNA polymerase (Novagen, UK) was used for PCR amplifications. For cloning reactions, restriction enzymes were supplied by Roche (Germany) and ligations were performed with T4 DNA ligase (Promega, USA). For genes cloned using pNZ8048, NcoI and SpeI were used, while for pTX8049, BamHI and SpeI were employed. In both cases, primers were used so as to facilitate the incorporation of a hexahistidine-tag at the C-terminus of the encoded protein, while pTX8049-generated proteins also incorporate an N-terminally fused thioredoxin tag (70). Proteins were expressed and purified as outlined previously (70). Rabbit antibody production was carried out by Harlan laboratories (Leicester, UK.). Immunization was initially carried out with individual proteins supplemented with Freund's

adjuvant at a concentration of ~200 μg/ml; this was followed by five booster injections over the 112 day protocol.

Tuc2009 Protein Expression. Antibodies against Tuc2009 tail proteins were previously generated and proved successful in detecting the homo-immune structural proteins of Tuc2009 and TP901-1 (71). Therefore, the following Tuc2009 capsid proteins were overproduced and purified in *L. lactis* NZ9000: $Portal_{Tuc2009}$, $MCP1_{Tuc2009}$, $MCP2_{Tuc2009}$, $Sfp_{Tuc2009}$, $MHP_{Tuc2009}$, $MCP3_{Tuc2009}$, $HTC1_{Tuc2009}$ and $gpZ_{Tuc2009}$. Attempts to express recombinant proteins of $HTC2_{Tuc2009}$ and $gpU_{Tuc2009}$ were unsuccessful (data not shown).

SDS-PAGE and Western Blotting. SDS-PAGE separation of proteins was performed using 12 % or 15 % (v/v) acrylamide gels following standard procedures (72). After protein separation by SDS-PAGE, the proteins were transferred to a nitrocellulose membrane by electro-blotting using 10 mM CAPS [3-(cyclohexylamino)-1-propanesulfonic acid; pH 11], 10 % methanol transfer buffer. Membranes were blocked for 1 hr at RT with 50 % (v/v) Odyssey blocking buffer in PBS (Licor, USA). Optimal conditions for each rabbit polyclonal antibody were determined empirically (data not shown). Membranes were incubated with primary rabbit polyclonal antibodies, in a range from 1:500-1:15000, in blocking buffer containing 0.12-0.25 % (v/v) Tween-20 for 1 hr at RT or overnight at 4 °C. Following 4 × 5 min washes in 0.1 % (v/v) Tween-20, the membranes were incubated in 1:7500 fluorescently labelled goat anti-rabbit IgG secondary antibody (Licor, USA) in blocking buffer containing 0.12-0.25 % Tween-20. Membranes were washed again and visualised on a Licor Odyssey imaging system (USA).

Tuc2009 Infection Assay. The Tuc2009 infection time-course assay was adapted from a previously described approach (71). *L. lactis* UC509.9 was grown to A_{600} of ~0.3 and infected with Tuc2009 to give a multiplicity of infection of 0.1. Aliquots of UC509.9 were taken 10 min before infection and every 10 min post-infection for 70 min. Samples were rapidly frozen by immediately placing them in an alcohol dry ice bath. Samples were subsequently thawed on ice and the cells collected by centrifugation in a 4 ^{o}C pre-chilled Thermo Scientific SL16R centrifuge at 19,000 × *g*. Cell pellets were resuspended in ice-cold TBT buffer (10 mM Tris-HCl, 300 mM NaCl, 50 mM $CaCl_2$; pH 8), before acid-washed glass beads (≤ 106 µm; Sigma) were added and then bead-beaten (Mini beadbeater-8, Biospec Products) for 10 × 1 min periods, with 1 min intervals on ice between each beat period. Clarified cellular extracts were prepared by centrifugation as described above and the protein concentration of these extracts were established using the Bio-Rad protein assay (73). Equal protein concentrations, 20 µg per well, were loaded into SDS-PAGE gel lanes.

Electron Microscopy. For negative staining, 5 ul of each sample were applied onto glow-discharged carbon-coated grids (Agar Scientific) and incubated for 1 min. The grids were washed with 5 µl of deionized water before incubating for 30 sec in 1% (w/v) of Uranyl Acetate (Agar Scientific, Stansted, UK.). CCD images were collected using a Tecnai Spirit operated at 120 Kv and a 2Kx2K CCD camera.

RESULTS

Annotation of TP901-1 Structural Module. In order to predict the function(s) of TP901-1 structural module-encoded proteins, a detailed annotation was performed. Protein sequences of TP901-1 ORFs were subjected to BLAST non-redundant database, Pfam and HHpred searches (51, 52, 74). Also, due to the conserved nature of phage structural proteins and genetic synteny of lambdoid-like phages (75, 76), TP901-1 was compared with the closely related lactococcal phage Tuc2009, *Bacillus subtilis* phage SPP1, and enterobacterial phage λ. The individual structural modules of TP901-1 and Tuc2009 are represented by polycistronic operons, each approximately 20 kb in length, transcribed from a promoter located upstream of the gene encoding their predicted small terminase subunits (77-79). The deduced structural module of TP901-1 encodes 22 proteins (Fig. 1) (77), while the structural module of phage Tuc2009 contains one additional gene, which specifies the host adsorption-enhancing baseplate protein BppA, and which is located between the baseplate-encoding genes $bppU_{Tuc2009}$ and $bppL_{Tuc2009}$ (78).

The majority of TP901-1 and Tuc2009 structural proteins are conserved between these phages. Eleven of the twenty TP901-1 and Tuc2009 common proteins analysed in this study share > 90 % amino acid (aa) similarity, while a further six proteins share > 60 % aa similarity. As the GenBank ORF numbers assigned to TP901-1 and Tuc2009 differ, we have assigned, and will henceforth refer to, a common name to proteins that are presumed (based on their similarity, or other reasons, as outlined below) to fulfil similar functions (Table 1).

Previous studies have determined the functions for several conserved TP901-1 and Tuc2009 structural proteins (13, 16, 34, 37, 47, 63, 69, 80, 81). Due to the

conserved nature of phage structural proteins, it is possible to predict the functions of several other TP901-1 and Tuc2009 proteins based on well characterized phage proteins (33, 46, 82-84). However, a number of TP901-1 and Tuc2009 proteins do not exhibit significant similarity to known proteins and thus their function is as yet undefined. These proteins are: the Minor Capsid Protein 1 (MCP1), Minor Capsid Protein 2 (MCP2), Scaffolding-like protein (Sfp), Minor Capsid Protein 3 (MCP3) tail-activator protein (gpZ), and tail chaperone proteins (gpG and gpT), each of which will be discussed below.

The N-terminal portion of $MCP1_{TP901-1}$ and the complete $MCP1_{Tuc2009}$ protein share a very high level (94.3 %) of aa similarity. However, $MCP1_{TP901-1}$ has an approximate 28 kDa C-terminal extension, or an additional 248 aa, compared to $MCP1_{Tuc2009}$. The initial ~100 aa of this 248 aa extension of $MCP1_{TP901-1}$ appears to be due to a genetic duplication of the 5'-part of the $MCP1_{TP901-1}$-encoding gene (Supplementary Fig. 1). The remaining C-terminally located 145 amino acids of $MCP1_{TP901-1}$ are lacking in its equivalent in Tuc2009, and no significant function can be ascribed to this C-terminal region based on database searches. The genes specifying $MCP1_{TP901-1}$ and $MCP1_{Tuc2009}$, as well as SPP1 capsid protein gp7 ($gp7_{SPP1}$) all occur immediately downstream of the genes encoding their corresponding portal proteins, and have a strong positive charge (pI values of ~9.0), and are largely α-helical (data not shown). Furthermore, ClustalW alignment of $MCP1_{TP901-1}$, $MCP1_{Tuc2009}$ and $gp7_{SPP1}$ was performed and revealed aa conservation, with the exception of the C-terminal extension of $MCP1_{TP901-1}$ (Supplementary Fig. 2).

$MCP2_{TP901-1}$ and $MCP2_{Tuc2009}$, though their corresponding genes are located at equivalent positions within the reciprocal TP901-1 and Tuc2009 genomes, do not share significant homology with LALIGN using default settings. Therefore, it is unknown if these phage proteins fulfil analogous functions. HHpred analysis of the $MCP2_{Tuc2009}$ detected strong homology to the winged helix DNA-binding protein YjcQ of *B. subtilis*, *p*-value 1.1e-43. Winged helix DNA-binding proteins are part of a larger class of DNA binding helix-turn-helix proteins. While no known function is ascribed to these proteins, YjcQ homologues have previously been identified within prophage capsid-encoding regions (85).

ORF35 of TP901-1 and ORF36 of Tuc2009 are annotated in this study as specifying putative scaffolding-like proteins (Sfp) based on their localization within the capsid-encoding structural module region and their predicted secondary structure. Scaffolding proteins are involved in assisting the correct assembly of viral capsid structures by ensuring specific protein-protein interactions, nucleating the assembly within cells, and preventing/excluding non-specific interactions between proteins during the molecular construction of phage capsids (25). $Sfp_{TP901-1}$ and $Sfp_{Tuc2009}$ are predominantly α-helical and part of their structure is predicted to be intrinsically disordered (Supplementary Fig. 3), features commonly associated with phage scaffolding proteins (25).

Significant similarity between $MCP3_{TP901-1}$ and $MCP3_{Tuc2009}$ could not be detected using the default settings of LALIGN. Therefore, these minor capsid proteins are putatively assigned a common protein name solely based on their genetic synteny, although they may not fulfil analogous biological functions. Interestingly, HHpred analysis of $MCP3_{Tuc2009}$ revealed its relatedness to GP10 of a

T7-like *Podoviridae* phage infecting *Prochlorococcus*, with a *p*-value 1.5e-12, and also the major capsid protein gp5 of the well characterized phage HK97, with a *p*-value 5.8e-05. $MCP3_{Tuc2009}$ exhibits similarity to a 52 aa region located at the C-terminus of HK97 gp5, corresponding to residues 332-384 of this viral procapsid protein. Amino acid Asn356 of gp5, conserved in $MCP3_{Tuc2009}$, is involved in covalently bonding neighbouring HK97 capsid subunits via an isopeptide bond (Supplementary Fig. 4) (86). Covalent cross-linkage of capsid protein subunits has been previously observed for lactococcal phage r1t (87); however, it has not been documented to date for phages TP901-1 or Tuc2009.

The gene products of ORF40 of TP901-1 and ORF43 of Tuc2009 share 66.7 % aa similarity. Although no significant similarities to other proteins could be detected by means of database searches, the proteins are annotated in this study as tail-activator proteins (gpZ), based solely on their gene synteny with the structural module of phage λ (42). Similarly, the chaperone proteins gpG and gpT of TP901-1, while not returning significant structural similarity to characterized (phage) chaperone proteins through HHpred analysis, are expected to fulfil a similar role as phage λ gpG and gpT tail chaperone proteins based on genetic synteny with lambdoid-like phages and the presence of a conserved 'slippery' sequence that is associated with gpGT production (40, 41).

Optimizing TP901-1erm host background. Several advances were made to optimize the host background for analysing mutations made in TP901-1*erm*. A lysogenized derivative of strain *L. lactis* NZ9000 had previously been isolated (63), allowing the use of an erythromycin marker present on the TP901-1*erm* genome (68, 69). This lysogenized NZ9000 derivative offered an advantage over the original

TP901-1*erm* host, *L. lactis* 901-1*erm*, as the latter strain contains a second prophage similar in sequence to TP901-1 which may supply proteins *in trans* (Finn Vogensen, personal communication). In addition, NZ9000 contains an integrated copy of the NisRK cassette for protein expression using the nisin-inducible promotor system (88), which was used for the overproduction of RecT for recombineering mutagenesis (65). In order to further optimize NZ9000 as a background for analysing mutations in TP901-1*erm*, a BamHI restriction site was introduced into the deduced -10 promoter sequence of the lysogenic/lytic switch of prophage t712 (thereby generating strain NZ9000-Cro_{t712}), aimed at causing ablation of this promoter, thereby preventing t712 prophage induction and virion production. Induction of prophage t712 was identified as a potential problem as; (i) t712 particle production would interfere with TP901-1 phage particle study, (ii) t712 induction would increase the metabolic burden on the host when only TP901-1 induction was desired, and (iii) it was deemed possible that, due to the conserved nature of *Siphoviridae* proteins, t712 may supply TP901-1 with structural proteins being analysed during this study *in trans* (66). In agreement with the potential problems posed by prophage t712 during this study, mutagenesis of the deduced promoter sequence for t712 Cro protein caused an approximate 100-fold increase in the TP901-1*erm* titre obtained following mitomycin C induction from the NZ9000-Cro_{t712} background as compared to that obtained when strain NZ9000 was used (data not shown; indicator strain used for TP901-1*erm* titre determination was *L. lactis* 3107).

Efficiency of Plaquing of TP901-1erm Mutants. The nomenclature assigned to TP901-1*erm* structural module mutants is based on the common name assigned to the phage open reading frames (Table 1), followed by a designation signifying the insertion of a translation terminating stop codon or restriction site (e.g. $MCP1_{TP901-}$

$_{1}$::Ter or gpT$_{TP901-1}$::BamHI, respectively). The mutant that carries a mutation causing the translation fusion of ORF *gpG* and *gpT* was termed gpG*fs*T$_{TP901-1}$.

TP901-1*erm*-derived structural module mutants were induced from their lysogenic host and then tested for their ability to infect host strain *L. lactis* 3107 by spot and plaque assays to determine which genes/proteins are essential for TP901-1 phage assembly and/or infection. Of the 14 mutants created in the TP901-1*erm* structural module, only mutants carrying a non-sense mutation in the genes for MCP2$_{TP901-1}$ or MCP3$_{TP901-1}$ retained some residual infectious ability (Fig. 1). Plaque assays of three independently induced lysates of MCP2$_{TP901-1}$::Ter and MCP3$_{TP901-1}$::Ter displayed reductions in Efficiency of Plaquing (EOP) of 1.8e-02 $^{+}$/- 9.0e-04 and 3.0e-04 $^{+}$/- 4.0e-05, respectively, which are statistically significant compared to the TP901-1*erm* control (*p*-values < 0.01). Plaques formed by MCP2$_{TP901-1}$::Ter and MCP3$_{TP901-1}$::Ter were confirmed to possess the desired mutation by PCR and restriction digest using individual plaques as the template material. In addition, plaque sizes of TP901-1*erm* mutants MCP2$_{TP901-1}$::Ter and MCP3$_{TP901-1}$::Ter were visually smaller than the control wild-type phage (data not shown).

In Vitro Assembly of TP901-1erm and Frequency of Lysogeny. Previous studies of phage λ has shown that capsids and tails can be produced independently, and following purification and subsequent combining, can assemble spontaneously *in vitro* (89). In order to determine if the capsid and tail structures of lactococcal phage TP901-1 are similarly capable of *in vitro* assembly, purified capsids and tails of TP901-1 were mixed and tested for the ability to lysogenize host strain *L. lactis* 3107. Unable to establish lysogeny on their own, following the incubation of purified MHP$_{TP901-1}$::Ter and MTP$_{TP901-1}$::Ter, which donated phage tails and

capsids, respectively, infection and subsequent lysogenization of *L. lactis* 3107 resulted in erythromycin resistant colonies (Table 2). This is indicative of lactococcal phage TP901-1*erm* capsids and tails assembling *in vitro* to form fully functional DNA-injecting virions.

Sucrose gradient-purified samples of TP901-1*erm*, $MCP2_{TP901-1}$::Ter and $MCP3_{TP901-1}$::Ter phages, which represent the control and the only structural module mutants of TP901-1*erm* phage analysed in this study which retained their ability to form plaques on *L. lactis* 3107, lysogenized their host following infection and conferred erythromycin resistance as expected. The positive control, TP901-1*erm*, displayed the highest levels of lysogeny, yielding 7.51e06 $^{+}$/- 4.25e05 Erm5 cfu/ml (or approximately 1.64 % Erm-resistant lysogens relative to total cfu/ml).

By complementing TP901-1*erm* structural module mutants through the addition of either functional capsids or tails, it is possible to distinguish gene products required for functional capsid formation from those that are crucial for tail production. There is a clear distinction between TP901-1*erm* mutants that establish lysogeny at higher levels following *in vitro* assembly with either phage capsids or phage tails (shaded grey in Table 2). TP901-1*erm* derivatives carrying individual mutations targeting the portal protein-encoding gene through to the gene specifying the HTC2 protein were able to lysogenize *L. lactis* 3107 following incubation with purified capsids (donated from TP901-1*erm* MTP mutant, see Materials and Methods section) to a much higher level as compared to the mutant on its own or when mixed with purified tails. In contrast, TP901-1*erm* mutations in genes *gpZ* to *gpT* were significantly increased in their frequency of lysogeny following the addition of purified tails compared to lysogeny by the purified samples on their own

or mixed with purified capsids. Notably, mutant gpG*fs*$T_{TP901-1}$ did not give an increase in its frequency of lysogeny following the addition of purified capsids or tails.

TP901-1 and Tuc2009 Structural Protein Detection and Assembly. Polyclonal antibodies, generated against Tuc2009 structural proteins (see Materials and Methods section), were confirmed as reacting against Tuc2009 purified virions (Fig. 2, lane indicated as φ) and during an infection time-course assay (Fig. 2, lanes labelled '-10' to '70'). All detected structural proteins of Tuc2009 appeared at molecular masses consistent with their calculated molecular weight. Immunoblots with the polyclonal antibodies directed against $MCP3_{Tuc2009}$ and $gpZ_{Tuc2009}$ did not result in a detectable immunological reaction using either whole phage particles or host cell extracts obtained during infection (data not shown), while the signal from $MCP1_{Tuc2009}$ is obscured during the time-course assay by a non-specific cross reaction (i.e. present in -10 lane).

Immunoblot detection was performed on chloroform-treated PEG 8000-mediated precipitates of induced TP901-1*erm* mutants to determine the role of the different proteins during structural assembly (Fig. 3). The portal protein of TP901-1 is detectable to some degree in all TP901-1*erm* mutants, and the purified phage controls, except for the lysate obtained from the TP901-1*erm* mutant that carried a non-sense mutation in the gene encoding the portal protein (Fig. 3). This result validated the recombineering mutagenesis strategy, confirming that the insertion of an in-frame stop codon successfully prevented its production.

In addition to the absence of the portal protein in PEG-mediated precipitates of mutant $Portal_{TP901-1}$::Ter, $MCP1_{TP901-1}$ and $HTC1_{TP901-1}$ are also undetectable in

Western blots of Portal$_{TP901-1}$::Ter, while this being accompanied by a decreased amount of proteins Sfp$_{TP901-1}$ and MHP$_{TP901-1}$. As expected, MCP1$_{TP901-1}$ was not observed in Western blots using isolated virions of a TP901*erm*-derived mutant carrying an introduced translation terminating stop codon in the MCP1$_{TP901-1}$-specifying gene, while no noticeable effect was caused by this mutation in any of the additional proteins tested by Western blot. The translation terminating mutation in Sfp$_{TP901-1}$ resulted in a decreased amount of Portal$_{TP901-1,}$ and very little MHP$_{TP901-1}$ detected. The lysate of mutant MHP$_{TP901-1}$ also resulted in a decrease in the level of Portal$_{TP901-1}$ detected and a complete absence of Sfp$_{TP901-1}$. In addition to HTC1 being absent, as expected, in mutant HTC1$_{TP901-1}$::Ter, this predicted head-tail connector protein is also absent in Portal$_{TP901-1}$::Ter, Sfp$_{TP901-1}$::Ter and MHP$_{TP901-1}$::Ter.

Analysing the TP901-1*erm* mutants for the production of MTP showed that lysates of all presumed capsid mutants contained tail proteins, as expected. No MTP was detected in mutants of MTP$_{TP901-1}$, gpG$_{TP901-1}$ and gpT$_{TP901-1}$. However, MTP was detected in the TP901-1 mutant permanently fusing gpG and gpT production together (gpG*fs*T$_{TP901-1}$). In addition to detecting MTP$_{TP901-1}$, Western blots were performed to detect the distal tail protein Dit$_{TP901-1}$ mutant lysates, as it is known to be crucial for the formation of functional phage tails (47). Interestingly, Dit$_{TP901-1}$ was undetectable in all mutant virions where MTP$_{TP901-1}$ was also absent.

Electron Microscopy Analysis of TP901-1 Mutants. EM analysis of sucrose gradient-purified phage lysates of TP901-1*erm* mutants was performed to visualize the structural assemblages predicted through immunoblotting. In several instances, sucrose gradient-purification resulted in two distinct bands (Supplementary Fig. 5).

The appearance, and subsequent analysis, of the two bands obtained by sucrose gradient ultracentrifugation revealed the lower more dense band consisted of DNA-packaged capsids and mature phage virions, while the upper less dense band contained free tail structures and capsids (which may have partially DNA-filled capsids). Where pertinent, EM images obtained from either an upper or lower sucrose gradient band are indicated.

Purified fractions of Portal$_{TP901-1}$::Ter showed only tail structures produced, and no visible capsid structures (Fig. 4B). EM images of MCP1$_{TP901-1}$::Ter lower sucrose gradient band showed DNA-packaged capsids only, with no tails attached (Fig. 4C). A fully assembled phage virion was visualized for mutant MCP2$_{TP901-1}$::Ter; this was expected as this mutant phage was shown to be still capable of forming plaques on host *L. lactis* strain 3107 (Fig. 4D). Analysis of the lower sucrose gradient band of mutant HTC2$_{TP901-1}$::Ter revealed DNA-packaged capsids (Fig. 4E).

Individual DNA-packaged capsids and separate tails were visualized for the lower and upper sucrose gradient-purified fractions of mutant gpZ$_{TP90-1}$::Ter, respectively (Fig. 4F and 4G, respectively; Supplementary Fig. 5). Interestingly, EM images of gpU$_{TP901-1}$::Ter showed DNA-packaged capsids in the lower sucrose gradient fraction in addition to long polytail structures formed by aberrant tail-tube polymerization in the upper sucrose gradient fraction (Fig. 4H and 4I, respectively). Mutagenesis of TP901-1*erm* MTP resulted, as expected, in the production of tail-less capsids (Fig. 4J). Notably, EM visualization of the TP901-1*erm* carrying the gpG$_{TP901-1}$::Ter mutation revealed DNA-containing capsid and large disordered aggregates (Fig. 4K). We propose that these aggregates, not present in other phage

preparations, are hydrophobic TMP molecules incapable of remaining in solution in the absence of the chaperone protein gpG. Further EM analysis of all sucrose gradient bands of TP901-1*erm* structural module mutants is planned, which is predicted to identify the presence of full phage, detached capsid and detached phage tail structures that are predicted through Western blot and frequency of lysogeny assays (Supplementary Table 4).

DISCUSSION

Despite advances in our understanding of *Siphoviridae* virions and their role for initiating the phage infection cycle, many aspects require further investigation. This is essential to ascertain the unique features of particular phages, while generating a unified model of characteristics distinct to *Siphoviridae*.

Approximately half of the genome of TP901-1 is dedicated to creating the structural parts and efficient assembly of an infectious virion. *In vitro* assembly of TP901-1*erm* structural module mutants, assessed by monitoring the frequency of lysogeny, delineated the boundaries between the TP901-1 genetic modules associated with capsid and tail morphogenesis. The organization of the TP901-1 genome co-locates all essential capsid protein-encoding genes at the beginning of the phage structural module, while gene products required for tail production begin at *gpZ* and continue as far as the host-specificity determining receptor binding protein-encoding gene *bppL* (Fig. 1). This efficient and function-specific organization supports the modular theory of phage genome organization (90).

Previous studies of phage λ showed that capsids and tails assemble independently, and that they can be joined together *in vitro* to form functional phages (89). *In vitro* assembly of TP901-1*erm* similarly demonstrated that the MHP mutant can produce functional tail structures, while the MTP mutant formed complete DNA-packaged capsids. No additional host factors appear to be required for the assembly of phage capsids and tails, as these individual multiprotein complexes appear to be ready to assemble once formed. Lysogeny frequency of *in vitro* assembled TP901-1*erm* virions, compared to that of identically prepared wild-type TP901-1*erm* phage control, tells us that the efficiency of assembly under the tested conditions is around

2.5-3.0 % (Table 2). The assembly of TP901-1*erm* capsids and tails is remarkably efficient, however, this is assuming that the MHP mutant (donating phage tails) and the MTP mutant (donating phage capsids) behave similarly to complete TP901-1*erm* virions during purification and contribute the necessary structures at a ratio of 1:1. Therefore, by optimizing the *in vitro* assembly parameters, the achievable efficiency of phage capsid and tail assembly may even be higher.

Western blot detection of TP901-1 capsid proteins $Portal_{TP901-1}$, $Sfp_{TP901-1}$ and $MHP_{TP901-1}$ in various TP901-1*erm* mutants demonstrated that these proteins are essential for orchestrating the correct assembly and/or stabilization of capsid structures (Fig. 3). There is a pronounced lack of the $Sfp_{TP901-1}$ in mutant $MHP_{TP901-1}$::Ter, and *vice versa*, suggesting that these two proteins are closely connected during TP901-1 capsid assembly. The presence of $Sfp_{TP901-1}$ is markedly more affected than $MHP_{TP901-1}$ in virion preparations of $Portal_{TP901-1}$::Ter. This is consistent with other well-characterized phage systems where scaffolding proteins interact with the portal protein complex to initiate assembly (91). No procapsid structures were visualized following purification of $Portal_{TP901-1}$::Ter, only tail structures (Fig. 4B), possibly due to the instability of capsid structures lacking packaged DNA.

The portal protein of TP901-1 appears to act as a nucleus for capsid assembly. This is in agreement with the role of portal proteins in other viral assembly models (26-29). $MCP1_{TP901-1}$, absent in $Portal_{TP901-1}$::Ter, is a putative homologue of phage SPP1 capsid protein gp7 (see Results section). The SPP1 gp7 protein has been identified as an internal virion protein that associates with the portal protein, being present at an estimated three copies per virion, and proposed to bind

DNA in a non-specific manner to limit the rate of genome ejection at the onset of infection (92-94). $HTC1_{TP901-1}$, absent in $Portal_{TP901-1}$::Ter, $Sfp_{TP901-1}$::Ter and $MHP_{TP901-1}$::Ter, appears to be a functional analogue of SPP1 gp15; and can fit into the electron density map of the TP901-1 head-tail connector region during the pseudoatomic model construction of this phage (11, 33). Interestingly, SPP1 gp15 only associates with its cognate portal protein once DNA packaging is complete (33); therefore, the detection of $HTC1_{TP901-1}$ in $MCP1_{TP901-1}$::Ter, $MCP2_{TP901-1}$::Ter and $MCP3_{TP901-1}$::Ter suggests that these mutants all produce DNA packaged capsids. Our EM analyses of $MCP1_{TP901-1}$::Ter, showing DNA packaged capsids (Fig. 4C), and $MCP2_{TP901-1}$::Ter, displaying complete phages with a DNA-packaged capsids (Fig. 4D), are in full agreement with this notion.

Western blots aimed at detecting $Sfp_{TP901-1}$ and $Sfp_{Tuc2009}$, using either crude lysates or purified phage virions, demonstrated the presence of these putative scaffolding-like proteins at expected full-lenght molecular weights (Fig. 2, Lane φ; Fig. 3). While in crude PEG-precipitated lysates visualization of $Sfp_{TP901-1}$ may be an artefact, the occurrence and clearly visible Western blot signal of Sfp proteins in purified phage preparations indicate that the putative scaffolding-like proteins of TP901-1 and Tuc2009 are part of the mature phage virions. Remarkably, previously characterized scaffolding proteins are reported to assist in phage assembly, yet are not associated with mature virions (25, 95, 96); therefore, the presence and precise function of the putative scaffolding proteins in mature TP901-1 and Tuc2009 virions will require further investigation.

The experimental approach of PEG-precipitation of lysates containing TP901-1*erm* mutant virions coupled to immunoblotting also allowed us to putatively

eliminate possible protein interactions. For instance, the portal protein of TP901-1 is predicted to interact with both Sfp and MCP1. However, the proposed scaffolding protein mutant does not appear to significantly affect the levels of MCP1 detected, and *vice versa*. Therefore, these proteins may interact with distinct areas of the internal face of the presumed dodecameric portal complex during capsid assembly, explaining the observed results.

Polyclonal antibodies were also used to observe the intracellular production of the capsid proteins during infection. $MCP2_{Tuc2009}$ and $Sfp_{Tuc2009}$ are the first proteins detected, produced 30 minutes post-infection. This result would agree with $Sfp_{Tuc2009}$ acting as a scaffold-like protein, being produced in large quantities early in the infection cycle in order to orchestrate capsid assembly (Fig. 2). The detection of MCP2 early in the infection cycle, coupled with the observed similarity of $MCP2_{Tuc2009}$ to DNA binding proteins, suggests that this small protein may localize within the phage capsid structure.

Recent structural studies and biochemical investigations of tail chaperone proteins have provided a model of the roles of gpG and gpGT (39, 40, 97, 98). Preventing translation of TP901-1*erm* gpG or gpT (mutants $gpG_{TP901-1}$::Ter and $gpT_{TP901-1}$::BamHI, respectively) resulted in a failure to detect MTP or baseplate protein Dit. Interestingly, analysis of a mutation that mimicks the natural frame-shift occurring in gene *gpG* to produce protein gpGT (mutant gpG*fs*$T_{TP901-1}$) showed that all tail proteins tested by Western blotting are present; however, no functional phages were produced. This confirms previous observations, whereby the 'T' domain is involved in binding to the phage MTP and the specific levels of gpG and gpGT produced are essential for forming functional tail structures (40, 41).

Immunoblots detecting $MTP_{TP901-1}$ showed a significant increase in the amount of MTP observed for mutant $gpU_{TP901-1}$::Ter (Fig. 3). In agreement, EM analysis highlighted the production of long tail structures for the mutant $gpU_{TP901-1}$::Ter. This corroborates previous studies, whereby mutations in the tail-terminator protein resulted in abnormally long polytail structures (42). As stated above, gpZ of TP901-1 could not be ascribed a function through database searches or by detecting homology to other well characterized phages. The nomenclature and proposed function of TP901-1 protein gpZ was based purely on synteny to phage λ tail module. By means of EM analysis, $gpZ_{TP901-1}$ appears to have a similar role to gpZ_{λ} in tail morphogenesis, as mutations in the encoding genes do not join their capsids and tails (42).

Based on the current and previously obtained information, we present a model for the assembly and function of TP901-1 and Tuc2009 phage virion proteins (Fig. 5). Despite TP901-1 and Tuc2009 infecting the Gram-positive bacterium *L. lactis*, the assembly and function of their virion structural proteins demonstrate remarkable conservation with phages infecting divergent host organisms, indicating *Siphoviridae* have likely evolved from a common origin of efficient DNA-delivery machines.

REFERENCES

1. **Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, Rohwer F.** 2003. Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol **185:**6220-6223.
2. **Edwards RA, Rohwer F.** 2005. Viral metagenomics. Nature Rev Microbiol **3:**504-510.
3. **Letarov A, Kulikov E.** 2009. The bacteriophages in human- and animal body-associated microbial communities. J Appl Microbiol **107:**1-13.
4. **Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD.** 2011. The human gut virome: inter-individual variation and dynamic response to diet. Genome Res **21:**1616-1625.
5. **Aksyuk AA, Rossmann MG.** 2011. Bacteriophage assembly. Viruses **3:**172-203.
6. **Whitehead HR, Cox GA.** 1936. The Occurrence of Bacteriophage in Cultures of Lactic Streptococci: J Dairy Res **7**:55-62.
7. **Brussow H.** 2001. Phages of dairy bacteria. Annu Rev Microbiol **55:**283-303.
8. **Desiere F, Lucchini S, Canchaya C, Ventura M, Brussow H.** 2002. Comparative genomics of phages and prophages in lactic acid bacteria. Antonie Van Leeuwenhoek **82:**73-91.
9. **Garneau JE, Moineau S.** 2011. Bacteriophages of lactic acid bacteria and their impact on milk fermentations. Microbial Cell Factories **10 Suppl 1:**S20.
10. **Deveau H, Labrie SJ, Chopin MC, Moineau S.** 2006. Biodiversity and classification of lactococcal phages. Appl Environ Microbiol **72:**4338-4346.

11. **Bebeacua C, Lai L, Vegge CS, Brondsted L, van Heel M, Veesler D, Cambillau C.** 2013. Visualizing a complete *Siphoviridae* member by single-particle electron microscopy: the structure of lactococcal phage TP901-1. J Virol **87:**1061-1068.

12. **Spinelli S, Desmyter A, Verrips CT, de Haard HJ, Moineau S, Cambillau C.** 2006. Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nat Struct Mol Biol **13:**85-89.

13. **Spinelli S, Campanacci V, Blangy S, Moineau S, Tegoni M, Cambillau C.** 2006. Modular structure of the receptor binding proteins of *Lactococcus lactis* phages. The RBP structure of the temperate phage TP901-1. J Biol Chem **281:**14256-14262.

14. **Sciara G, Blangy S, Siponen M, Mc Grath S, van Sinderen D, Tegoni M, Cambillau C, Campanacci V.** 2008. A topological model of the baseplate of lactococcal phage Tuc2009. J Biol Chem **283:**2716-2723.

15. **Siponen M, Spinelli S, Blangy S, Moineau S, Cambillau C, Campanacci V.** 2009. Crystal structure of a chimeric receptor binding protein constructed from two lactococcal phages. J Bacteriol **191:**3220-3225.

16. **Bebeacua C, Bron P, Lai L, Vegge CS, Brondsted L, Spinelli S, Campanacci V, Veesler D, van Heel M, Cambillau C.** 2010. Structure and molecular assignment of lactococcal phage TP901-1 baseplate. J Biol Chem **285:**39079-39086.

17. **Campanacci V, Veesler D, Lichiere J, Blangy S, Sciara G, Moineau S, van Sinderen D, Bron P, Cambillau C.** 2010. Solution and electron microscopy characterization of lactococcal phage baseplates expressed in *Escherichia coli*. J Struct Biol **172:**75-84.

18. **Sciara G, Bebeacua C, Bron P, Tremblay D, Ortiz-Lombardia M, Lichiere J, van Heel M, Campanacci V, Moineau S, Cambillau C.** 2010. Structure of lactococcal phage p2 baseplate and its mechanism of activation. Proc Natl Acad Sci U S A **107:**6852-6857.

19. **Shepherd DA, Veesler D, Lichiere J, Ashcroft AE, Cambillau C.** 2011. Unraveling lactococcal phage baseplate assembly by mass spectrometry. Mol Cell Proteomics **10:**M111 009787.

20. **Veesler D, Spinelli S, Mahony J, Lichiere J, Blangy S, Bricogne G, Legrand P, Ortiz-Lombardia M, Campanacci V, van Sinderen D, Cambillau C.** 2012. Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. Proc Natl Acad Sci U S A **109:**8954-8958.

21. **Collins B, Bebeacua C, Mahony J, Blangy S, Douillard FP, Veesler D, Cambillau C, van Sinderen D.** 2013. Structure and functional analysis of the host recognition device of lactococcal phage Tuc2009. J Virol **87:**8429-8440.

22. **Mahony J, Kot W, Murphy J, Ainsworth S, Neve H, Hansen LH, Heller KJ, Sorensen SJ, Hammer K, Cambillau C, Vogensen FK, van Sinderen D.** 2013. Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. Appl Environ Microbiol **79:**4385-4392.

23. **Veesler D, Cambillau C.** 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol Mol Biol Rev **75:**423-433.

24. **Mateu MG.** 2013. Assembly, stability and dynamics of virus capsids. Arch Biochem Biophys **531:**65-79.

25. **Dokland T.** 1999. Scaffolding proteins and their role in viral assembly. Cell Mol Life Sci **56:**580-603.

26. **Bazinet C, King J.** 1988. Initiation of P22 procapsid assembly *in vivo*. J Mol Biol **202:**77-86.

27. **Guo P, Erickso S, Xu W, Olson NH, Baker TS, Anderson D.** 1991. Regulation of the phage φ29 prohead shape and size by the portal vertex. Virology **183:**366-373.

28. **Alonso JC, Tavares P, Lurz R, Trautner TA.** 2006. Bacteriophage SPP1. The Bacteriophages**:**331-349.

29. **Newcomb WW, Thomsen DR, Homa FL, Brown JC.** 2003. Assembly of the herpes simplex virus capsid: Identification of soluble scaffold-portal complexes and their role in formation of portal-containing capsids. J Virol **77:**9862-9871.

30. **Prevelige PE, Jr., Thomas D, King J.** 1993. Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. Biophys J **64:**824-835.

31. **Thuman-Commike PA, Greene B, Jakana J, Prasad BV, King J, Prevelige PE, Jr., Chiu W.** 1996. Three-dimensional structure of scaffolding-containing phage P22 procapsids by electron cryo-microscopy. J Mol Biol **260:**85-98.

32. **Johnson JE, Chiu W.** 2007. DNA packaging and delivery machines in tailed bacteriophages. Curr Opin Struct Biol **17:**237-243.

33. **Lhuillier S, Gallopin M, Gilquin B, Brasiles S, Lancelot N, Letellier G, Gilles M, Dethan G, Orlova EV, Couprie J, Tavares P, Zinn-Justin S.** 2009. Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. Proc Natl Acad Sci U S A **106:**8507-8512.

34. **Vegge CS, Neve H, Brondsted L, Heller KJ, Vogensen FK.** 2006. Analysis of the collar-whisker structure of temperate lactococcal bacteriophage TP901-1. Appl Environ Microbiol **72:**6815-6818.

35. **Plisson C, White HE, Auzat I, Zafarani A, Sao-Jose C, Lhuillier S, Tavares P, Orlova EV.** 2007. Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. EMBO J **26:**3720-3728.

36. **Bebeacua C, Fajardo L, Carlos J, Blangy S, Spinelli S, Bollmann S, Neve H, Cambillau C, Heller KJ.** 2013. X-ray structure of a superinfection exclusion lipoprotein from phage TP-J34 and identification of the tape measure protein as its target. Mol Microbiol **89**:152-165.

37. **Pedersen M, Ostergaard S, Bresciani J, Vogensen FK.** 2000. Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. Virology **276:**315-328.

38. **Katsura I, Hendrix RW.** 1984. Length determination in bacteriophage lambda tails. Cell **39:**691-698.

39. **Xu J, Hendrix RW, Duda RL.** 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol Cell **16:**11-21.

40. **Xu J, Hendrix RW, Duda RL.** 2013. Chaperone-Protein Interactions That Mediate Assembly of the Bacteriophage Lambda Tail to the Correct Length. J Mol Biol **426**:1004-1018.

41. **Xu J, Hendrix RW, Duda RL.** 2013. A Balanced Ratio of Proteins from Gene G and Frameshift-Extended Gene GT Is Required for Phage Lambda Tail Assembly. J Mol Biol **425:**3476-3487.

42. **Katsura I.** 1976. Morphogenesis of bacteriophage lambda tail. Polymorphism in the assembly of the major tail protein. J Mol Biol **107:**307-326.

43. **Katsura I, Tsugita A.** 1977. Purification and characterization of the major protein and the terminator protein of the bacteriophage lambda tail. Virology **76:**129-145.

44. **Pell LG, Liu A, Edmonds L, Donaldson LW, Howell PL, Davidson AR.** 2009. The X-ray crystal structure of the phage lambda tail terminator protein reveals the biologically relevant hexameric ring structure and demonstrates a conserved mechanism of tail termination among diverse long-tailed phages. J Mol Biol **389:**938-951.

45. **Edmonds L, Liu A, Kwan JJ, Avanessy A, Caracoglia M, Yang I, Maxwell KL, Rubenstein J, Davidson AR, Donaldson LW.** 2007. The NMR structure of the gpU tail-terminator protein from bacteriophage lambda: identification of sites contributing to Mg(II)-mediated oligomerization and biological function. J Mol Biol **365:**175-186.

46. **Chagot B, Auzat I, Gallopin M, Petitpas I, Gilquin B, Tavares P, Zinn-Justin S.** 2012. Solution structure of gp17 from the *Siphoviridae* bacteriophage SPP1: Insights into its role in virion assembly. Proteins **80:**319-326.

47. **Vegge CS, Brondsted L, Neve H, Mc Grath S, van Sinderen D, Vogensen FK.** 2005. Structural characterization and assembly of the distal tail structure of the temperate lactococcal bacteriophage TP901-1. J Bacteriol **187:**4187-4197.

48. **Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW.** 2010. GenBank. Nucleic Acids Res **38:**D46-51.

49. **Ventura M, Zomer A, Canchaya C, O'Connell-Motherway M, Kuipers O, Turroni F, Ribbera A, Foroni E, Buist G, Wegmann U, Shearman C, Gasson MJ, Fitzgerald GF, Kok J, van Sinderen D.** 2007. Comparative analyses of

prophage-like elements present in two *Lactococcus lactis* strains. Appl Environ Microbiol **73:**7771-7780.

50. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: a fast phage search tool. Nucleic Acids Res **39:**W347-352.

51. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

52. **Soding J, Biegert A, Lupas AN.** 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res **33:**W244-248.

53. **Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A.** 2010. The Pfam protein families database. Nucleic Acids Res **38:**D211-222.

54. **Soding J.** 2005. Protein homology detection by HMM-HMM comparison. Bioinformatics **21:**951-960.

55. **Huang XQ, Miller W.** 1991. A Time-Efficient, Linear-Space Local Similarity Algorithm. Adv Appl Math **12:**337-357.

56. **Thompson JD, Higgins DG, Gibson TJ.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:**4673-4680.

57. **Jones DT.** 1999. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol **292:**195-202.

58. **Cuff JA, Barton GJ.** 2000. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins **40:**502-511.

59. **Ouali M, King RD.** 2000. Cascaded multiple classifiers for secondary structure prediction. Protein Sci **9:**1162-1176.

60. **Rost B.** 2001. Review: protein secondary structure prediction continues to rise. J Struct Biol **134:**204-218.

61. **Dosztanyi Z, Csizmok V, Tompa P, Simon I.** 2005. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol **347:**827-839.

62. **Vegge CS, Vogensen FK, McGrath S, Neve H, van Sinderen D, Brondsted L.** 2006. Identification of the lower baseplate protein as the antireceptor of the temperate lactococcal bacteriophages TP901-1 and Tuc2009. J Bacteriol **188:**55-63.

63. **Stockdale SR, Mahony J, Courtin P, Chapot-Chartier MP, van Pijkeren JP, Britton RA, Neve H, Heller KJ, Aideh B, Vogensen FK, van Sinderen D.** 2013. The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for infection specialization. J Biol Chem **288:**5581-5590.

64. **van Pijkeren JP, Neoh KM, Sirias D, Findley AS, Britton RA.** 2012. Exploring optimization parameters to increase ssDNA recombineering in *Lactococcus lactis* and *Lactobacillus reuteri*. Bioengineered **3:**209-217.

65. **van Pijkeren JP, Britton RA.** 2012. High efficiency recombineering in lactic acid bacteria. Nucleic Acids Res **40:**e76.

66. **Roces C, Wegmann U, Campelo AB, Garcia P, Rodriguez A, Martinez B.** 2013. Lack of the host membrane protease FtsH hinders release of the *Lactococcus lactis* bacteriophage TP712. J Gen Virol **94**:2814-2818.

67. **Lillehaug D.** 1997. An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. J Appl Microbiol **83:**85-90.

68. **Koch B, Christiansen B, Evison T, Vogensen FK, Hammer K.** 1997. Construction of specific erythromycin resistance mutations in the temperate lactococcal bacteriophage TP901-1 and their use in studies of phage biology. Appl Environ Microbiol **63:**2439-2441.

69. **Vegge CS, Vogensen FK, Mc Grath S, Neve H, van Sinderen D, Brondsted L.** 2006. Identification of the lower baseplate protein as the antireceptor of the temperate lactococcal bacteriophages TP901-1 and Tuc2009. J Bacteriol **188:**55-63.

70. **Douillard FP, O'Connell-Motherway M, Cambillau C, van Sinderen D.** 2011. Expanding the molecular toolbox for *Lactococcus lactis*: construction of an inducible thioredoxin gene fusion expression system. Microb Cell Fact **10:**66.

71. **Mc Grath S, Neve H, Seegers JF, Eijlander R, Vegge CS, Brondsted L, Heller KJ, Fitzgerald GF, Vogensen FK, van Sinderen D.** 2006. Anatomy of a lactococcal phage tail. J Bacteriol **188:**3972-3982.

72. **Laemmli UK.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.

73. **Bradford MM.** 1976. Rapid and Sensitive Method for Quantitation of Microgram Quantities of Protein Utilizing Principle of Protein-Dye Binding. Anal Biochem **72:**248-254.

74. **Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD.** 2012. The Pfam protein families database. Nucleic Acids Res **40:**D290-301.

75. **Brüssow H, Hendrix RW.** 2002. Phage Genomics. Cell **108:**13-16.

76. **Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull GF.** 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci U S A **96:**2192-2197.

77. **Brondsted L, Ostergaard S, Pedersen M, Hammer K, Vogensen FK.** 2001. Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of lactococcal bacteriophages. Virology **283:**93-109.

78. **Seegers JF, Mc Grath S, O'Connell-Motherway M, Arendt EK, van de Guchte M, Creaven M, Fitzgerald GF, van Sinderen D.** 2004. Molecular and transcriptional analysis of the temperate lactococcal bacteriophage Tuc2009. Virology **329:**40-52.

79. **Ainsworth S, Zomer A, Mahony J, van Sinderen D.** 2013. Lytic infection of *Lactococcus lactis* by bacteriophages Tuc2009 and c2 trigger alternative transcriptional host responses. Appl Environ Microb.

80. **Kenny JG, McGrath S, Fitzgerald GF, van Sinderen D.** 2004. Bacteriophage Tuc2009 encodes a tail-associated cell wall-degrading activity. J Bacteriol **186:**3480-3491.

81. **Veesler D, Robin G, Lichiere J, Auzat I, Tavares P, Bron P, Campanacci V, Cambillau C.** 2010. Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): a baseplate hub paradigm in Gram-positive infecting phages. J Biol Chem **285:**36666-36673.

82. **Pell LG, Kanelis V, Donaldson LW, Howell PL, Davidson AR.** 2009. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. Proc Natl Acad Sci U S A **106:**4160-4165.

83. **Olia AS, Prevelige PE, Jr., Johnson JE, Cingolani G.** 2011. Three-dimensional structure of a viral genome-delivery portal vertex. Nat Struct Mol Biol **18:**597-603.

84. **White HE, Sherman MB, Brasiles S, Jacquet E, Seavers P, Tavares P, Orlova EV.** 2012. Capsid Structure and Its Stability at the Late Stages of Bacteriophage SPP1 Assembly. J Virol **86:**6768-6777.

85. **Ventura M, Canchaya C, Pridmore RD, Brussow H.** 2004. The prophages of *Lactobacillus johnsonii* NCC 533: comparative genomics and transcription analysis. Virology **320:**229-242.

86. **Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE.** 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. Science **289:**2129-2133.

87. **van Sinderen D, Karsens H, Kok J, Terpstra P, Ruiters MH, Venema G, Nauta A.** 1996. Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage r1t. Molecular microbiology **19:**1343-1355.

88. **Kuipers OP, de Ruyter PGGA, Kleerebezem M, de Vos WM.** 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. J Biotechnol **64**:15-21.

89. **Weigle J.** 1966. Assembly of phage lambda *in vitro*. Proc Natl Acad Sci U S A **55:**1462-1466.

90. **Brussow H, Canchaya C, Hardt WD.** 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev : MMBR **68:**560-602, table of contents.

91. **Rao VB, Feiss M.** 2008. The bacteriophage DNA packaging motor. Ann Rev Gen **42:**647-681.

92. **Droge A, Santos MA, Stiege AC, Alonso JC, Lurz R, Trautner TA, Tavares P.** 2000. Shape and DNA packaging activity of bacteriophage SPP1 procapsid: protein components and interactions during assembly. J Mol Biol **296:**117-132.

93. **Vinga I, Droge A, Stiege AC, Lurz R, Santos MA, Daugelavicius R, Tavares P.** 2006. The minor capsid protein gp7 of bacteriophage SPP1 is required for efficient infection of *Bacillus subtilis*. Molecular microbiology **61:**1609-1621.

94. **Stiege AC, Isidro A, Droge A, Tavares P.** 2003. Specific targeting of a DNA-binding protein to the SPP1 procapsid by interaction with the portal oligomer. Mol Microbiol **49:**1201-1212.

95. **Helgstrand C, Wikoff WR, Duda RL, Hendrix RW, Johnson JE, Liljas L.** 2003. The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 A resolution. J Mol Biol **334:**885-899.

96. **Chen DH, Baker ML, Hryc CF, DiMaio F, Jakana J, Wu W, Dougherty M, Haase-Pettingell C, Schmid MF, Jiang W, Baker D, King JA, Chiu W.** 2011. Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. Proc Natl Acad Sci U S A **108:**1355-1360.

97. **Siponen M, Sciara G, Villion M, Spinelli S, Lichiere J, Cambillau C, Moineau S, Campanacci V.** 2009. Crystal structure of ORF12 from *Lactococcus lactis* phage p2 identifies a tape measure protein chaperone. J Bacteriol **191:**728-734.

98. **Pell LG, Cumby N, Clark TE, Tuite A, Battaile KP, Edwards AM, Chirgadze NY, Davidson AR, Maxwell KL.** 2013. A conserved spiral structure for highly diverged phage tail assembly chaperones. J Mol Biol **425:**2436-2449.

99. **Braun V, Hertwig S, Neve H, Geis A, Teuber M.** 1989. Taxonomic Differentiation of Bacteriophages of *Lactococcus lactis* by Electron-Microscopy, DNA-DNA Hybridization, and Protein Profiles. J Gen Microbiol **135:**2551-2560.

100. **Christiansen B, Johnsen MG, Stenby E, Vogensen FK, Hammer K.** 1994. Characterization of the lactococcal temperate phage TP901-1 and its site-specific integration. J Bacteriol **176:**1069-1076.

101. **Arendt EK, Daly C, Fitzgerald GF, van de Guchte M.** 1994. Molecular characterization of lactococcal bacteriophage Tuc2009 and identification and analysis of genes encoding lysin, a putative holin, and two structural proteins. Appl Environ Microbiol **60:**1875-1883.

102. **Costello VA.** 1988. Characterization of bacteriophage-host interactions in *Streptococcus cremoris* UC503 and related lactic streptococci. Ph.D. Thesis. University College Cork.

103. **Ainsworth S, Zomer A, de Jager V, Bottacini F, van Hijum SA, Mahony J, van Sinderen D.** 2013. Complete Genome of *Lactococcus lactis* subsp. *cremoris* UC509.9, Host for a Model Lactococcal P335 Bacteriophage. Genome Announc **1**.

104. **Kuipers OP, Beerthuyzen MM, Siezen RJ, De Vos WM.** 1993. Characterization of the nisin gene cluster nisABTCIPR of *Lactococcus lactis*. Requirement of expression of the *nisA* and *nisI* genes for development of immunity. Eur J Biochem **216:**281-291.

105. **Koch B, Christiansen B, Evison T, Vogensen FK, Hammer K.** 1997. Construction of specific erythromycin resistance mutations in the temperate lactococcal bacteriophage TP901-1 and their use in studies of phage biology. Appl Environ Microbiol **63:**2439-2441.

106. **de Ruyter PG, Kuipers OP, de Vos WM.** 1996. Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. Appl Environ Microb **62:**3662-3667.

Table 1. Comparative analysis of TP901-1 and Tuc2009 structural module proteins. *HHpred *p*-values shown are for TP901-1 protein sequences, except for Tuc2009 ORF35, ORF40, and ORF52.

TP901-1 ORF	TP901-1 Protein Size (aa - kDa)	Tuc2009 ORF	Tuc2009 Protein Size (aa – kDa)	Assigned Common Protein Name	TP901-1 & Tuc2009 Protein Similarity (aa overlap)	HHpred Significant Hit* (Localization; *p*-value)	PDB Codes
ORF32	453 - 51.8	ORF33	446 - 51.1	**Portal protein**	71.9 % (423)	Portal protein, phage SPP1 (9.8e-70)	2jes_A
ORF33	565 - 64.3	ORF34	347 - 39.7	**Minor capsid protein 1 (MCP1)**	94.3 % (316)	-	-
ORF34	76 - 9.3	ORF35	111 - 12.8	**Minor capsid protein 2 (MCP2)**	-	YjcQ protein, *B. subtilis* (1.1e-43)	2hgc_A
ORF35	221 - 24.5	ORF36	221 - 24.5	**Scaffold-like protein (Sfp)**	96.8 % (220)	-	-
ORF36	273 - 28.7	ORF37 + 39	271 (55 + 216) - 28.6 (6.0 + 22.6)	**Major Head Protein (MHP)**	98.2 % (272)	Viral coat protein, phage SPP1 (2.1e-38)	4an5_A
ORF37	66 - 7.5	ORF40	59 - 6.5	**Minor capsid protein 3 (MCP3)**	-	GP10, T7-like capsid protein (1.5e-12)	2xd8_A
ORF38	111 - 12.8	ORF41	111 - 12.8	**Head-tail connector 1 (HTC1)**	90.9 % (110)	GP15, phage SPP1 (5.4e-24)	2kbz_A
ORF39	104 - 12.2	ORF42	103 - 11.9	**Head-tail connector 2 (HTC2)**	93.1 % (102)	GP16, phage SPP1 (8.8e-07)	2kca_A
ORF40	113 - 12.5	ORF43	109 - 12.0	**Tail activator (gpZ)**	66.7 % (108)	-	-
ORF41	130 - 14.8	ORF44	130 - 14.8	**Tail-terminator (gpU)**	99.2 % (129)	GP17, phage SPP1 (1.6e-37)	2lfp_A
ORF42	170 - 18.6	ORF45	166 - 18.2	**Major Tail Protein (MTP)**	88.2 % (169)	Major tail protein V, phage λ (1.1e-12)	2k4q_A
ORF43	112 - 12.5	ORF46	112 - 12.5	**gpG**	100 % (111)	-	-
ORF44	131 - 15.5	ORF47	131 - 15.6	**gpT**	100 % (129)	-	-

TP901-1 ORF	TP901-1 Protein Size (aa - kDa)	Tuc2009 ORF	Tuc2009 Protein Size (aa – kDa)	Assigned Common Protein Name	TP901-1 & Tuc2009 Protein Similarity (aa overlap)	HHpred Significant Hit* (Localization; *p*-value)	PDB Codes
ORF45	938 - 100.3	ORF48	1026 - 110.3	**Tape Measure Protein (TMP)**	86.1 % (1025)	-	-
ORF46	254 - 29.1	ORF49	254 - 29.1	**Distal Tail Protein (Dit)**	95.7 % (253)	ORF46 Distal tail protein, phage TP901-1 (1.1e-59)	4div_S
ORF47	919 - 102.1	ORF50	898 - 100.1	**Tail-Associated Lysin (Tal)**	95.5 % (909)	Prophage protein GP18, *L. monocytogenes* (N-terminus; 1.6e-45); Morphogenesis protein 1, phage phi-29 hydrolase (C-terminus; 3.6e-56)	3gs9_A 3csq_A
ORF48	300 - 33.8	ORF51	323 - 36.4	**Upper Baseplate Protein (BppU)**	80.7 % (296)	ORF48 baseplate protein, phage TP901-1 (9.0e-114)	4div_A
-	-	ORF52	287 - 31.9	**Basplate Associated Protein (BppA)**	-	Carbohydrate-binding module, Endo-1,4-beta-xylanase (1.3e-05)	1dyo_A
ORF49	164 - 17.2	ORF53	174 - 18.9	**Lower Baseplate Protein (BppL)**	56.5 % (62)	ORF49 baseplate protein, phage TP901-1 (1.6e-88)	3s8m_A
ORF50	75 - 8.7	ORF54	75 - 8.7	**Hypothetical protein**	95.9 % (74)	-	-
ORF51	668 - 71.6	ORF55	677 - 73.3	**Neck Passage Structure (NPS)**	71.9 % (690)	ORF48 immunoglobulin fold, phage TP901-1 (N-terminus; 3.1e-09); Endoglucanase C, carbohydrate-binding module (C-terminus; 1.7e-4)	3uh8_A 1gu3_A

Table 2. Lysogeny by TP901-1*erm* and structural module purified mutants mixed with phage capsids and tails. The lysogeny of purified samples resulting in the highest numbers of erythromycin resistant *L. lactis* 3107 colonies, either by purified sample on their own or mixed with phage capsids or tails, is highlighted in grey. The lack of erythromycin resistant colonies arising following lysogeny by gpG*fs*T$_{TP901-1}$ on its own or mixed with capsids and tails suggests tail tube polymerization, but not tail assembly to capsids, is affected.

Sample	Lysogeny by Purified Sample[1]		
	Purified Sample	Mixed with Phage Capsids	Mixed with Phage Tails
TP901-1*erm*	7.51e06 +/- 4.25e05	NT	NT
Portal$_{TP901-1}$::Ter	0.00	5.90e04 +/- 4.36e03	0.00
MCP1$_{TP901-1}$::Ter	2.70e02 +/- 3.00e01	7.80e02 +/- 8.72e01	1.33e02 +/- 1.53e01
MCP2$_{TP901-1}$::Ter	6.52e06 +/-3.07e05	NT	NT
Sfp$_{TP901-1}$::Ter	6.00e01 +/- 6.93e01	5.43e04 +/- 1.57e04	5.33e01 +/- 2.31e01
MHP$_{TP901-1}$::Ter	0.00	2.30e05 +/- 5.01e04	0.00
MCP3$_{TP901-1}$::Ter	5.05e06 +/- 9.1e05	NT	NT
HTC1$_{TP901-1}$::Ter	2.62e04 +/- 1.88e03	9.53e04 +/- 2.48e04	1.38e04 +/- 2.75e03
HTC2$_{TP901-1}$::Ter	7.33e01 +/- 5.51e01	1.43e05 +/- 2.55e04	2.00e01 +/- 2.00e01
gpZ$_{TP901-1}$::Ter	4.87e02 +/- 1.02e02	1.97e02 +/- 4.73e01	2.46e05 +/- 2.01e04
gpU$_{TP901-1}$::Ter	0.00	0.00	2.78e05 +/- 2.96e04
MTP$_{TP901-1}$::Ter	0.00	0.00	2.01e05 +/- 2.98e04
gpG$_{TP901-1}$::Ter	0.00	0.00	2.54e05 +/- 3.80e04
gpT$_{TP901-1}$::BamHI	0.00	0.00	1.30e05 +/- 2.41e04
gpG*fs*T$_{TP901-1}$	0.00	0.00	0.00

[1] *L. lactis* 3107 Erm5 cfu/ml

NT = Not Tested

Supplementary Table 1. Bacterial strains, phages and plasmids used in this study.

Strains, phages and plasmids	Relevant features	Source
L. lactis		
NZ9000	MG1363 *pepN::nisRK*	(88)
NZ9000-Cro_{t712}	NZ9000 with point mutations inserting a BamHI site into prophage t712 anti-repressor (Cro) promotor sequence; contains plasmid pJP005	This study
NZ9000-Cro_{t712}-TP901-1*erm*	NZ9000-Cro_{t712} displaying $t712^-$ phenotype lysogenized with TP901-1*erm* prophage; contains plasmid pJP005	This study
901-1*erm*	Lysogenic host and source of TP901-1*erm* prophage	(99)
3107	Lytic host for phage TP901-1*erm*	(100)
UC509	Lysogenic host and source of Tuc2009 prophage	(101)
UC509.9	Lytic host for phage Tuc2009	(102, 103)
NZ9700	Nisin over-producing strain	(104)
NZ9000-Cro_{t712}-$Portal_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* portal protein (ORF32)	This study
NZ9000-Cro_{t712}-$MCP1_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* MCP1 (ORF33)	This study
NZ9000-Cro_{t712}-$MCP2_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* MCP2 (ORF34)	This study
NZ9000-Cro_{t712}-$Sfp_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* Sfp (ORF35)	This study
NZ9000-Cro_{t712}-$MHP_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* MHP (ORF36)	This study
NZ9000-Cro_{t712}-$MCP3_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* MCP3 (ORF37)	This study
NZ9000-Cro_{t712}-$HTC1_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* HTC1 (ORF38)	This study
NZ9000-Cro_{t712}-$HTC2_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* HTC2 (ORF39)	This study
NZ9000-Cro_{t712}-$gpZ_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* gpZ (ORF40)	This study
NZ9000-Cro_{t712}-$gpU_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* gpU (ORF41)	This study
NZ9000-Cro_{t712}-$MTP_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* MTP (ORF42)	This study
NZ9000-Cro_{t712}-$gpG_{TP901-1}$::Ter	Translation termination codon insertion in TP901-1*erm* gpG (ORF43)	This study
NZ9000-Cro_{t712}-$gpT_{TP901-1}$::BamHI	Insertion mutation in TP901-1*erm* gpG (ORF43) disrupting gpT (ORF44) production	This study
NZ9000-Cro_{t712}-gpG*fs*$T_{TP901-1}$	Insertion and nucleotide altering mutation in TP901-1*erm* gpG (ORF43) causing a frameshift and permanently fusing it in frame to gpT (ORF44)	This study
Phages		
TP901-1*erm*	Temperate P335 species phage infecting 3107; contains erythromycin marker	(69, 99, 105)
Tuc2009	Temperate P335 species phage infecting UC509.9	(102)
Plasmids		
pNZ8048	Nisin-inducible protein expression vector	(106)
pTX8049	pNZ8048 derivative expression vector; recombinant N-terminal thioredoxin fusion	(70)
pJP005	pNZ8048 derivative expressing RecT protein required for recombineering mutagenesis	(65)
pNZ8048-$Portal_{Tuc2009}$	Tuc2009 Portal protein (ORF33) expression construct; Genbank coordinates 16486-17847	This study
pNZ8048-$MCP1_{Tuc2009}$	Tuc2009 MCP1 (ORF34) expression construct; Genbank coordinates 17844-18884	This study
pTX8049-$MCP2_{Tuc2009}$	Tuc2009 MCP2 (ORF35) expression construct; Genbank coordinates 18930-19262	This study
pNZ8048-$Sfp_{Tuc2009}$	Tuc2009 Sfp (ORF36) expression construct; Genbank coordinates 19388-20050	This study
pTX8049-$MHP_{Tuc2009}$	Tuc2009 MHP (ORF37-39) expression construct; Genbank coordinates 20052-22183	This study
pTX8049-$MCP3_{Tuc2009}$	Tuc2009 MCP3 (ORF40) expression construct; Genbank coordinates 22183-22383	This study
pTX8049-$HTC1_{Tuc2009}$	Tuc2009 HTC1 (ORF41) expression construct; Genbank coordinates 22367-22699	This study
pTX8049-$gpZ_{Tuc2009}$	Tuc2009 gpZ (ORF43) expression construct; Genbank coordinates 23004-23330	This study

Supplementary Table 2. Detailed description of mutations created in the various strains during this study, and their effects.

Mutant strain and phages	Accession numbers, ORF numbers and genomic coordinates of mutation region	Original DNA sequence	Original protein sequence[1, 2]	DNA sequence following recombineering[3]	Protein sequence following recombineering
***L. lactis* NZ9000**	**NC_017949**				
NZ9000-Cro_{t712}	782805..782833	TTGACAAGTTGCAACGA CGATGTTATAAT	N/A	TTGACAAGTTGCAACGA CGATCAGGATCC	N/A
TP901-1	**NC_002747**				
NZ9000-Cro_{t712}-$Portal_{TP901-1}$::Ter	ORF32; 15816..15839	AAACAAGAGCCTTTATT TGCCGTG	KQEPLFAV	AAACAAGAGCCTTGAA TTCTATTTGCCGTG	KQEP*ILFAV
NZ9000-Cro_{t712}-$MCP1_{TP901-1}$::Ter	ORF33; 17151..17174	GGAATTCTTGGTTTAAG CGTTCCA	GILGLSVP	GGAATTCTTGGTTGAAT TCATAAGCGTTCCA	GILG*IHKRS
NZ9000-Cro_{t712}-$MCP2_{TP901-1}$::Ter	ORF34; 18500..18523	TATCATTACTATTTTGA CGGTGAT	YHYYFDGD	TATCATTACTATTGAAT TCATTGACGGTGAT	YHYY*IH*R*
NZ9000-Cro_{t712}-$Sfp_{TP901-1}$::Ter	ORF35; 19028..19051	GTTAAAGAAAAATCTG ACGAAGAA	VKEKSDEE	GTTAAAGAAAAATGAA TTCCTGACGAAGAA	VKEK*IPDEE
NZ9000-Cro_{t712}-$MHP_{TP901-1}$::Ter	ORF36; 19812..19835	GCTAAGACTACCTCTCA AACTGTT	AKTTSQTV	GCTAAGACTACCTGAAT TCGCTCAAACTGTT	AKTT*IRSNC
NZ9000-Cro_{t712}-$MCP3_{TP901-1}$::Ter	ORF37; 20318..20341	GATGAACTAACGACGCT TACCGTT	DELTTLTV	GATGAACTAACGTGAAT TCACGCTTACCGTT	DELT*IHAYR
NZ9000-Cro_{t712}-$HTC1_{TP901-1}$::Ter	ORF38; 20526..20549	CTTATTCTTGGTTCTGA CATTAAA	LILGSDIK	CTTATTCTTGGTTGAAT TCCTGACATTAAA	LILG*IPDIK
NZ9000-Cro_{t712}-$HTC2_{TP901-1}$::Ter	ORF39; 20900..20923	AGTGTAGAAGTTTTGG AGATATT	SVEVFGDI	AGTGTAGAAGTTTAGAA TTCTTGGAGATATT	SVEV*NSWRY
NZ9000-Cro_{t712}-$gpZ_{TP901-1}$::Ter	ORF40; 21107..21130	CTTGTAAAGCATTTGGA TAAAGCA	LVKHLDKA	CTTGTAAAGCATTGAAT TCTGGATAAAGCA	LVKH*ILDKA
NZ9000-Cro_{t712}-$gpU_{TP901-1}$::Ter	ORF41; 21444..21467	TTCAAACGAATCCAAGC TTTGGGG	FKRIQALG	TTCAAACGATGATGAGC TTTGGGG	FKR**ALG
NZ9000-Cro_{t712}-$MTP_{TP901-1}$::Ter	ORF42; 22061..22084	ATGCAGCAATTATTGAA GTGTGGG	MQQLLKCG	ATGCAGCAATTATGAAT TCATGAAGTGTGGG	MQQL*IHEVW
NZ9000-Cro_{t712}-$gpG_{TP901-1}$::Ter	ORF43; 22579..22602	AACATTGCAACTTTATC GAATGTA	NIATLSNV	AACATTGCAACTTGAAT TCTATCGAATGTA	NIAT*ILSNV
NZ9000-Cro_{t712}-$gpT_{TP901-1}$::BamHI	ORF44; 22731..22737	GAAAGCAATACGGGAA AGTTAATC	ESNTGKLI..[6 aa]..*	GAAAGCAATACGGGATC CGGAAAGTTAATC	ESNTGSGKLI..[6 aa]..*
NZ9000-Cro_{t712}-gpG*fs*$T_{TP901-1}$	ORF43-44; 22731..22737	GAAAGCAATACGGGAA AGTTAATC	ESNTGKLI..[6 aa]..*	GAAAGCAATACTGGTA AAGCTGATC	ESNTGKVN..[126 aa]..*

1 The original protein sequence of NZ9000-Cro_{t712}-$gpT_{TP901-1}$::Ter mutant shows only the amino acid sequence of gpG. Protein sequence of gpGT (produced through a -1 frame-shift in *gpG* DNA sequence GGGAAAG, fusing it to *gpT*) is not shown.

2 The original protein sequence of NZ9000-Cro_{t712}-gpG*fs*$T_{TP901-1}$ mutant shows only the amino acid sequence of gpG.

3	Introduced DNA bases causing mutations are underlined once, while introduced bases resulting in a novel restriction site are double underlined.
N/A	Not applicable
*	Translation terminating stop codon

Supplementary Table 3. Oligonucleotides used in this study.

Oligonucleotide name	Oligonucleotide sequence (5' → 3')
Recombineering primers[ψ]	
NZ9000-Cro$_{t712}$::BamHI	C*T*C*C*A*TGATTGTTCGCTTGTTTGACTTTATGAGTTAGGATCCTGATCGTCGTTGCAACTTGTCAACCATTAAGGTTGTTTT
Portal$_{TP901-1}$::Ter	T*T*T*T*T*GTCCTCGTCAACACCATATCTCACGGCAAATAGAATTCAAGGCTCTTGTTTGACTGTATCGTCATAGACCATAAA
MCP1$_{TP901-1}$::Ter	T*C*A*A*C*TAAACTGTTATATCCTTCCTTTGGAACGCTTATGAATTCAACCAAGAATTCCGGCTTGTCTTTCAAATTCTGTGAG
MCP2$_{TP901-1}$::Ter	G*C*T*A*T*AATTTCTTCAACAACATTTCTATCACCGTCAATGAATTCAATAGTAATGATAGAATTTATTTTCTTTTTTAAAGATGA
Sfp$_{TP901-1}$::Ter	T*T*A*G*C*TTTTTGAAGTTCGGCAGCTTTTTCTTCGTCAGGAATTCATTTTTCTTTAACTGACTTTTTGCCACCTTTTTCAAG
MHP$_{TP901-1}$::Ter	C*C*C*C*G*TCAACGTTTGCTTTAGTAGAAACAGTTTGAGCGAATTCAGGTAGTCTTAGCTGCGCTCAATAAGTCGTCATCGA
MCP3$_{TP901-1}$::Ter	T*A*G*T*T*TCAAGAAGCTCTTTTAGCTGATTAACGGTAAGCGTGAATTCACGTTAGTTCATCATTTTTCACTTGCTTGGTCGCATTTATG
HTC1$_{TP901-1}$::Ter	A*C*A*T*A*TTCTAGTTCTGGCGGTACTTCTTTAATGTCAGGAATTCAACCAAGAATAAGCAATAAACGGTCACGAGTGCGTTT
HTC2$_{TP901-1}$::Ter	T*C*A*T*T*CGCATGACTTTTGCCCCTTTTTTAATATCTCCAAAGAATTCAAACTTCTACACTTCTGTCAGTTCCAATATCAGTGATATTT
gpZ$_{TP901-1}$:: Ter	G*T*T*G*A*ACACCCTTTAAAGATGCTGCTTTATCCAGAATTCAATGCTTTACAAGTTGGTCAATCCCTTTAAAAGATAAGCTA
gpU$_{TP901-1}$::Ter	T*C*A*T*T*GGCTTATAATCATAAACGGTATACCCCAAAGCTCATCATCGTTTGAACAATTCGTCAAAAATAGATTGGTCTCGAG
MTP$_{TP901-1}$::Ter	A*G*T*T*G*CTTTTTCTGCTTTATCAATTTCCCACACTTCATGAATTCATAATTGCTGCATCATCAAACGCTTTGTCCATTTCGT
gpG$_{TP901-1}$::Ter	G*T*T*T*C*TGTTCGATTTCCTAAAAATAATACATTCGATAGAATTCAAGTTGCAATGTTAGCCATTTCTAGTTCAGGAATGAT
gpT$_{TP901-1}$::BamHI	T*C*A*G*C*TATTTGGTCATTTTTGCTTTGATTAACTTTCCGGATCCCGTATTGCTTTCAGTAATTTCTTTCAAAACATCATCA
gpG*fs*T$_{TP901-1}$	C*A*A*A*C*TTTTCAGCTATTTGGTCATTTTTGCTTTGATCAGCTTTACCAGTATTGCTTTCAGTAATTTCTTTCAAAACATCATCA
Screening primers[¥]	
Cro-t712-Fw	CTTCAAGATATTCATCGCTCAGACGG
Cro-t712-MidFw	TTGCAACGACGATCAGGATCC
Cro-t712-Rv	TCGTAATCATCAAGCGCATGTTCC
Scrnseq-orf32-Fw	GCAATGACAACTTATCAGCGTGA
Scrn-orf32-MidRv	CACCATATCTCACGGCAAATAGAATTC
Seq-orf32-Rv	ACTGGTAAATCTGGATATGGGTTG
Scrnseq-orf33-Fw	GAGAAAGCTTGGCAAGAGCA
Scrn-orf33-MidRv	CTTCCTTTGGAACGCTTATGAATTC
Seq-orf33-Rv	ATGCATGTTTGGTGCATTAATACCA
Scrnseq-orf34-Fw	CATTCAGTTTTAGGTTATGGGAACAT
Scrn-orf34-MidRv	AACATTTCTATCACCGTCAATGAATTC
Seq-orf34-Rv	TTCAGCCTGCTCTTTACCAA
Scrnseq-orf35-Fw	GACGTGAAGAGAATCATTGTACCAGA
Scrn-orf35-MidRv	GCAGCTTTTTCTTCGTCAGGAATTC

Seq-orf35-Rv	TTCATCAGCTTGACTGGTAATGT
Scrnseq-orf36-Fw	GAATTGGTAAAGAGCAGGCTGA
Scrn-orf36-MidFw	GCTTTAGTAGAAACAGTTTGAGGAATTCG
Seq-orf36-Rv	GGCTTGTGCATCCTCATCAT
Scrnseq-orf37-Fw	ACTGTTTCTACTAAAGCAAACGTTG
Scrn-orf37-MidRv	TGATTAACGGTAAGCGTGAATTCA
Seq-orf37-Rv	AACAGTTCTGGTAGGCTCCTT
Scrnseq-orf38-Fw	ACGGAACTTACGCTGATGT
Scrn-orf38-MidRv	CGGTACTTCTTTAATGTCAGGAATTC
Seq-orf38-Rv	CTTCTTGTGAGTAGGACTGCAT
Scrnseq-orf39-Fw	TTGCTTATGGCTATCACTGATGA
Scrn-orf39-MidRv	CCCTTTTTTAATATCTCCAAAGAATTCA
Seq-orf39-Rv	GGATTCGTTTGAACAATTCGTCA
Scrnseq-orf40-Fw	GTAATCGAAAAACGCACTCGTG
Scrn-orf40-MidRv	CTTTAAAGATGCTGCTTTATCCAGAATTC
Seq-orf40-Rv	ACAAAAGGTTGAGCAGATTGA
Scrnseq-orf41-Fw	TGATTTAGGCGAATGGGTTGA
Scrn-orf41-MidRv	GGTATACCCCAAAGCTCATCA
Seq-orf41-Rv	TGTATTTGTTGTTGTATCGTCCA
Scrnseq-orf42-Fw	CACTTCAAATATGACAGCGAATATGCAG
Scrn-orf42-MidFw	TTGATGATGCAGCAATTATGAATTCA
Seq-orf42-Rv	CAATGTTAGCCATTTCTAGTTCAGGA
Scrnseq-orf43-Fw	GGAAACTTGCATTCCAAACAGA
Scrn-orf43-MidRv	CGATTTCCTAAAAATAATACATTCGATAGAATTC
Seq-orf43-Rv	TTCGCATTCATCAATAAAATCATCAA
Scrnseq-orf44-Fw	TCTCTTATGAACCTAACTCAGAAGATGCG
Scrn-orf44-MidRv	CATTTTTGCTTTGATTAACTTTCCGGATCC
Seq-orf44-Rv	AAGTTTTGAACTTGTTTAGCAGCGTC
Scrnseq-orf43-44-Fw	ACTTGCTGTTAAAATCATTCCTGAACT
Scrn-orf43-44-MidRv	ACTGAAAGCAATACTGGTAAAGCTG
Seq-orf43-44-Rv	TTCTCTGGGCTTTCGATTCCTA

Recombinant protein primers[♮,#]

pNZ-orf33Tuc-Fw	AGCAGCCCATGGTGAATTTGAAACCAATAAA
pNZ-orf33Tuc-Rv	AGCAGCACTAGTTTA**ATGGTGATGGTGATGGTG**CCCCTCATTTGTTTCATTCTC
pNZ-orf34Tuc-Fw	AGCAGCCCATGGATTCATCTGATTACTGGAG

pNZ-orf34Tuc-Rv	AGCAGCACTAGTTTA**GTGATGGTGATGGTGATG**TCCTTTTCTTTTATTGC
pTX-orf35Tuc-Fw	AGCAGCGGATCCATGGCTAAAGATGATTTCTT
pTX-orf35Tuc-Rv	AGCAGCACTAGTTTA**GTGATGGTGATGGTGATG**AGTTCCTGGTATCCAGTCTTTT
pNZ-orf36Tuc-Fw	AGCAGCCCATGGAACAAACAGAACTTTT
pNZ-orf36Tuc-Rv	AGCAGCACTAGTTTA**GTGATGGTGATGGTGATG**ATAGCCTCCTGTTAATTTT
pTX-orf37Tuc-Fw	AGCAGCCCATGGCAAAACAAAAAACAACACTTAC
pTX-orf39Tuc-Rv	AGCAGCACTAGTTTA**GTGATGGTGATGGTGATG**TACACCCGTAAATGTGATATTA
pTX-orf40Tuc-Fw	AGCAGCGGATCCATGGGGCGGCTACTAAGTCG
pTX-orf40Tuc-Rv	AGCAGCACTAGTTTA**GTGATGGTGATGGTGATG**TAAGTGATAGCCATAAGCAA
pTX-orf41Tuc-Fw	AGCAGCGGATCCATGGCTATCACTTATGAAAT
pTX-orf41Tuc-Rv	AGCAGCACTAGTTTA**GTGATGGTGATGGTGATG**ATACAATCTGAACCTCCCTATCTTC
pTX-orf43Tuc-Fw	AGCAGCGGATCCATGAAAATAACTGGAATTGA
pTX-orf43Tuc-Rv	AGCAGCACTAGTTTA**GTGATGGTGATGGTGATG**TTTCGTCAACCTTTCTAA

* Phosphorothioate linkages of recombineering oligos.

Ψ Nucleotides of recombineering oligos which are designed to introduce or alter the base sequence of TP901-1*erm* are underlined once, while introduced restriction enzyme sites are double underlined.

¥ Nucleotides of screening oligos which bind to mutated bases are underlined.

ǂ Restriction enzymes sites of recombinant protein construct oligos are double underlined.

\# Hexahistidine tags are indicated in bold.

Supplementary Table 4. Predicted and observed phenotypic properties of TP901-1*erm* structural module mutants. A selection of electron micrographs are available which confirm predicted structural assemblies formed by the specific mutants (Fig. 4).

Sample analysed	Predicted phenotypes of TP901-1*erm* structural mutants[1]	Brief description for predicted phenotypes of TP901-1*erm* mutants; based on literature and, where required, experimental work	EM observed phenotype[2]
TP901-1*erm*	N/A	N/A	Complete wild-type TP901-1 phage
$Portal_{TP901-1}$::Ter	Tail structures	Portal proteins are important for capsid morphogenesis (5)	Tail structures (Fig. 4B)
$MCP1_{TP901-1}$::Ter	Unknown	SPP1 homologue has a role in capsid DNA packaging and/or ejection (93)	LSGB - Capsids only, no attached tails (Fig. 4C)
$MCP2_{TP901-1}$::Ter	Wild-type resembling phage	Minor capsid protein of TP901-1 and Tuc2009, mutant still forms plaques	Complete phage (Fig. 4D)
$Sfp_{TP901-1}$::Ter	Tail structures	Scaffolding-like proteins are important for capsid morphogenesis (5)	To be completed
$MHP_{TP901-1}$::Ter	Tail structures	Major capsid proteins are important for capsid morphogenesis (5)	To be completed
$MCP3_{TP901-1}$::Ter	Wild-type resembling phage	Minor capsid protein of TP901-1 and Tuc2009, mutant still forms plaques	To be completed
$HTC1_{TP901-1}$::Ter	Detached capsid and tail structures	Head-tail connector proteins required as an attachment site for phage tails (33)	To be completed
$HTC2_{TP901-1}$::Ter	Detached capsid and tail structures	Head-tail connector proteins required as an attachment site for phage tails (33)	LSGB - Capsids only, no attached tails (Fig. 4E)
$gpZ_{TP901-1}$::Ter	Detached capsid and tail structures	Important for phage λ capsid-tail attachment (42)	LSGB & USGB - Detached capsid and tail structures (Fig. 4F & G)
$gpU_{TP901-1}$::Ter	Capsids and polytail structures	Important for preventing aberrant MTP polymerization (42)	LSGB & USGB - Capsids and polytails (Fig. 4H & I)
$MTP_{TP901-1}$::Ter	Capsids	Major tail proteins are essential for tail morphogenesis (89)	Capsids (Fig. 4J)
$gpG_{TP901-1}$::Ter	Capsids	Chaperone gpG required to maintain hydrophobic TMP in solution (40, 41)	Capsids and possibly insoluble TMP (Fig. 4K)
$gpT_{TP901-1}$::BamHI	Capsids	Chaperone gpGT require to recruit MTP to TMP for tail assembly (40, 41)	To be completed
$gpG\text{-}sT_{TP901-1}$	Unknown	Ratio of gpG and gpGT required for tail tube polymerization (40, 41)	To be completed

[1] The predicted phenotypes of TP901-1*erm* structural module mutants are based on their ability to form plaques, Western blotting data, frequency of lysogeny following *in vitro* assembly and conservation of protein function amongst *Siphoviridae*.

N/A Not applicable

[2] Lower sucrose gradient band (LSGB), upper sucrose gradient band (USGB)

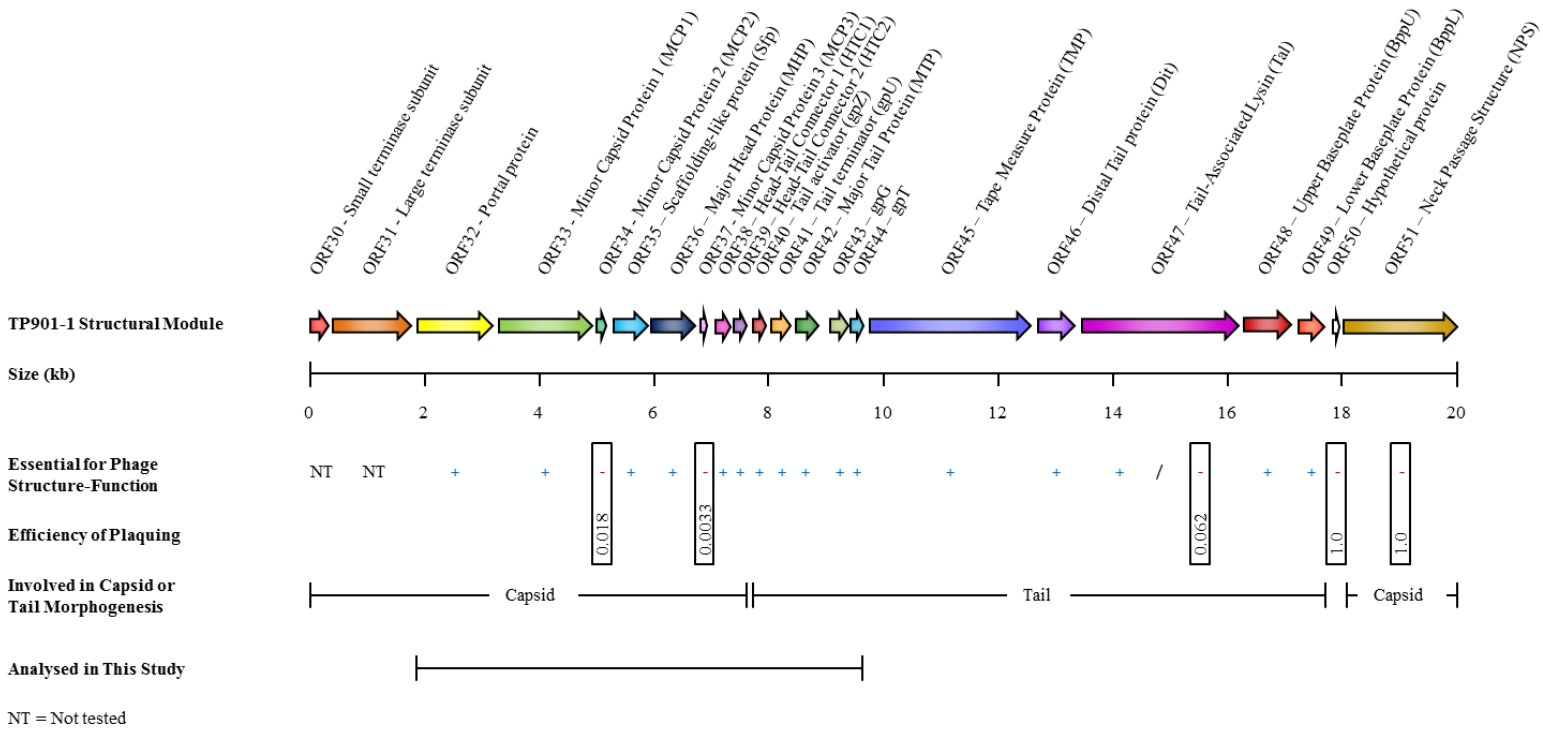


Figure 1. Diagrammatic representation of the 22 annotated structural module genes of phage TP901-1. Mutagenesis of TP901-1 structural module has determined gene products essential (+) or non-essential (-) for creating infectious phage virions. The efficiency of plaquing is displayed for mutants of non-essential gene products. The genomic regions encoding capsid and tail morphogenesis proteins are depicted.

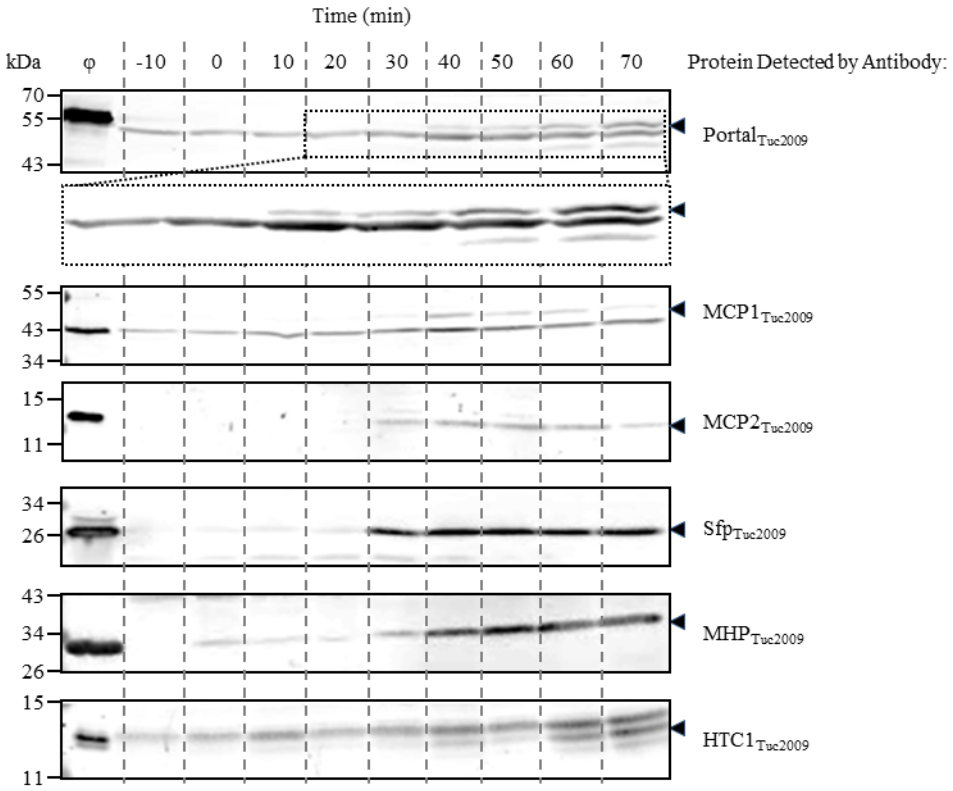


Figure 2. Western blot time-course detection of Tuc2009 capsid and head-tail connector proteins produced following infection of *L. lactis* UC509.9. A control lane containing purified phage virions, φ, shows the protein sizes as they are part of the mature virion. The protein detected by the primary polyclonal antibody is outlined to the right of the figure, with black-filled triangles highlighting the exact protein of interest.

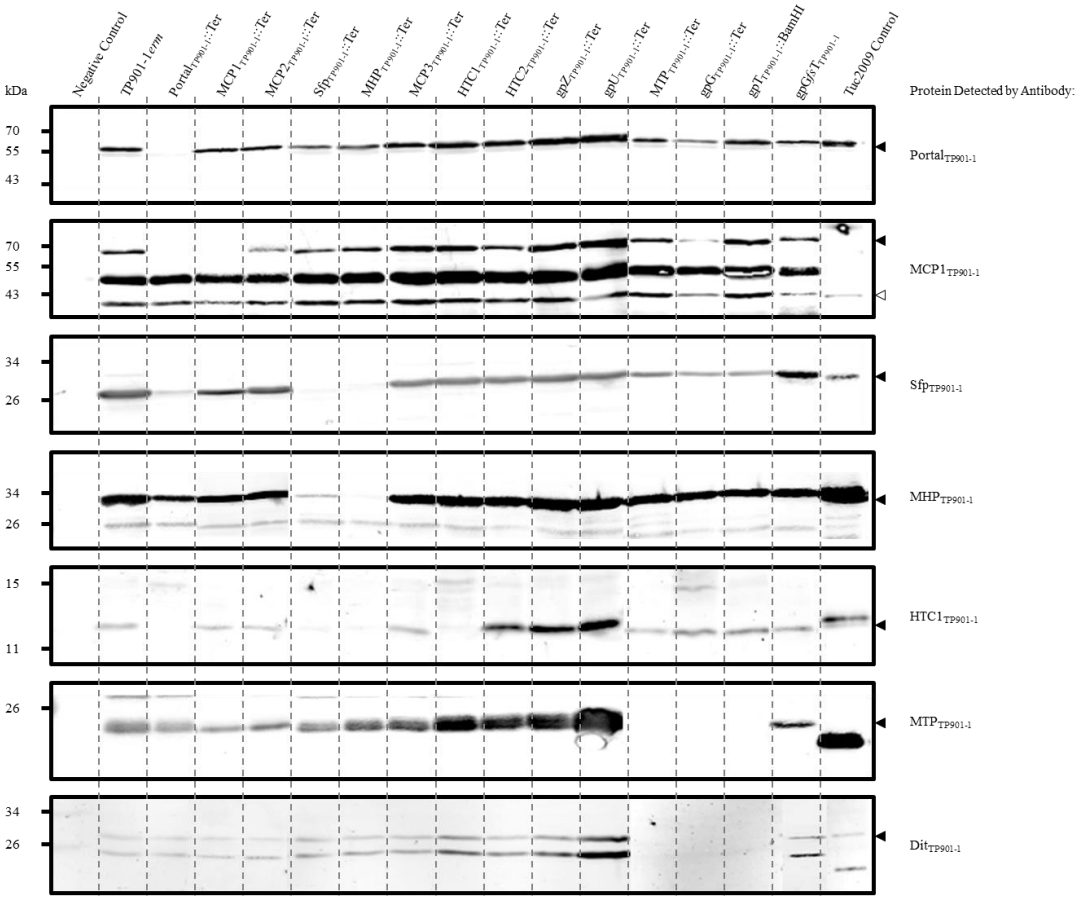
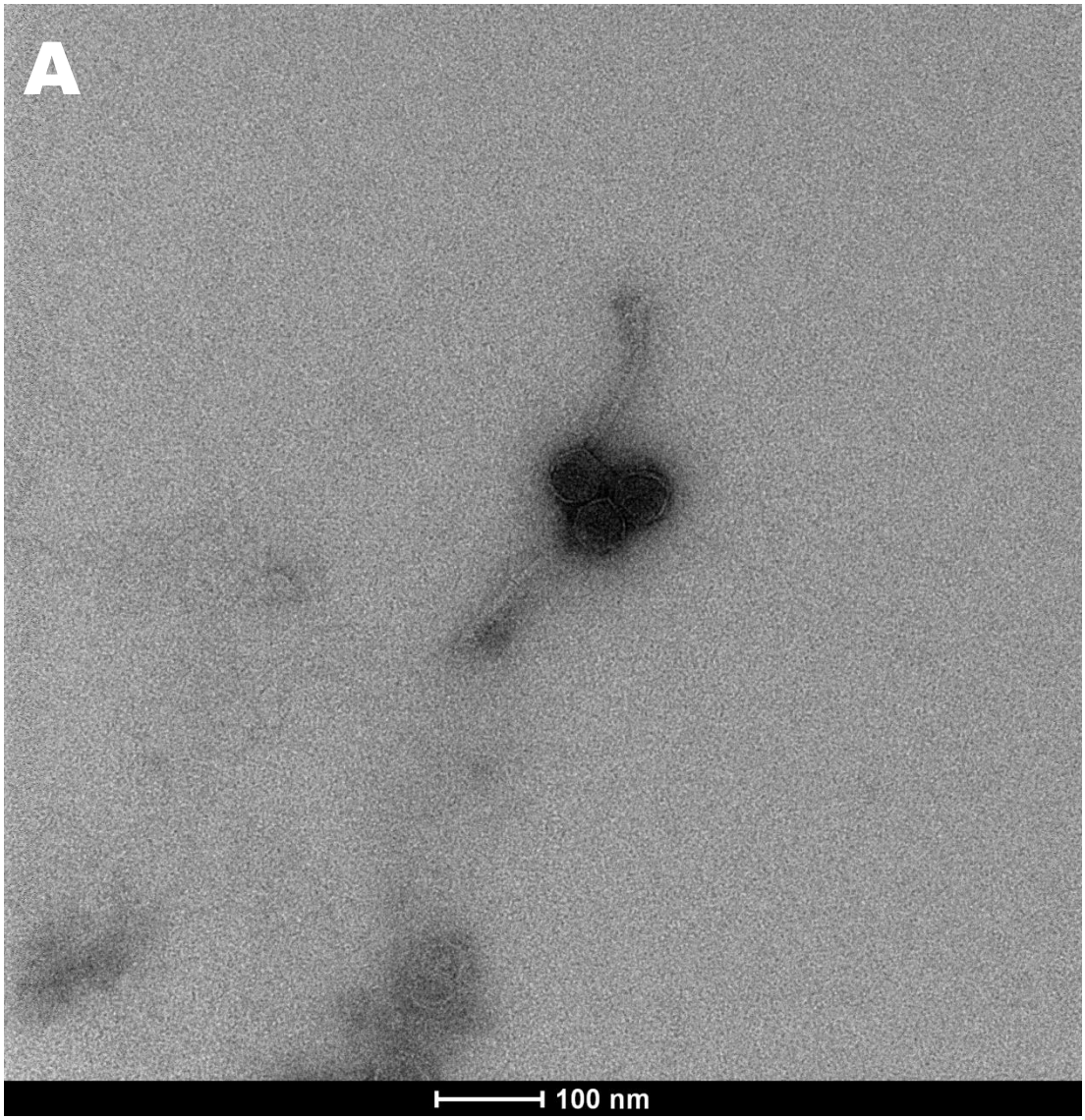
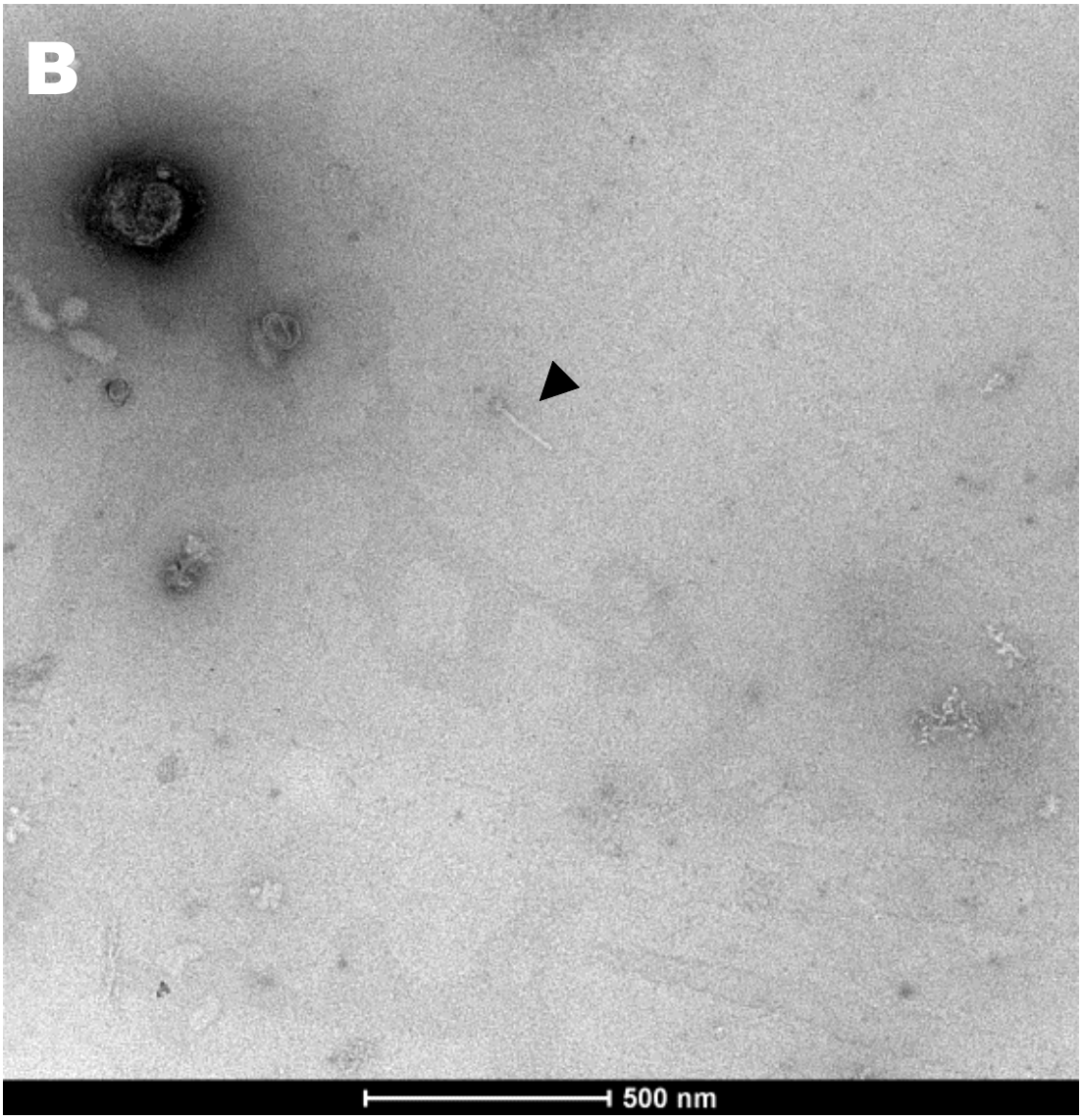
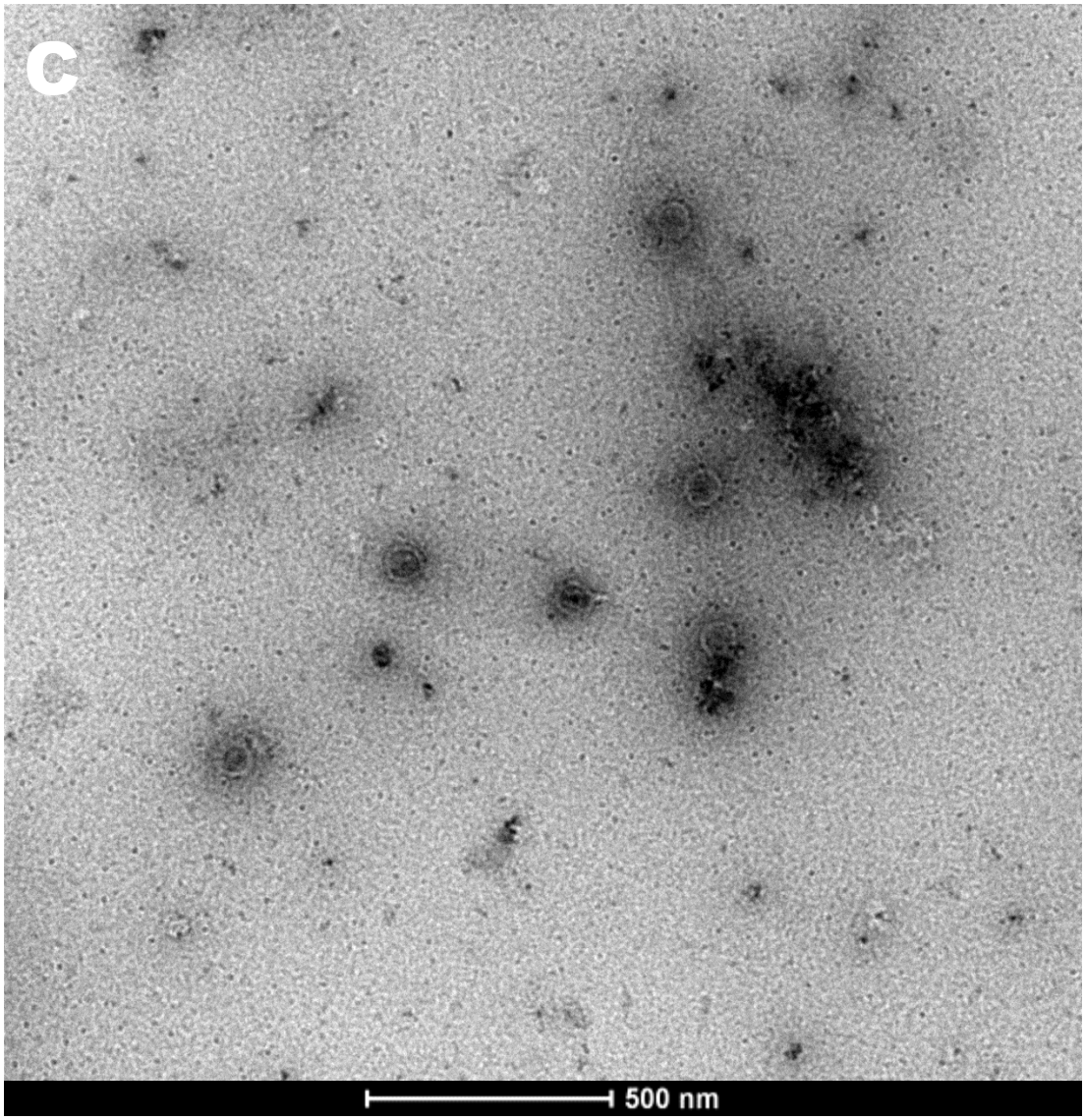
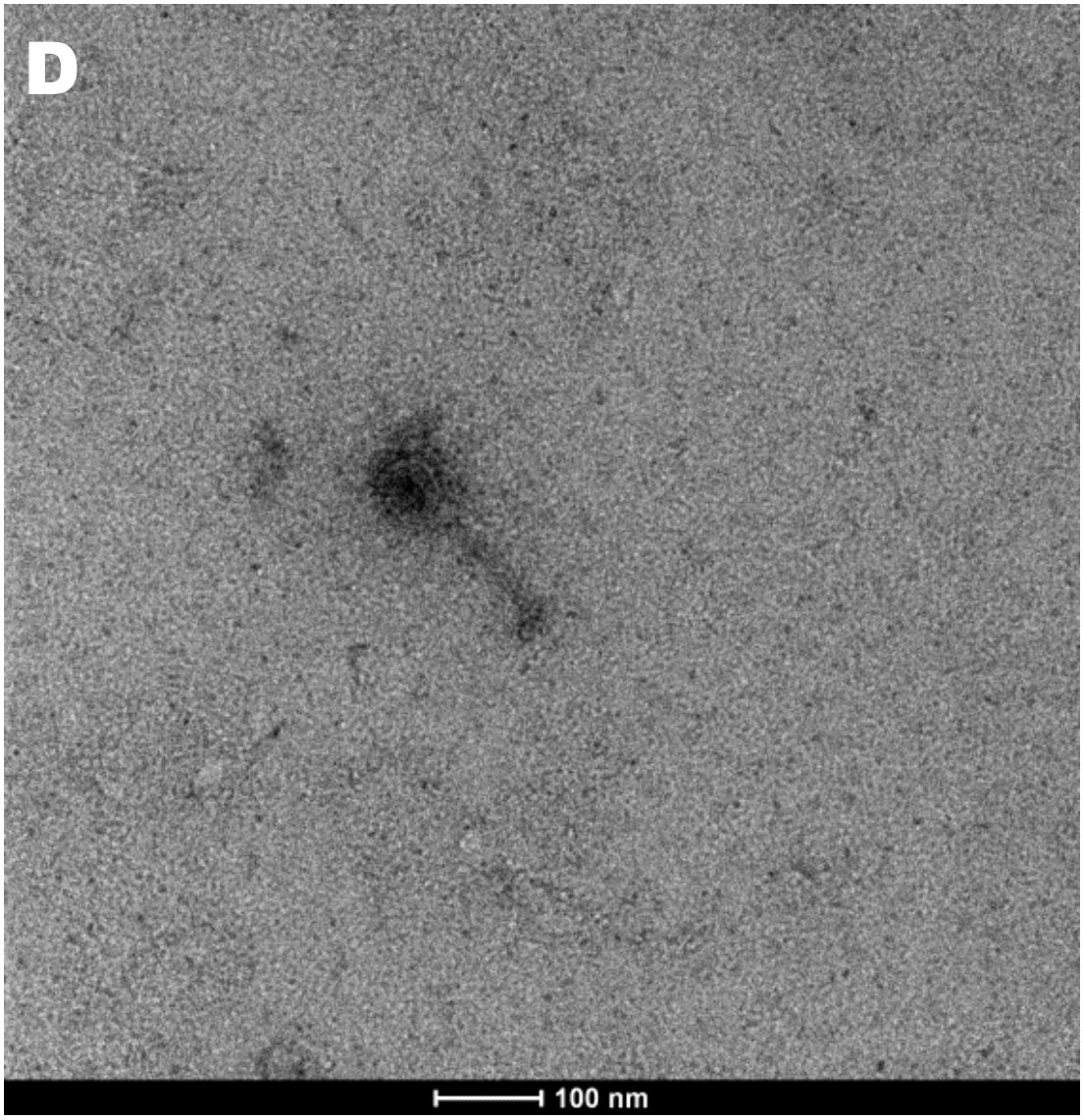


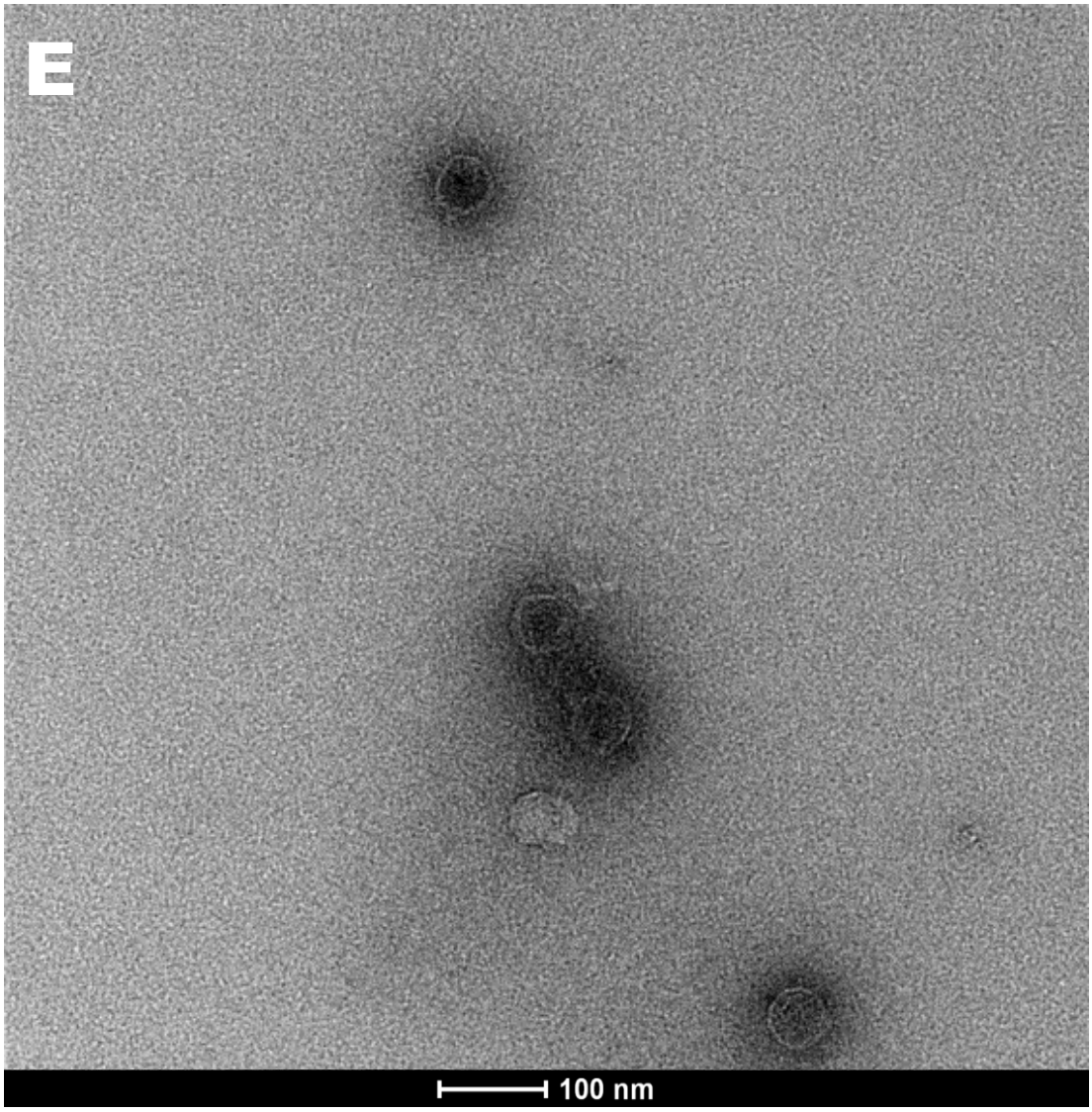
Figure 3. Immunological detection of TP901-1*erm* structural proteins precipitated from phage lysates by PEG 8000. Negative control represents PEG 8000 precipitation of mitomycin C induced strain NZ9000-Cro$_{t712}$. The protein detected by the primary antibody is depicted to the right of the image, with black-filled triangles highlighting protein sizes and a black-unfilled triangle indicating MCP1$_{Tuc2009}$ protein that is ~25 kDa smaller than MCP1$_{TP901-1}$.

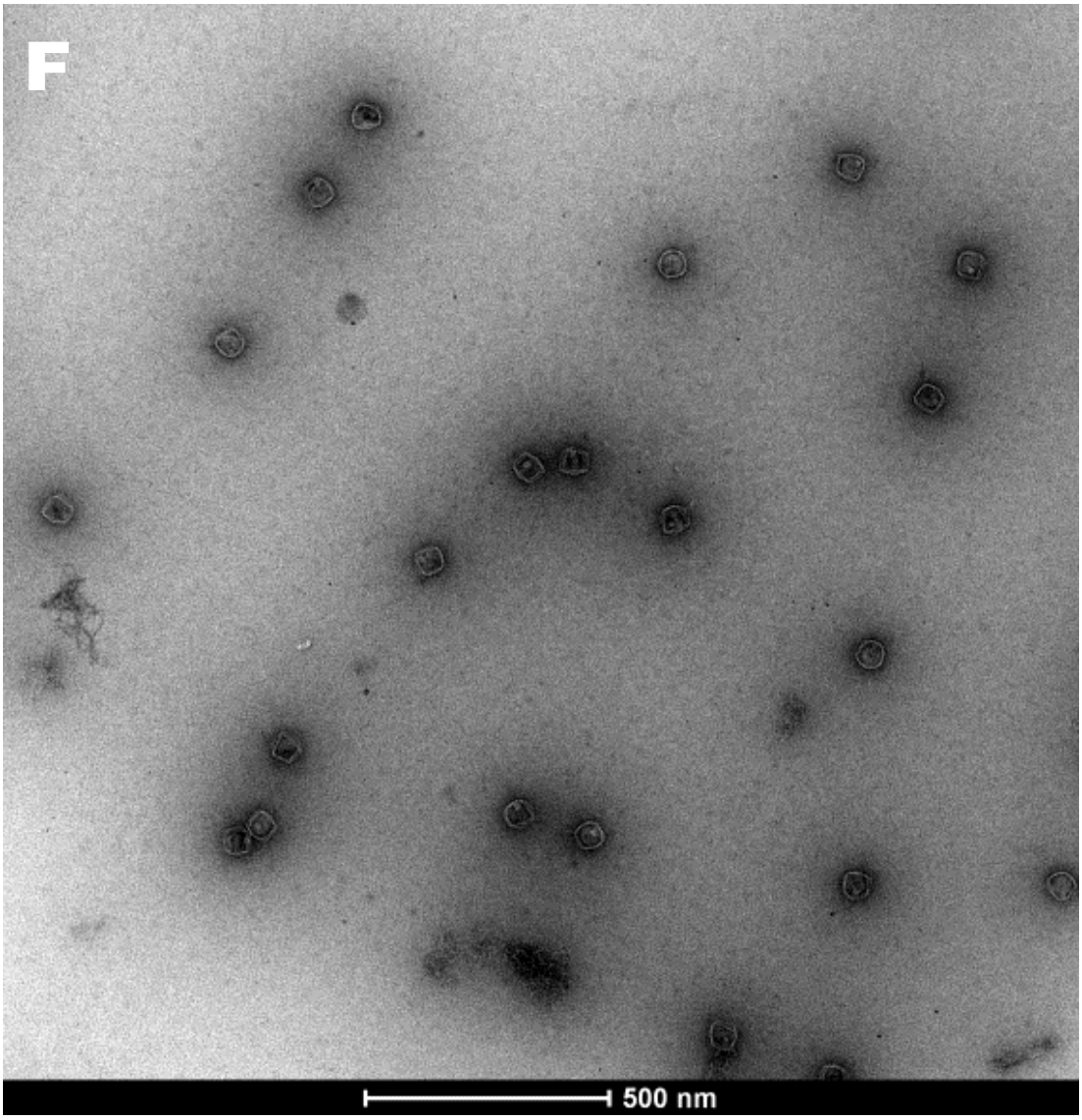

A
100 nm

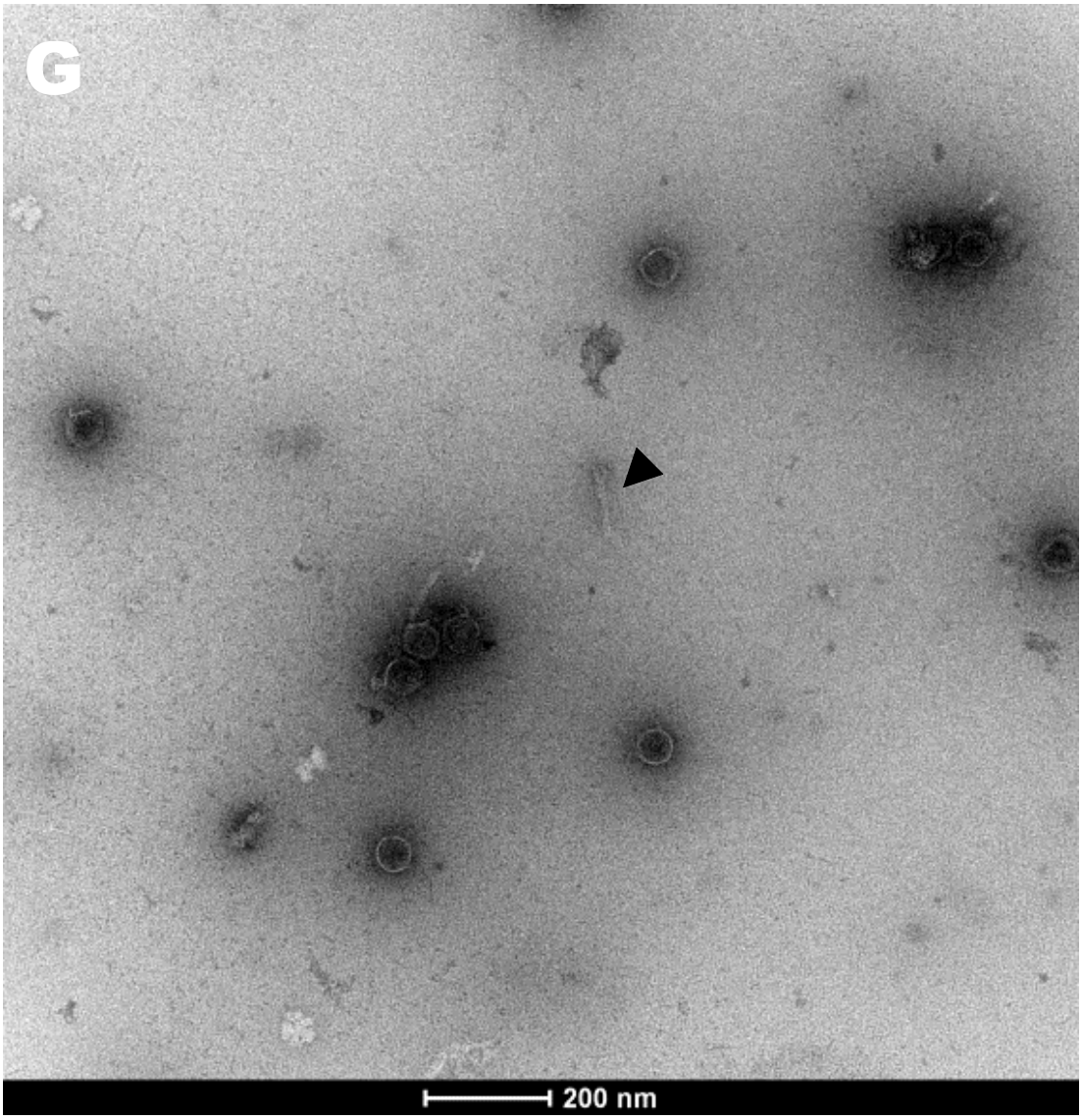

B
500 nm

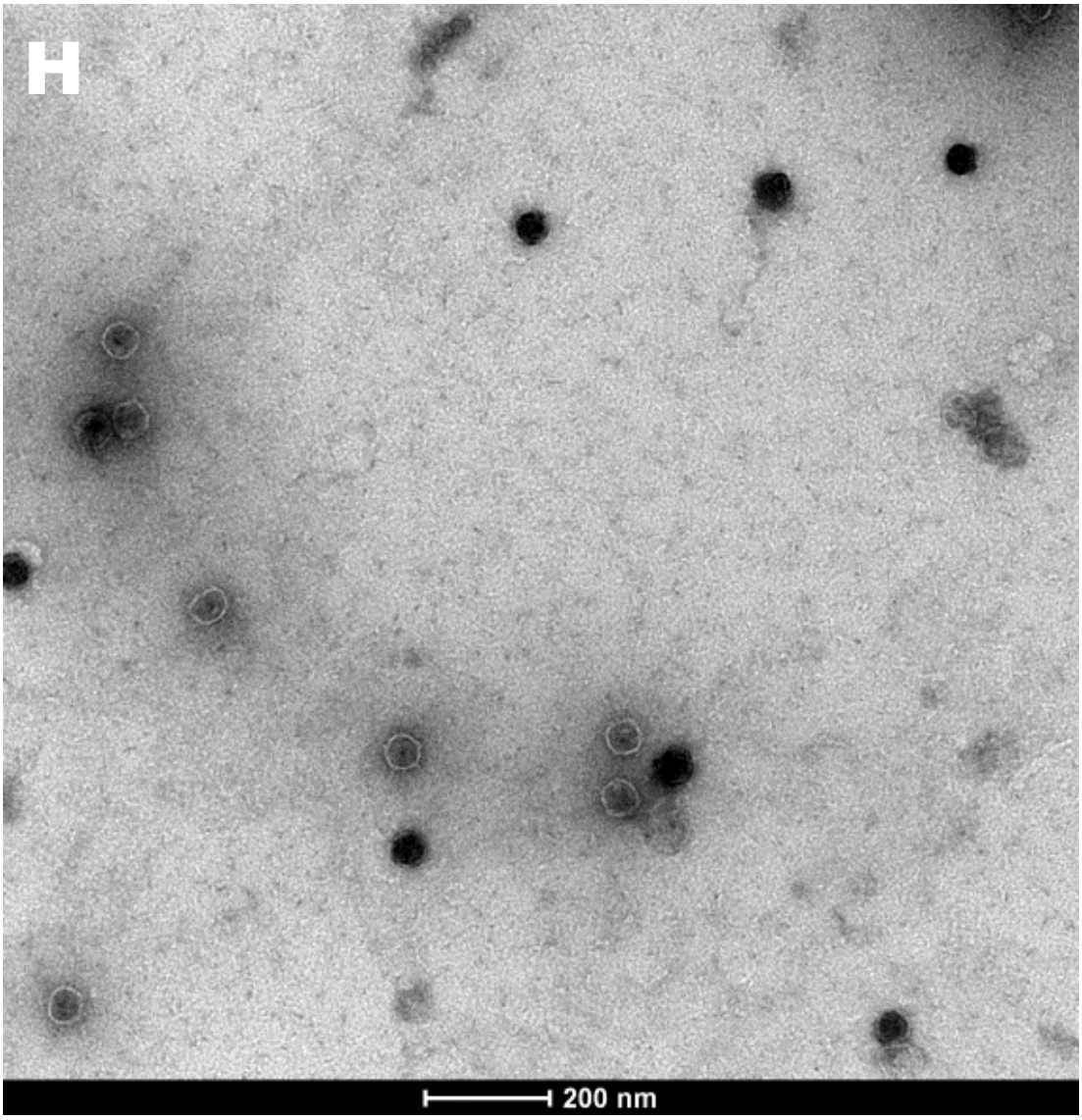

C
500 nm

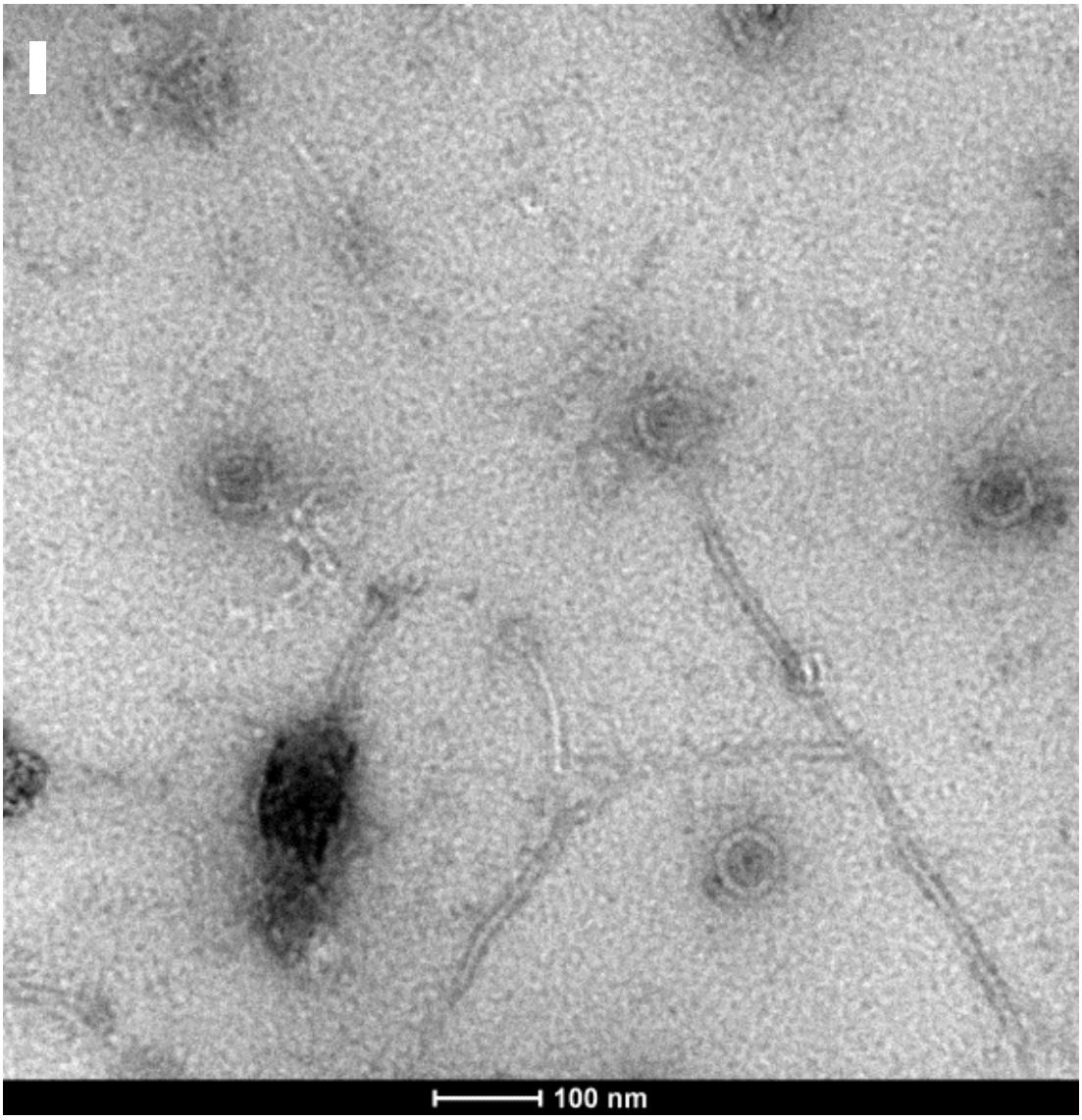

D
100 nm

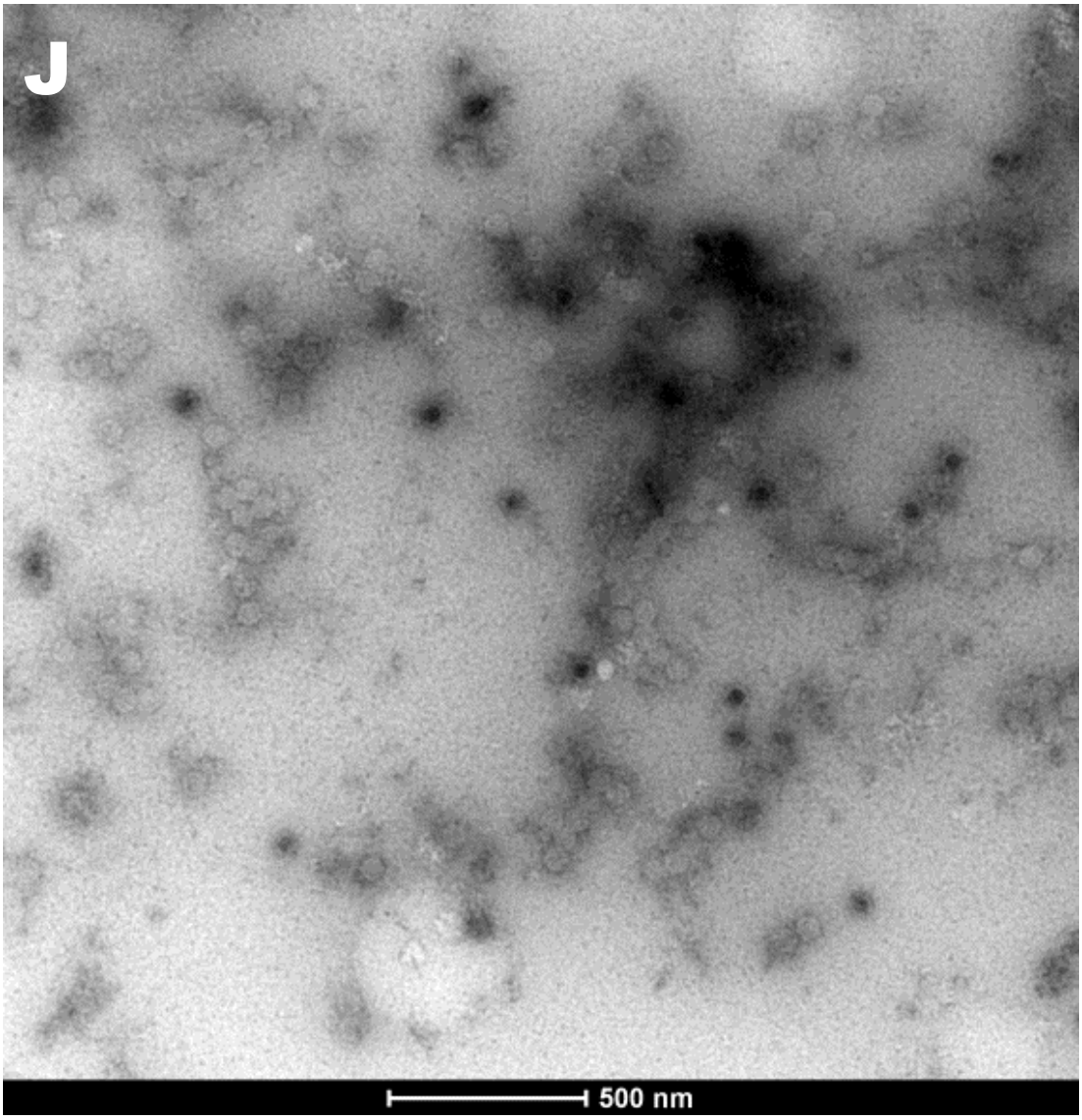

E
100 nm


F
500 nm


G
200 nm


H
200 nm


100 nm


J
500 nm

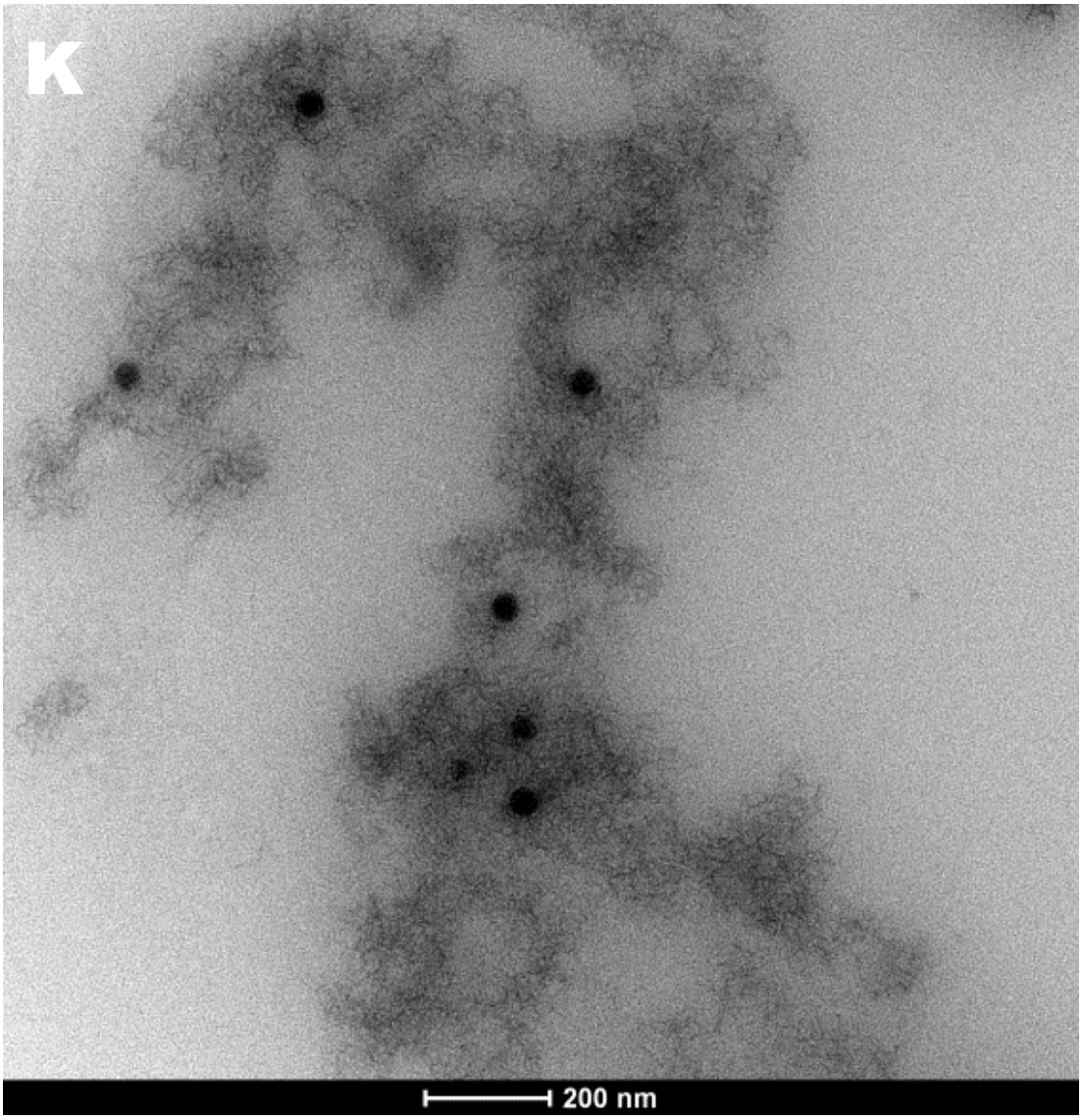

K
200 nm

Figure 4. EM images of (A) TP901-1*erm* and (B-K) structural module mutant derivatives. Where relevant, black-filled triangles indicate free tail structures. (B) Tails, but no capsids, produced by mutant $Portal_{TP901-1}$::Ter. (C) Capsids, without attached tails, visualised for mutant $MCP1_{TP901-1}$::Ter. (D) Full phage particles produced following mutation $MCP2_{TP901-1}$::Ter. (E) Capsids produced by $HTC2_{TP901-1}$::Ter. (F) Capsids isolated for mutant $gpZ_{TP901-1}$::Ter from lower sucrose gradient band (Supplementary Fig. 5). (G) Capsids and tails imaged for mutant $gpZ_{TP901-1}$::Ter from upper sucrose gradient band (Supplementary Fig. 5). (H) Capsids of mutant $gpU_{TP901-1}$::Ter observed in the lower sucrose gradient band (I) Capsids and polytails observed for mutant $gpU_{TP901-1}$::Ter in the upper sucrose gradient band. (J) Capsids only are observed for mutant $MTP_{TP901-1}$::Ter. (K) Capsids and insoluble aggregates, proposed to be the TMP, observed for mutant $gpG_{TP901-1}$::Ter. Refer to text for a more detailed description characterizing each mutant. Scale bars indicating the EM magnification is displayed below each image.

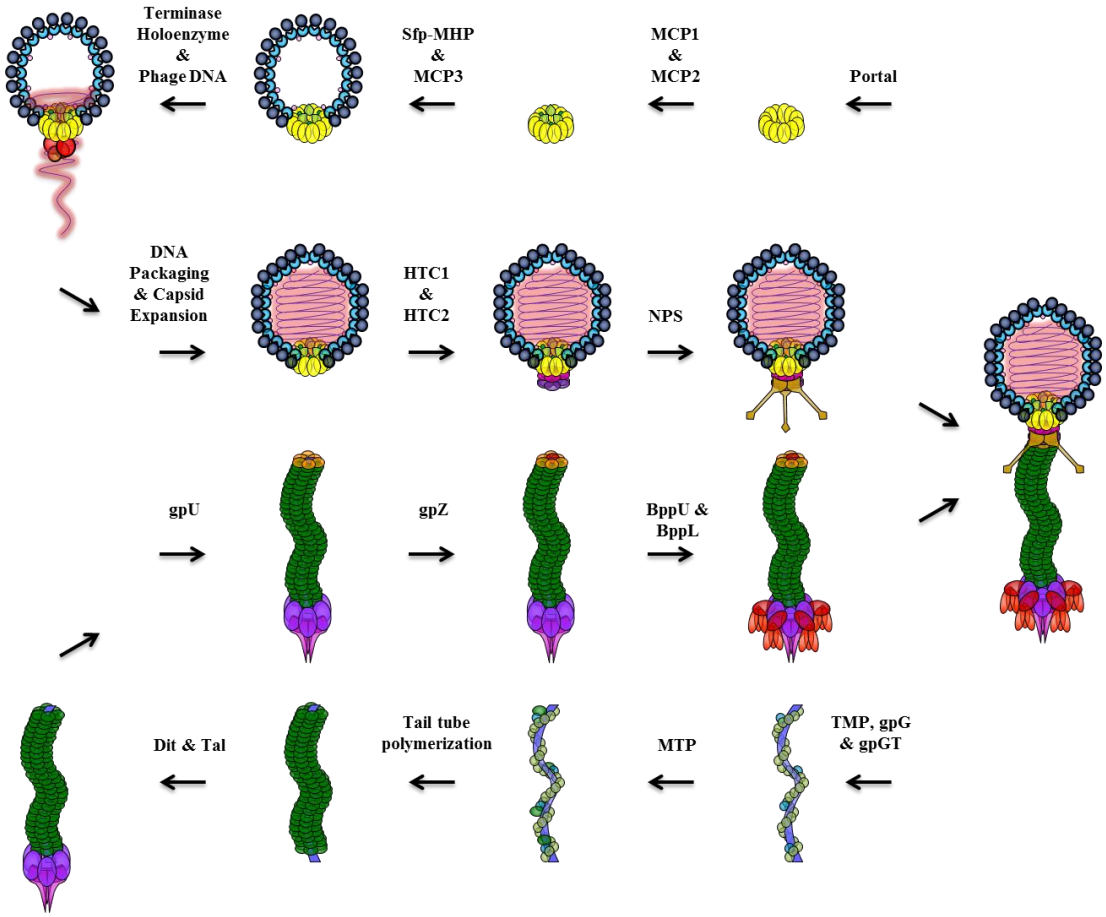
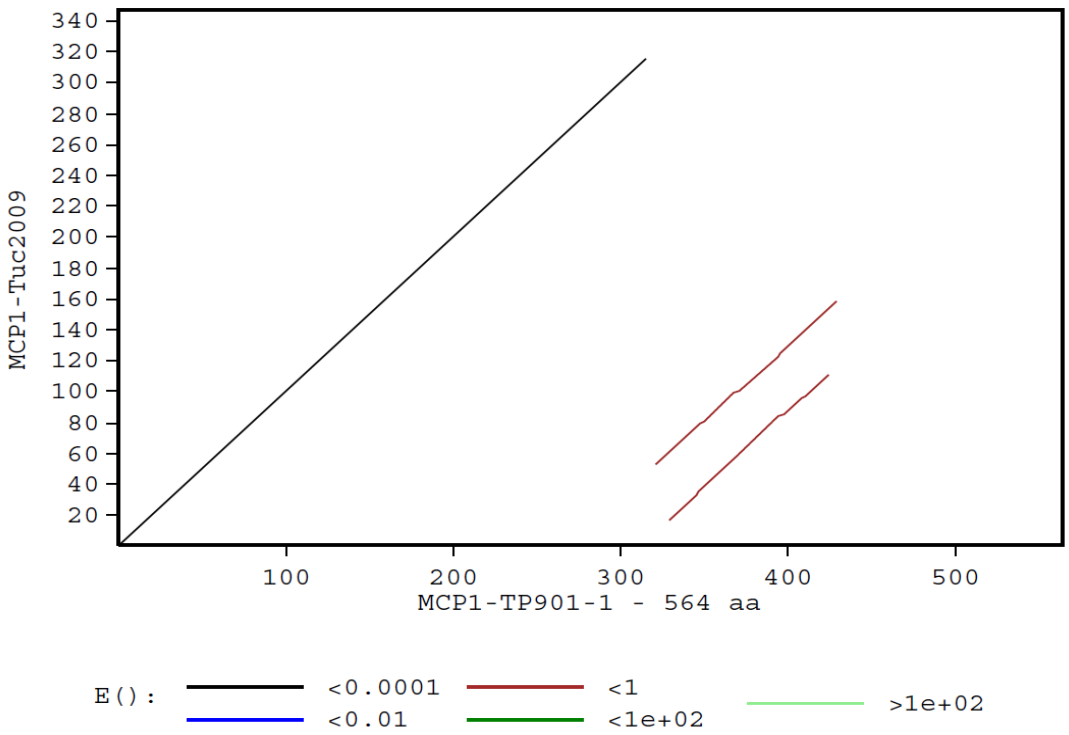


Figure 5. Molecular model putatively depicting TP901-1 and Tuc2009 phage assembly. Capsid and tail structures are produced independently and spontaneously assemble. Portal protein initiates capsid assembly, as mutant $Portal_{TP901\text{-}1}$::Ter fails to produce capsids. MCP1 is predicted to associate with the internal face of the portal protein complex. Due to $MCP2_{Tuc2009}$ being similar to YjcQ DNA-binding proteins, it is depicted as an internal capsid protein. Scaffolding-like proteins $Sfp_{TP901\text{-}1}$ and $Sfp_{Tuc2009}$ appear to be retained in their entirety in mature virions. Furthermore, Sfp and MHP appear to be added together to procapsids, as mutant $MHP_{TP901\text{-}1}$::Ter negatively impacts on the addition of $Sfp_{TP901\text{-}1}$ to assembling capsid complexes. Due to weak conservation of MCP3 with capsid protein domains involved in covalently bonding capsid subunits, MCP3 may have a capsid stabilizing function. DNA

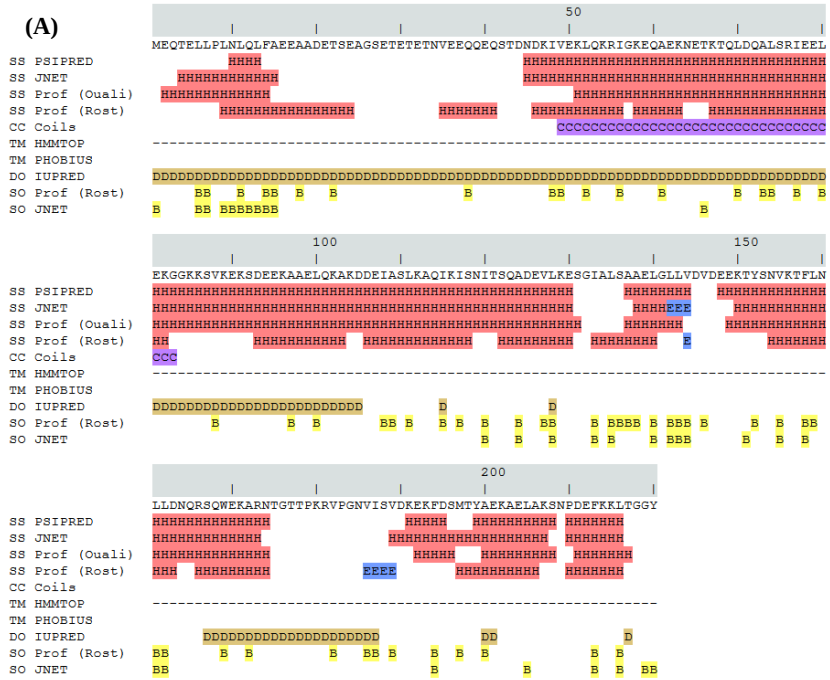
packaging by the terminase holoenzyme likely expands the capsids of TP901-1 and Tuc2009, creating more angular and stable structures. HTC1, and subsequently HTC2, are predicted to associate with capsids once DNA is packaged. TP901-1 and Tuc2009 phages contain non-essential NPS fibres associated with the portal vertex of capsids (34, 71). Tail assembly, illustrated as initiating with the TMP bound by chaperones gpG and gpGT, requires the C-terminus of the TMP, Dit and the N-terminus of Tal protein to form an initiator complex (47). These interactions are likely rapid, possibly occurring simultaneously. However, the 'G' domains of the chaperones bind and maintain the hydrophobic TMP in solution, facilitating initiator complex formation, while the 'T' domain of gpGT recruits MTP for tail tube polymerization. These results are supported by the lack of Dit protein in $gpG_{TP901-1}$::Ter mutant. Polymerization of the phage tail tube is concluded by the addition of tail terminator protein gpU to the final MTP ring. Protein gpZ is possibly required for the attachment of tails to mature capsids. During tail assembly, proteins BppU and BppL associate with Dit to form a baseplate structure which functions as the phage host-recognising receptor-binding complex.

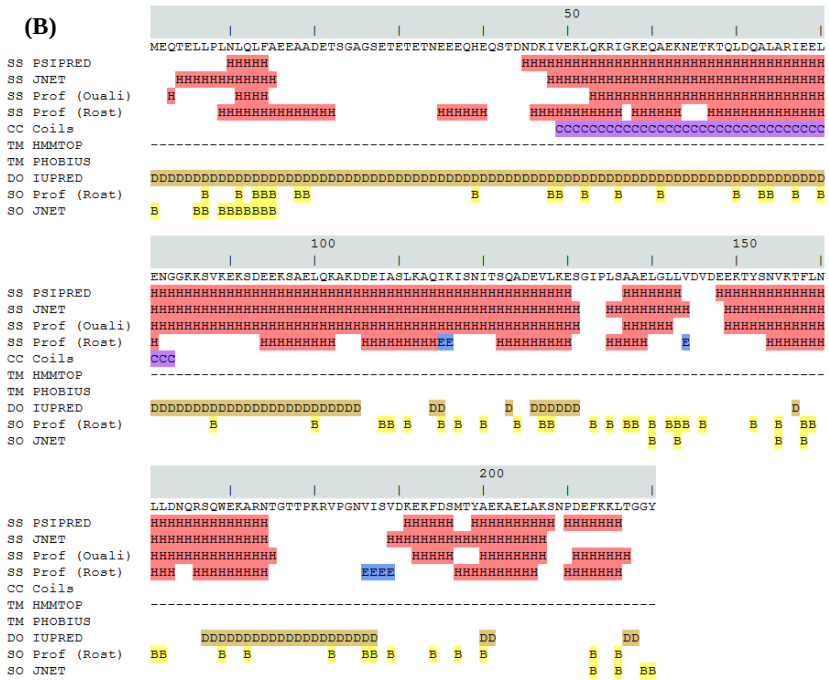


Supplementary Figure 1. Dot-plot analysis of $MCP1_{TP901-1}$ and $MCP1_{Tuc2009}$ using PLALIGN. The complete $MCP1_{Tuc2009}$ protein is homologous to $MCP1_{TP901-1}$ N-terminus; however, $MCP1_{TP901-1}$ possesses an approximately 28 kDa C-terminal extension (~250 aa). The initial ~100 amino acids of $MCP1_{TP901-1}$ C-terminal extension show it is derived from a possible genetic duplication.



Supplementary Figure 2. ClustalW alignment of $MCP1_{TP901\text{-}1}$, $MCP1_{Tuc2009}$ and phage SPP1 capsid protein gp7. The three proteins share significant conservation, except for the approximate 28 kDa C-terminal extension present in $MCP_{TP901\text{-}1}$.



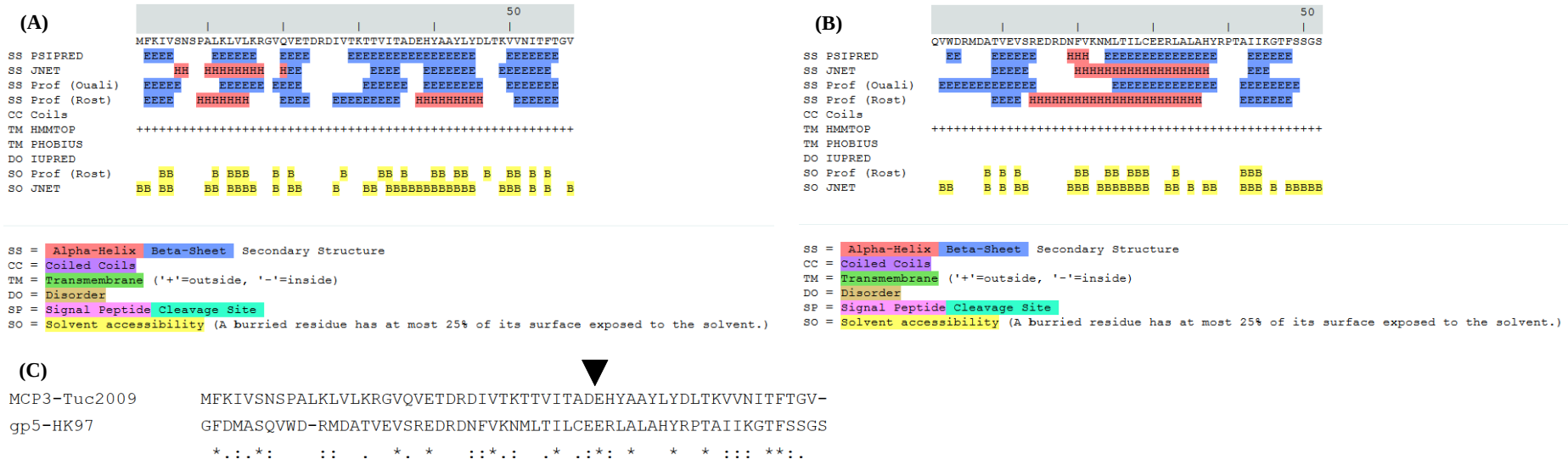


SS = Alpha-Helix Beta-Sheet Secondary Structure
CC = Coiled Coils
TM = Transmembrane ('+'=outside, '-'=inside)
DO = Disorder
SP = Signal Peptide Cleavage Site
SO = Solvent accessibility (A burried residue has at most 25% of its surface exposed to the solvent.)

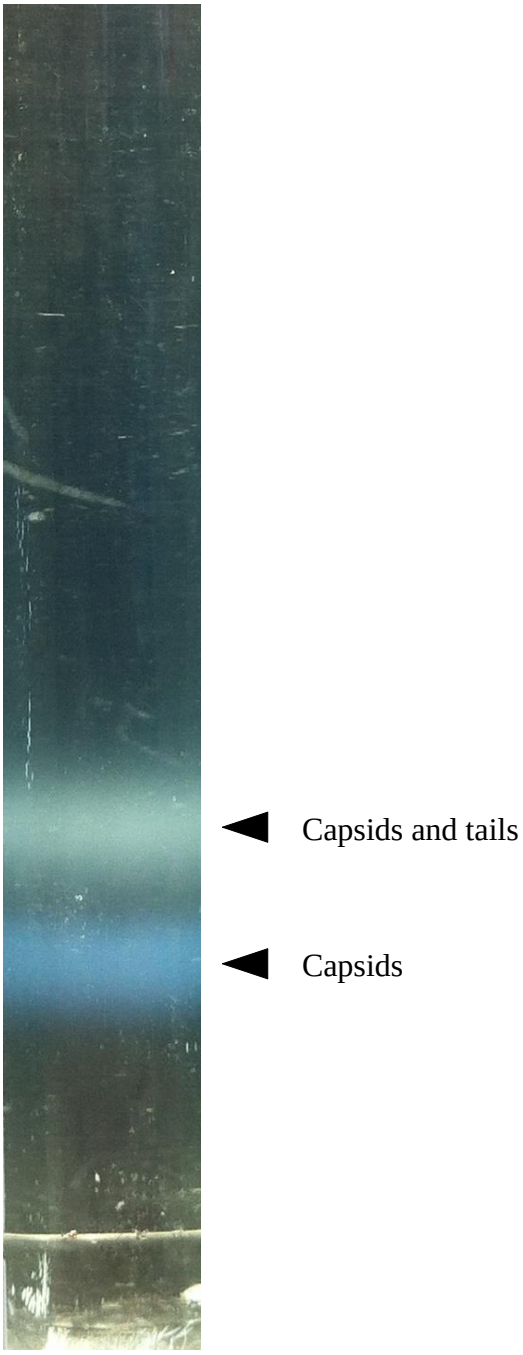
SS = Alpha-Helix Beta-Sheet Secondary Structure
CC = Coiled Coils
TM = Transmembrane ('+'=outside, '-'=inside)
DO = Disorder
SP = Signal Peptide Cleavage Site
SO = Solvent accessibility (A burried residue has at most 25% of its surface exposed to the solvent.)



Supplementary Figure 3. Secondary structure and intrinsically disordered-region predictions of (A) TP901-1, (B) Tuc2009, (C) phage λ and (D) phage SPP1 scaffolding, or scaffolding-like, proteins. $Sfp_{TP901\text{-}1}$, $Sfp_{Tuc2009}$, $gpNu3_{\lambda}$ and $gp11_{SPP1}$ are all predicted as largely α-helical in structure (PSIPRED) and contain large regions of intrinsically disordered protein structure (IUPRED).



Supplementary Figure 4. Secondary structure comparison and ClustalW alignment of $MCP3_{Tuc2009}$ and phage HK97 capsid protein gp5. Both $MCP3_{Tuc2009}$ and $gp5_{HK97}$ (A and B) show, albeit weakly, similar secondary structure (PSIPRED), and their ClustalW alignment (C) highlights several conserved amino acids. A conserved glutamic acid, highlighted by a black-filled triangle (C), is characterized in HK97 gp5 as involved in covalently crosslinking capsid subunits, and suggests a possible role for $MCP3_{Tuc2009}$ in capsid stabilization.



Supplementary Figure 5. Example of sucrose gradient purification (example shown is of TP901-1*erm* mutant $gpZ_{TP901\text{-}1}$::Ter). EM analysis of the bottom blue-grey band revealed phage capsids only, while the top yellow-grey band contained separate capsid and tail structures. Distinct bands, indicative of free capsids and tails, were analysed, where possible, separately.

CHAPTER IV

Mutational Analysis of a Lactococcal Tape Measure Protein Reveals its Multi-Functionality in Tail Morphology and DNA Injection

This chapter is in preparation for publication as a research article:

Mutational Analysis of a Lactococcal Tape Measure Protein Reveals its Multi-Functionality in Tail Morphology and DNA Injection.

ABSTRACT

The characteristic feature of *Siphoviridae* is their long non-contractile tail, which they use to efficiently recognise and initiate infection of their bacterial hosts. In this study, we examined the Tape Measure Protein (TMP) of the lactococcal phage TP901-1 in order to determine its role in assembling functional viral particles, in defining tail length, and in facilitating DNA injection into its host. Through the creation and analysis of 19 distinct deletion mutants in the TMP-encoding gene of TP901-1, we propose that TMP is located within the phage tail tube where its C- and N-terminal portion each possess distinct virion assembly functions at opposing ends of the tail. We furthermore identify repeats within the TMP primary sequence and propose their function is to determine the final phage tail length, while we also identify putative transmembrane-spanning regions located in the TMP and demonstrate their importance for the phage infection process.

INTRODUCTION

Bacteriophages are the most abundant biological entities on Earth, with an estimated 10^{31} viruses in the biosphere (1, 2). Three morphologically distinct families of tailed phages are most frequently observed in biological samples: *Myoviridae*, members of which possess long contractile tails; *Podoviridae*, which have short non-contractile tails; and *Siphoviridae*, which exhibit long non-contractile tails (3). *Siphoviridae* phages are believed to represent the dominant viral morphology (4), perhaps because their long flexible tails facilitate efficient infection of their bacterial hosts.

The length of a *Siphoviridae* tail is generally assumed to be dictated by the Tape Measure Protein (TMP) as in-frame deletions of, and insertions in, a *Siphoviridae* TMP-encoding gene proportionally decrease and increase, respectively, the length of the corresponding phage tail (5, 6). These findings suggest that the TMP acts as a molecular ruler, or scaffold, around which individual monomers of the major tail protein assemble to form the tail tube (7).

Probably the best understood *Siphoviridae* tail assembly model is that of phage λ, which proposes that six copies of the TMP, represented by gene product H (gpH), associate to form an 'initiator' complex with gpI, gpJ, gpK and gpL (8). However, the inclusion of gpH to the initiator complex occurs with the aid of two chaperone proteins, gpG and gpGT (8, 9), whose functions include maintaining the hydrophobic TMP in solution during tail assembly (10). Recently, Xu *et al.* (2013) demonstrated that chaperone gpG binds across the complete length of gpH. Chaperone gpGT, produced through a conserved -1 ribosomal frame-shift during mRNA translation at a specific sequence corresponding to the 3' end of *gpG* (11),

also binds to gpH through its N-terminal ‘G’ domain. However, gpGT also recruits Major Tail Protein (MTP or gpV) of phage λ to the assembling tail structure through its C-terminal ‘T’ domain, which is proposed to facilitate the assembly of the MTP-composed tail tube around and along the TMP (9, 12).

Polymerization of phage λ gpV into hexameric tail tube rings is proposed to pause when the end of gpH is reached (13, 14), at which point the tail terminator protein, gpU, associates with the last assembled gpV ring to terminate tail elongation (13, 15). Phage λ tails, being assembled independently of phage capsids (16), are then proposed to be activated by an unknown mechanism which involves the action of gpZ (13), and which facilitates subsequent association with individual phage capsids to produce fully functional infectious virions.

In addition to important roles in dictating tail length and assembly, the TMP of *Siphoviridae* phages is believed to form a channel in the bacterial membrane for DNA entry into the bacterial cytoplasm. Several studies have demonstrated that the TMPs of *Siphoviridae* are ejected from the tail tube at the onset of infection (17-19) and that they subsequently localise with cell wall and membranous fractions of infected bacteria (20). For both phage λ and T5, DNA injection into liposomes containing their proteinaceous receptors, LamB and FhuA, respectively, have been documented (21, 22). In both instances, transmembrane (Tm) domains associated with phage λ and T5 TMPs have been proposed to form the channel for DNA entry into membranous vesicles. However, the TMP of phage T5, designated Pb2, more clearly implicates *Siphoviridae* TMPs as creating Tm channels, as it could be purified from phage virions and incubated with lipid bilayer membranes to form pores (18).

In this study, we examined various functional roles of the TMP of lactococcal phage TP901-1 using a series of in-frame deletion mutants of $tmp_{TP901\text{-}1}$ targeting various domains and repeats. Based on the obtained results, we discuss the diverse roles of $TMP_{TP901\text{-}1}$ in assembling the long tail organelle of this *Siphoviridae*, and its importance in initiating infection.

MATERIALS

Bioinformatic Analysis. The TP901-1 genome sequence was downloaded from NCBI GenBank at accession number NC_002747 (23). The sequence of $TMP_{TP901-1}$, corresponding to the product of ORF45, was manipulated using DNASTAR software package (fifth edition, 2006; DNASTAR Inc., Madison, WI). Repeat sequences in $TMP_{TP901-1}$ were deduced by manually aligning Trp and Phe residues with 11 or 18 amino acid periodicity. Transmembrane domains were identified using TMHMM (24), while protein transmembrane topology was characterised using TOPO2 (25). The secondary structure of $TMP_{TP901-1}$ was predicted using SABLE (26) and alignment of TMP amino acid sequences was performed using ClustalW (27).

Bacterial Strains, Phages and Growth Conditions. Bacterial strains and phages used in this study are listed in Supplementary Table 1. Lactococcal strains were grown using standard conditions (28, 29). Phage lysates for the various biological characterisation assays were induced in an identical manner from three separate overnights of NZ9000-Cro_{t712} lysogenized with TP901-1*erm*, or a particular mutant derivative, and stored as outlined previously (Stockdale *et al.*, Chapter 3). Polyethylene glycol 8000 (PEG 8000) precipitation of phage particles (or tails and/or capsids) of the TP901-1*erm* phage or a particular mutant derivative, and their subsequent purification, was performed as described previously (Stockdale *et al.*, Chapter 3).

Mutant Generation and Screening. Recombineering mutagenesis was performed as described previously (29-31), with the following adjustments. Briefly, recombineering oligonucleotides designed to delete regions of $tmp_{TP901-1}$, which are

listed in Supplementary Table 2, bound to approximately 40-45 nucleotides of TP901-1*erm* genomic DNA flanking each deletion. Genomic deletions created in this study were on average designed to be approximately 100 nucleotides, although sequences as short as 57 nucleotide bases (phage mutant ΔTm3) and as long as 153 bases (mutant Δ1-9) were successfully removed in a single recombineering round. The corresponding amino acid sequences of $TMP_{TP901\text{-}1}$ mutations created in this study are outlined in Supplementary Table 3. Following transformations involving recombineering oligonucleotides, cells were allowed to recover for approximately 30 minutes before spread-plating the necessary dilutions on GM17 agar plates supplemented with 5 µg/ml erythromycin. Screening was performed by PCR amplification using NZ9000-Cro_{t712}-TP901-1*erm* colonies, or derivatives of this strain carrying the specific introduced mutations, as template DNA. Oligonucleotide pairs for screening were designed to generate an approximately 500 bp fragment for wild-type TP901-1*erm* genomic sequence, while generating an approximately 400 bp fragment following the successful deletion of a $tmp_{TP901\text{-}1}$ region. The desired sequence of each introduced mutation in $tmp_{TP901\text{-}1}$ was confirmed by Sanger sequencing the PCR product generated during screening (MWG Eurofins, Germany).

Phage Assays. Plaque assays of TP901-1*erm* and $TMP_{TP901\text{-}1}$ mutant derivatives were performed using standard protocols (32). The efficiency of plaquing (EOP) of TP901-1*erm* mutants, compared to the wild-type control, was calculated by dividing the plaque-forming units per millilitre (pfu/ml) of the particular TP901-1*erm* mutant by the pfu/ml of the un-mutated TP901-1*erm* control. EOP assays were performed on three separate replicates over three successive days, and the results averaged.

Adsorption assays of TP901-1*erm* and derived mutants were conducted as follows: *L. lactis* 3107 was grown to an A_{600} of ~0.3 at which point $CaCl_2$, 10 mM final concentration, was added to the culture. The *L. lactis* 3107 culture was dispensed into sterile eppendorfs in 450 µL aliqouts, before 50 µL TP901-1*erm* or a particular $TMP_{TP901-1}$ plaque-forming mutant was added to the culture. Phages were allowed to adsorb for 30 seconds before 1 M (v/v) final concentration of NaCl was added, after which bacteria were pelleted by centrifugation at 16,800 × g for 1 minute. The supernatant was aspirated into a fresh sterile eppendorf and tested for free phages to determine the final phage titre. In order to determine the initial titre of phages used for all adsorption assays, the protocol was conducted as described above using TP901-1*erm* and plaque-forming $TMP_{TP901-1}$ mutant derivatives except phages were added to sterile GM17 media containing 10 mM $CaCl_2$. The efficiency of phage adsorption, expressed as the percentage of phage adsorbing to their host in 30 seconds, was calculated by [(initial phage titre-final phage titre)/initial phage titre]*100 (33).

Frequency of lysogeny experiments were performed as described previously (Stockdale *et al.*, Chapter 3), except that all assays were performed using Mitomycin C-induced phage lysates of TP901-1*erm* and $TMP_{TP901-1}$ mutants rather than using sucrose gradient-purified fractions. The ability to measure frequency of lysogeny is based on bacterial acquisition of the adenine methylase gene present in the TP901-1*erm* phage genome or its mutant derivatives, conferring resistance to 5 µg/ml erythromycin (Erm5) (34, 35). A full list of results from the various phage assays characterising TP901-1*erm* and $TMP_{TP901-1}$ mutants is available in Table 1.

Protein Analysis. Separation of TP901-1*erm* or TP901-1*erm*-derivative virion proteins using 12 % polyacrylamide SDS-PAGE gels, their electro-blotting transfer to nitrocellulose membranes, and immunoblotting techniques were performed as described previously (Stockdale *et al.*, Chapter 3). PEG 8000 precipitates of TP901-1*erm* or derived mutant virions were subjected to Western blotting to determine the presence of Major Tail Protein (MTP) and Major Head Protein (MHP), indicative of tail tubes and capsids formed, respectively. A TP901-1*erm* mutant of the MTP-encoding gene (Stockdale *et al.*, Chapter 3) was used as a negative control for tail tube/MTP-detecting Western blots, while a MHP mutant of TP901-1*erm* (Stockdale *et al.*, Chapter 3) was used as a negative control for Western blots checking the production of phage capsids/MHP.

RESULTS

Characterisation of $TMP_{TP901-1}$. It has long been known that the TMPs of *Siphoviridae* control the length of their assembled tail structures, with deletions and insertions resulting in proportionally shorter and longer tails, respectively (5, 6). Katsura and Hendrix (1984) proposed that TMP proteins adopt a coiled-coil α-helical structure, with each amino acid of the TMP template protein roughly corresponding to 0.15 nm of the phage tail's final length (5). Secondary structure prediction of $TMP_{TP901-1}$ revealed, as expected, a largely α-helical content (Fig. 1). However, the C-terminal extremity of $TMP_{TP901-1}$ possesses a short coiled and β-sheet region, the potential significance of which is discussed below.

Bioinformatic programmes aimed at detecting repeat sequences within $TMP_{TP901-1}$ did not identify any strong candidates for repeating units within this protein (results not shown). Therefore, manual examination of the amino acid sequence of $TMP_{TP901-1}$ was conducted. Analysis was based on the characterisation of TMP proteins by Belcaid *et al.* (2011), who suggested specific repeat sequences in TMP proteins had a periodicity of 11 and 18 amino acids, and also by Siponen *et al.* (2009), who described repeating aromatic amino acids in the proposed TMP of lactococcal phage p2 (10, 36).

Manual alignment of Trp and Phe residues, spaced 11 or 18 amino acids apart, resulted in the identification of 29 aromatic-residue-containing repeat sequences (Fig. 2). Four of the deduced repeat sequences measured 18 amino acids in length, with each of these repeats beginning with a Trp residue. The remaining 25 repeats of $TMP_{TP901-1}$ are proposed to be 11 amino acids in length. While the majority of the twenty-five shorter repeats in $TMP_{TP901-1}$ are predicted to begin with

a Trp residue, four repeats commence with a Phe. Four of the 11 amino acid-long repeats begin with neither a tryptophan nor phenylalanine residue; however, these were still considered repeat sequences as the periodicity of aromatic amino acids continued following these repeats.

Analysis of $TMP_{TP901-1}$ revealed the presence of two Tm-containing domains, each with three Tm-spanning regions (Fig. 3). Interestingly, the carboxy-proximal Tm domain occurs within the deduced aromatic-residue-containing repeat sequences. The predicted topology of $TMP_{TP901-1}$, as it may occur following the protein's ejection from the tail tube, implicates both Tm domains as being inserted into the bacterial membrane (Fig. 4). In addition, the topological prediction of $TMP_{TP901-1}$ suggests that the stretch of 129 amino acids between the two Tm domains is located within the bacterial cytoplasm, while the N- and C-terminal extremities of $TMP_{TP901-1}$ are situated extracellularly.

Comparative analysis of the $TMP_{TP901-1}$ with its closely related homolog of lactococcal phage Tuc2009, $TMP_{Tuc2009}$, revealed that $TMP_{TP901-1}$ had suffered a deletion or that a duplication had occurred in $TMP_{Tuc2009}$ (Fig. 5). In either case, the difference between $TMP_{Tuc2009}$ and $TMP_{TP901-1}$ corresponds to 88 amino acids, immediately following the carboxy-proximal Tm-spanning domain. The differential 88 amino acids obey the characteristics identified in $TMP_{TP901-1}$ repeats, with aromatic residues occurring periodically at intervals of 11 amino acids (Fig. 5). In addition, studies examining TP901-1 and Tuc2009 tail structures proposed their tail tube lengths are 135 and 144 nm long, respectively, corresponding to $TMP_{TP901-1}$ (937 aa) being slightly smaller than $TMP_{Tuc2009}$ (1,025 aa) (6, 37).

Efficiency of plaquing. Mutations were introduced into the gene encoding $TMP_{TP901-1}$ in order to distinguish the various functions associated with this protein, and also to discover functionalities associated with particular regions of $TMP_{TP901-1}$. Deletions within the $TMP_{TP901-1}$-encoding gene, causing corresponding ablations of various aromatic residue-containing repeat sequences within $TMP_{TP901-1}$, were performed to presumably shorten the length of TP901-1*erm* tail. Such mutants are denoted by the symbol 'Δ' followed by the number assigned to the particular removed aromatic-residue-containing repeat sequences (Fig. 2). Deletions corresponding to the removal of Tm domain regions were carried out to determine their important in the phage infection process, as they are presumed to insert into the bacterial membrane following TMP ejection from the tail tube. These mutants were designated 'ΔTm', followed by the number given to individual Tm spanning regions of $TMP_{TP901-1}$ (Fig. 3B). Deletions affecting the N- or C-terminus of $TMP_{TP901-1}$, and named ΔN-terminus and ΔC-terminus, respectively, were conducted to investigate if the observed phenotype would be consistent with their respective presumed roles in either associating with capsid components or forming a tail-assembly initiator complex.

Lysates produced from mitomycin C-induced NZ9000-Cro_{t712}-TP901-1*erm* and derived strains each harbouring a TP901-1*erm* prophage carrying a particular mutation in the TMP-encoding gene were tested for their ability to form plaques on *L. lactis* 3107. Of the nineteen TMP-mutants created, only eight mutants displayed an Efficiency of Plaquing (EOP) at levels comparable to the wild-type TP901-1*erm* control (Fig. 6; Table 1). The eight $TMP_{TP901-1}$ mutants with EOPs comparable to TP901-1*erm* represent deletion mutants removing complete aromatic-residue-containing repeating units of 11 or 18 amino acids (Fig. 2). Interestingly, mutant Δ1-

7.5, in which one of the $TMP_{TP901-1}$ repeat sequences had partially been removed, formed no plaques (data not shown), whereas mutants Δ1-5 and Δ1-8, in which repeat sequences had been removed in full, exhibited EOPs of 1.96 $^{+}$/- 0.22 and 2.43 $^{+}$/- 0.26, respectively. TMP mutant Δ24-29 also formed plaques on host *L. lactis* 3107; however, its EOP was dramatically reduced (to an approximate EOP of 3.20e-4 relative to wild-type TP901-1*erm*). The possible reason for the dramatic decrease in EOP observed for Δ24-29 is discussed below. Analysis of the $TMP_{TP901-1}$ mutants which were shown to form plaques highlighted that only the EOP of mutant Δ1-2, which carries the smallest of the TMP repeat deletions (removing 29 amino acids; supplementary Table 3), was not statistically different from that of the wild-type TP901-1*erm* control. The EOP values obtained for all other TMP repeat deletion mutants were statistically different from that of the wild-type TP901-1*erm*, with *p*-values < 0.01.

Surprisingly, $TMP_{TP901-1}$ mutants Δ1-5, Δ1-8, Δ1-9, Δ24-26 and Δ20-26 displayed an increase in their EOP relative to that obtained for the TP901-1*erm* control. Mutant Δ20-26 yielded the highest EOP, approximately 12.5-fold higher than that obtained for TP901-1*erm.* Plaque sizes of TMP mutants that were capable of infecting host *L. lactis* 3107 displayed considerable variation (Fig. 7). The plaque sizes of TMP mutants Δ1-2, Δ1-5, Δ1-8 and Δ1-9 exhibited a gradual decrease in size compared to that produced by the TP901-1*erm* control, whereas plaques of TMP mutants Δ24-26 and Δ20-26 were similar in size to that of TP901-1*erm*. TMP mutants Δ1-5+24-26 and Δ1-9+20-26 formed pinprick plaques on their host. Possible reasons for the altered plaque morphologies are discussed below.

Adsorption Assays. In order to determine if the adsorption efficiency of TP901-1*erm* and $TMP_{TP901-1}$ mutants with predicted shorter tails differed, adsorption

assays were conducted (Fig. 8; Table 1). Phage adsorption to their lytic host, *L. lactis* 3107, is expressed as percentage of phage bound at 30 ^{o}C within 30 seconds. The adsorption of TP901-1*erm* control measured 34.05 $^{+}$/- 1.76 % (Fig. 8; Table 1). $TMP_{TP901-1}$ mutants Δ1-2, Δ1-5, Δ1-8 and Δ1-9, which steadily increase in their EOP with a presumed concomitant decrease in tail length (Fig. 6; Table 1), display a gradual decline in adsorption assay efficiency. Mutant Δ1-9, which represents the $tmp_{TP901-1}$ mutant removing the most DNA between the two encoded Tm-spanning domains, displayed an adsorption efficiency of 20.77 $^{+}$/- 1.57 %. In addition, $TMP_{TP901-1}$ deletion mutants Δ24-26 and Δ20-26, which exhibit a dramatically higher EOP than TP901-1*erm*, also displayed a substantial drop in their ability to adsorb to host *L. lactis* 3107; only 8.81 $^{+}$/- 0.18 % of the phages present in lysates of mutant Δ20-26 adsorbed under tested conditions. Mutants Δ1-5+24-26 and Δ1-9+20-26, which are predicted to produce the shortest TP901-1*erm* tail structures, while still forming visible plaques, display the lowest adsorption efficiencies, with just 5.09 $^{+}$/- 0.09 % of $TMP_{TP901-1}$ mutant Δ1-9+20-26 adsorbing to its host in 30 seconds at 30 ^{o}C.

Frequency of lysogeny. TP901-1*erm* and plaque forming $TMP_{TP901-1}$ mutant derivatives were tested for their ability to inject DNA into their hosts by performing frequency of lysogeny assays (Fig. 9A; Table 1). In addition, $TMP_{TP901-1}$ mutants which were shown to be unable to form plaques on their host were tested for their ability to inject DNA into a host bacterium following *in vitro* assembly with donated phage tails, in case the TMP mutation causes the formation of unattached capsids and tails, or a failure to produce tails. The infection-defective phenotype of such TMP mutants, as determined by frequency of lysogeny, can sometimes be rescued by *in vitro* assembly experiments (Fig. 9B; Table 1) (Stockdale *et al.*, Chapter 3).

While TP901-1*erm* control lysogenized its host at levels of 6.03e03 $^{+}$/- 5.24e2 Erm5 cfu/ml, $TMP_{TP901-1}$ mutants Δ1-2, Δ1-5, Δ1-8 and Δ1-9 steadily increased in their frequency of lysogeny (Fig. 9A; Table 1). $TMP_{TP901-1}$ mutants Δ1-5, Δ1-8 and Δ1-9, Δ24-26 and Δ20-26 displayed frequencies of lysogeny that were significantly higher than the TP901-1*erm* control (*p*-value < 0.01), with mutant Δ20-26 yielding 3.05e05 $^{+}$/- 4.23e04 Erm5 cfu/ml lysogenized cells under the same conditions as the control. Combinatorial $TMP_{TP901-1}$ mutants, including Δ1-9+24-26 whose plaques are too small to accurately count, demonstrated that the frequency of lysogeny steadily decreased for mutants Δ1-5+24-26, Δ1-9+24-26 and Δ1-9+20-26. This is in agreement with the EOP results for mutants Δ1-5+24-26 and Δ1-9+20-26, and expected EOP results for TMP mutant Δ1-9+24-26, with the shortest predicted tail produced by Δ1-9+20-26 lysogenizing its host at a frequency of 3.17e04 $^{+}$/- 6.19e03 Erm5 cfu/ml.

Lysates of non-plaque forming $TMP_{TP901-1}$ mutants were tested for their ability to lysogenize host *L lactis* 3107 on their own or following incubation with free phage tails (Fig. 9B), donated from $MHP_{TP901-1}$::Ter mutant (Stockdale *et al.*, Chapter 3). $TMP_{TP901-1}$ ΔN-terminus and ΔC-terminus yielded no erythromycin resistant colonies following incubation of the host strain with lysates of these mutants. However, frequency of lysogeny assays for $TMP_{TP901-1}$ mutants ΔN-terminus and ΔC-terminus following *in vitro* assembly with phage tails resulted in erythromycin colonies. Analysis of $TMP_{TP901-1}$ Tm spanning region mutants resulted in a few erythromycin resistant colonies observed on the undiluted plate for mutant ΔTm3, while no such colonies were observed for mutant ΔTm2-3. Lysogens of *L. lactis* 3107 were detected for ΔTm3 and ΔTm2-3 lysates mixed with phage tails donated from $MTP_{TP901-1}$::Ter; the possible reason for this observation is discussed

below. Unexpectedly, erythromycin resistant colonies, at a level statistically similar to the TP901-1*erm* control (*p*-value > 0.01), were observed for mutant ΔTm4 following the frequency of lysogeny assay. Furthermore, the subsequent incubation of $TMP_{TP901-1}$ mutant ΔTm4 with donated tail structures did not statistically result in an increase in its frequency of lysogeny compared to ΔTm4 lysates on their own (*p*-value > 0.01). A putative explanation for the behavioural phenotype of ΔTm4 is provided in the discussion.

TP901-1 immunoblotting. Western blots detecting TP901-1 MTP were conducted against PEG 8000 precipitated lysates of TP901-1*erm* and $TMP_{TP901-1}$ mutant derivatives which were unable to form plaques, as an indication of tail tubes produced (Fig. 10). MTP was clearly detectable in all $TMP_{TP901-1}$ mutants, but only weakly visible in mutant Δ1-7.5, while apparently absent for mutant ΔC-terminus. Western blots against the MHP of TP901-1*erm* TMP mutants was performed to confirm phage capsid proteins were present, signifying correct preparation of phage samples. The capsid protein was detectable for all TMP mutants, except for the negative control representing a mutant in the MHP (results not shown).

DISCUSSION

Bacteriophages, dominated by tailed phages of the order *Caudovirales*, are estimated to outnumber bacteria by about 10-fold (2). Detailed molecular models outlining the initiation of infection by *Myoviridae* and *Podoviridae* virions have previously been described (38, 39), yet a detailed model of how *Siphoviridae* adsorb, penetrate and inject DNA into their hosts is still lacking. In this study, we have examined the TMP of lactococcal phage TP901-1, in order to build a more detailed picture of this protein's role in assembly and length determination of the long non-contractile *Siphoviridae* tail organelle, while also to increase our understanding of how TMP proteins may facilitate DNA injection into their host.

$TMP_{TP901-1}$ is predicted as largely α-helical (Fig. 1), in agreement with the proposed coiled-coil secondary structure of TMP proteins (5). The C-terminus of $TMP_{TP901-1}$ contains a short region predicted to be less α-helical in structure, which is consistent with secondary structure predictions of phage λ TMP and also anticipated in the model of λ tail assembly to interact with various proteins to form an initiator complex for tail assembly (7). In the current study, removing the terminal 30 amino acids at the C-terminus of $TMP_{TP901-1}$ in mutant ΔC-terminus, appears to block tail formation as no MTP could be detected by means of Western blot analysis (Fig. 10). This finding corroborates the role of the $TMP_{TP901-1}$ C-terminus in forming an initiator-type complex for tail assembly (37, 40, 41) and also indicates that the C-terminal end of TMP within *Siphoviridae* tail structures is located distal to the capsid.

If the TMP C-terminus is located at the tip of *Siphoviridae* phage tails, the corresponding N-terminus is expected to be located proximal to phage capsids within

tail structures. $TMP_{TP901-1}$ mutant ΔN-terminus, removing 30 amino acids immediately downstream of the initiating methionine, produced tail proteins and capsid components as detected through Western blotting (Fig. 10). Frequency of lysogeny assays showed no infection of *L. lactis* by $TMP_{TP901-1}$ ΔN-terminus lysates; however, following the mixing of ΔN-terminus with a TP901-1*erm* mutant donating functional phage tails (Stockdale *et al.*, Chapter 3), erythromycin resistant lysogens of *L. lactis* 3107 were obtained (Fig. 9B; Table 1). This suggests that the N-terminus of $TMP_{TP901-1}$ is involved in phage head-tail connection, through either direct contact with capsid proteins or indirectly via an interaction with the tail-terminator protein which attaches to the terminal hexameric ring of the MTP tail tube (13, 15, 42).

While the TMP of *Siphoviridae* phages are known to dictate the final length of their tail structures (5, 6), to the best of our knowledge, no studies have specifically tried to shorten the length of *Siphoviridae* phage tails while maintaining their functionality. The identification of regularly occurring aromatic repeat sequences represented an opportunity to test the effect of shortening a *Siphoviridae* phage tail, while preserving the phage's infectivity, and understand why they possess such long tail organelles. Interestingly, deletions in $tmp_{TP901-1}$, presumed to shorten the final length of the phage's tail, result in an increase in the observed phage EOP and frequency of lysogeny, while simultaneously decreasing the phage's adsorption efficiency. Two possible factors which may explain the observed increase in titre of $TMP_{TP901-1}$ mutants with presumably shorter tails is: (i) recombinant TP901-1 phage, TP901-1*erm*, contains an additional adenine methylase gene for erythromycin resistance that may reduce the optimal packaging size of its genome within capsids; or (ii) due to their decreased adsorption efficiency, fewer $TMP_{TP901-1}$ mutants bind to cellular debris following lysis and become inactivated. However, combinatorial

mutations removing larger sections of *tmp*$_{TP901-1}$, while continuing to decrease in their adsorption efficiencies, also results in the gradual decrease in mutant EOPs. These results suggest the efficiency of adsorption, and subsequent optimal efficiency at which phage establishing infection, is linked with the length of *Siphoviridae* phage tails.

In order to confirm the nature of specific aromatic-residue-containing repeat sequences in assembling and determining the length of phage tails, TMP$_{TP901-1}$ mutant Δ1-7.5, out of sync with the 11 and 18 amino acid periodicity of tryptophan and phenylalanine residues, was created. Western blot detection of MTP present in Δ1-7.5 highlighted little MTP precipitated from phage lysates. In addition, the combination of Δ1-7.5 lysates with tails donated from MHP$_{TP901-1}$::Ter mutant resulted in substantially more erythromycin colonies during frequency of lysogeny assays, compared to lysogeny by TMP$_{TP901-1}$ Δ1-7.5 mutant on its own, signifying inefficient tail assembly for mutant Δ1-7.5. Potential reasons for the decreased production of tails with mutant Δ1-7.5 include the requirement of chaperone proteins gpG and gpGT to efficiently encompass the TMP during tail assembly. This is supported by structural studies of tail assembly chaperone proteins which demonstrated they form spiral structures in solution (10, 43), with chaperone ORF12 of lactococcal phage p2 containing regularly spaced hydrophobic patches along the inside of its chaperone spiral structure which is anticipated to coincide with the positions of aromatic residues along the phage TMP α-helical structure (10).

In addition to TMP$_{TP901-1}$ mutant Δ1-7.5, several TP901-1*erm* TMP mutants displayed a dramatic increase in their frequency of lysogeny following incubation with tails donated from mutant MHP$_{TP901-1}$::Ter. This indicates these mutants

inefficiently produce tail structures following induction from their host. All mutants created in this study analysing the annotated repeats 27-29, despite deletions obeying the 11 amino acid periodicity of repeat sequences, displayed either a dramatic reduction in EOP or low frequency of lysogeny. Two potential reasons for these observations are; (i) repeats 28 and 29 are in fact not repeat sequences, as they do not start with a tryptophan or phenylalanine amino acid; or (ii) repeat 29 is eighteen rather than eleven amino acids long, as a tryptophan residue occurs seven amino acids following the residue which was deemed to represent the end of this particular repeat sequence. However, without a better understanding of structure of TMP proteins as they occur within phage virions, including how repeats contribute to tail assembly and why they can be of different length (36), it is hard to predict phenotypes of particular mutants which do not assembly tails as efficiently as the wild-type phage.

Mutations in Tm spanning regions of $TMP_{TP901-1}$ substantiate the presumed importance of these domains for DNA injection into their host. Mutants ΔTm3 and ΔTm2-3, which are expected to produce phages with connected capsids and tails, yielded very few and zero erythromycin resistant colonies, respectively, following frequency of lysogeny experiments (Fig. 9B; Table 1). These findings support the presumed role of the N-terminal proximal Tm domain of $TMP_{TP901-1}$ in facilitating DNA injection into its host. However, frequency of lysogeny assays of ΔTm3 and ΔTm2-3 following incubation with functional tails yielded erythromycin resistant colonies, highlighting that, to a limited degree, these mutants inefficiently produced tail structures. This is in agreement with other $TMP_{TP901-1}$ mutants, as ΔTm3 and ΔTm2-3 do remove amino acids with a denomination of 11 or 18 residues (Supplementary Table 3).

Unexpectedly, $TMP_{TP901-1}$ mutant ΔTm4 lysogenized *L. lactis* 3107 at levels comparable to wild-type TP901-1*erm* during frequency of lysogeny assays (Fig. 9B; Table 1), despite not forming plaques on its host. Without structural data on the nature of the predicted Tm pores formed by TMPs during *Siphoviridae* DNA injecting into their hosts, it is hard to draw a conclusion on the role(s) of the C-terminal proximal Tm domain of TP901-1. However, from results obtained in this study, it seems that the C-terminal proximal Tm domain of $TMP_{TP901-1}$ is not essential for DNA injection into its host. In addition, the distance between the N-terminal and C-terminal proximal Tm domains is predicted as important for phage DNA injection and/or replication. Plaques produced by mutants Δ1-2, Δ1-5, Δ1-8 and Δ1-9 decrease in size when the intervening gap between the two Tm domains is decreased, and plaques are abolished for mutant Δ1-10 which overlaps with the start of the forth predicted Tm spanning region.

In order to confirm the predicted phenotypes of $TMP_{TP901-1}$ phage mutants produced in this study, electron microscopy (EM) images are required. Based on the biological characterization assays of TMP_{TP90-1} mutants, EM analysis, which is currently underway, is predicted to visualize: detached capsids and tail structures for mutant ΔN-terminus; capsids only for ΔC-terminus, excess capsid structures for mutants which remove sections of $tmp_{TP901-1}$ out of sync with 11 or 18 amino acid periodicity, due to inefficient tail production; and TP901-1*erm* with shorter tails for mutants deleting aromatic-residue-containing repeating units in the $TMP_{TP901-1}$-encoding gene.

In conclusion, several distinct functions of the TMP of lactococcal phage TP901-1 have been demonstrated in this study, including; (i) the role of the N- and

C-terminal regions of $TMP_{TP901-1}$ in tail assembly, (ii) the potential role of aromatic-residue-containing repeat sequences in controlling phage tail length, (iii) the necessity of repeats of specific length for efficient tail production, and (iv) the role of certain $TMP_{TP901-1}$ Tm spanning regions for DNA injection into a host bacterium. While this studies adds to the knowledge of *Siphoviridae* TMP tail assembly and function, more biological and, if possible, structural studies are required to further disentangle the complicated multi-functionality of *Siphoviridae* TMPs and their essential role during host infection.

REFERENCES

1. **Breitbart M, Rohwer F.** 2005. Here a virus, there a virus, everywhere the same virus? Trends Microbiol **13:**278-284.
2. **Wommack KE, Colwell RR.** 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev : MMBR **64:**69-114.
3. **Ackermann HW.** 1998. Tailed bacteriophages: the order *Caudovirales*. Adv Virus Res **51:**135-201.
4. **Ackermann HW.** 2007. 5500 Phages examined in the electron microscope. Arch Virol **152:**227-243.
5. **Katsura I, Hendrix RW.** 1984. Length determination in bacteriophage lambda tails. Cell **39:**691-698.
6. **Pedersen M, Ostergaard S, Bresciani J, Vogensen FK.** 2000. Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. Virology **276:**315-328.
7. **Katsura I.** 1990. Mechanism of length determination in bacteriophage lambda tails. Adv Biophys **26:**1-18.
8. **Katsura I, Kuhl PW.** 1975. Morphogenesis of the tail of bacteriophage lambda. III. Morphogenetic pathway. J Mol Biol **91:**257-273.
9. **Xu J, Hendrix RW, Duda RL.** 2013. Chaperone-Protein Interactions That Mediate Assembly of the Bacteriophage Lambda Tail to the Correct Length. J Mol Biol **426**:1004-1018.
10. **Siponen M, Sciara G, Villion M, Spinelli S, Lichiere J, Cambillau C, Moineau S, Campanacci V.** 2009. Crystal structure of ORF12 from

Lactococcus lactis phage p2 identifies a tape measure protein chaperone. J Bacteriol **191:**728-734.

11. **Xu J, Hendrix RW, Duda RL.** 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol Cell **16:**11-21.
12. **Xu J, Hendrix RW, Duda RL.** 2013. A Balanced Ratio of Proteins from Gene G and Frameshift-Extended Gene GT Is Required for Phage Lambda Tail Assembly. J Mol Biol **425:**3476-3487.
13. **Katsura I.** 1976. Morphogenesis of bacteriophage lambda tail. Polymorphism in the assembly of the major tail protein. J Mol Biol **107:**307-326.
14. **Pell LG, Kanelis V, Donaldson LW, Howell PL, Davidson AR.** 2009. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. Proc Natl Acad Sci U S A **106:**4160-4165.
15. **Pell LG, Liu A, Edmonds L, Donaldson LW, Howell PL, Davidson AR.** 2009. The X-ray crystal structure of the phage lambda tail terminator protein reveals the biologically relevant hexameric ring structure and demonstrates a conserved mechanism of tail termination among diverse long-tailed phages. J Mol Biol **389:**938-951.
16. **Weigle J.** 1966. Assembly of phage lambda *in vitro*. Proc Natl Acad Sci U S A **55:**1462-1466.
17. **Roa M.** 1981. Receptor triggered ejection of DNA and protein in phage lambda. FEMS Microbiol Lett **11:**257-262.

18. **Feucht A, Schmid A, Benz R, Schwarz H, Heller KJ.** 1990. Pore formation associated with the tail-tip protein pb2 of bacteriophage T5. J Biol Chem **265:**18561-18567.

19. **Plisson C, White HE, Auzat I, Zafarani A, Sao-Jose C, Lhuillier S, Tavares P, Orlova EV.** 2007. Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. EMBO J **26:**3720-3728.

20. **Marinelli LJ.** 2008. Characterization of the role of the Rpf motif in mycobacteriophage tape measure proteins. Ph.D. Thesis. University of Pittsburgh.

21. **Roessner CA, Ihler GM.** 1986. Formation of transmembrane channels in liposomes during injection of lambda DNA. J Biol Chem **261:**386-390.

22. **Bohm J, Lambert O, Frangakis AS, Letellier L, Baumeister W, Rigaud JL.** 2001. FhuA-mediated phage genome transfer into liposomes: a cryo-electron tomography study. Curr Biol **11:**1168-1175.

23. **Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW.** 2010. GenBank. Nucleic Acids Res **38:**D46-51.

24. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567-580.

25. **Moller S, Croning MD, Apweiler R.** 2001. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics **17:**646-653.

26. **Adamczak R, Porollo A, Meller J.** 2005. Combining prediction of secondary structure and solvent accessibility in proteins. Proteins **59:**467-475.

27. **Thompson JD, Higgins DG, Gibson TJ.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:**4673-4680.

28. **Vegge CS, Vogensen FK, McGrath S, Neve H, van Sinderen D, Brondsted L.** 2006. Identification of the lower baseplate protein as the antireceptor of the temperate lactococcal bacteriophages TP901-1 and Tuc2009. J Bacteriol **188:**55-63.

29. **Stockdale SR, Mahony J, Courtin P, Chapot-Chartier MP, van Pijkeren JP, Britton RA, Neve H, Heller KJ, Aideh B, Vogensen FK, van Sinderen D.** 2013. The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for infection specialization. J Biol Chem **288:**5581-5590.

30. **van Pijkeren JP, Neoh KM, Sirias D, Findley AS, Britton RA.** 2012. Exploring optimization parameters to increase ssDNA recombineering in *Lactococcus lactis* and *Lactobacillus reuteri*. Bioengineered **3:**209-217.

31. **van Pijkeren JP, Britton RA.** 2012. High efficiency recombineering in lactic acid bacteria. Nucleic Acids Res **40:**e76.

32. **Lillehaug D.** 1997. An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. J Appl Microbiol **83:**85-90.

33. **Sanders ME, Klaenhammer TR.** 1980. Restriction and modification in group N streptococci: effect of heat on development of modified lytic bacteriophage. Appl Environ Microbiol **40:**500-506.

34. **Koch B, Christiansen B, Evison T, Vogensen FK, Hammer K.** 1997. Construction of specific erythromycin resistance mutations in the temperate

lactococcal bacteriophage TP901-1 and their use in studies of phage biology. Appl Environ Microbiol **63:**2439-2441.

35. **Vegge CS, Vogensen FK, Mc Grath S, Neve H, van Sinderen D, Brondsted L.** 2006. Identification of the lower baseplate protein as the antireceptor of the temperate lactococcal bacteriophages TP901-1 and Tuc2009. J Bacteriol **188:**55-63.

36. **Belcaid M, Bergeron A, Poisson G.** 2011. The evolution of the tape measure protein: units, duplications and losses. BMC Bioinforma **12 Suppl 9:**S10.

37. **Mc Grath S, Neve H, Seegers JF, Eijlander R, Vegge CS, Brondsted L, Heller KJ, Fitzgerald GF, Vogensen FK, van Sinderen D.** 2006. Anatomy of a lactococcal phage tail. J Bacteriol **188:**3972-3982.

38. **Hu B, Margolin W, Molineux IJ, Liu J.** 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science **339:**576-579.

39. **Rossmann MG, Mesyanzhinov VV, Arisaka F, Leiman PG.** 2004. The bacteriophage T4 DNA injection machine. Curr Opin Struct Biol **14:**171-180.

40. **Vegge CS, Brondsted L, Neve H, Mc Grath S, van Sinderen D, Vogensen FK.** 2005. Structural characterization and assembly of the distal tail structure of the temperate lactococcal bacteriophage TP901-1. J Bacteriol **187:**4187-4197.

41. **Sciara G, Blangy S, Siponen M, Mc Grath S, van Sinderen D, Tegoni M, Cambillau C, Campanacci V.** 2008. A topological model of the baseplate of lactococcal phage Tuc2009. J Biol Chem **283:**2716-2723.

42. **Chagot B, Auzat I, Gallopin M, Petitpas I, Gilquin B, Tavares P, Zinn-Justin S.** 2012. Solution structure of gp17 from the *Siphoviridae* bacteriophage SPP1: Insights into its role in virion assembly. Proteins **80:**319-326.

43. **Pell LG, Cumby N, Clark TE, Tuite A, Battaile KP, Edwards AM, Chirgadze NY, Davidson AR, Maxwell KL.** 2013. A conserved spiral structure for highly diverged phage tail assembly chaperones. J Mol Biol **425:**2436-2449.

44. **Koch B, Christiansen B, Evison T, Vogensen FK, Hammer K.** 1997. Construction of specific erythromycin resistance mutations in the temperate lactococcal bacteriophage TP901-1 and their use in studies of phage biology. Appl Environ Microbiol **63:**2439-2441.

45. **Braun V, Hertwig S, Neve H, Geis A, Teuber M.** 1989. Taxonomic Differentiation of Bacteriophages of *Lactococcus lactis* by Electron-Microscopy, DNA DNA Hybridization, and Protein Profiles. J Gen Microbiol **135:**2551-2560.

Table 1. Results of biological assays analysing TP901-1*erm* and $TMP_{TP901-1}$ mutants.

TP901-1*erm* or $TMP_{TP901-1}$ Tested	Result Average	Standard Deviation
Efficiency of Plaquing (EOP)		
TP901-1*erm*	1.00	+/- 0.046
Δ1-2	0.67	+/- 0.067
Δ1-5	1.96	+/- 0.224
Δ1-8	2.43	+/- 0.258
Δ1-9	2.89	+/- 0.266
Δ24-26	7.04	+/- 0.298
Δ20-26	12.52	+/- 1.483
Δ1-5+24-26	0.44	+/- 0.066
Δ1-9+24-26	N/A	N/A
Δ1-9+20-26	0.35	+/- 0.008
Percentage Phage Adsorption	**(%)**	**(%)**
TP901-1*erm*	34.05	+/- 1.76
Δ1-2	27.84	+/- 5.60
Δ1-5	26.37	+/- 1.82
Δ1-8	21.98	+/- 2.40
Δ1-9	20.77	+/- 1.57
Δ24-26	11.49	+/- 1.39
Δ20-26	8.81	+/- 0.18
Δ1-5+24-26	8.63	+/- 0.19
Δ1-9+20-26	5.09	+/- 0.09
Frequency of Lysogeny	**(Erm5 cfu/ml)**	**(Erm5 cfu/ml)**
TP901-1*erm*	6.03e03	+/- 5.24e02
Δ1-2	8.43e03	+/- 3.33e02
Δ1-5	3.20e04	+/- 2.90e03
Δ1-8	4.62e04	+/- 3.31e03
Δ1-9	6.08e04	+/- 3.76e03
Δ24-26	1.34e05	+/- 9.07e03
Δ20-26	3.05e05	+/- 4.32e04
Δ1-5+24-26	7.27e04	+/- 1.00e04
Δ1-9+24-26	5.73e04	+/- 3.67e03
Δ1-9+20-26	3.17e04	+/- 6.19e03

TP901-1*erm* or $TMP_{TP901-1}$ Tested	Result Average	Standard Deviation
Frequency of Lysogeny	**(Erm5 cfu/ml)**	**(Erm5 cfu/ml)**
ΔN-terminus	0.00	+/- 0.00
ΔN-terminus + Tails	1.19e04	+/- 3.47e03
ΔC-terminus	0.00	+/- 0.00
ΔC-terminus + Tails	3.23e03	+/- 1.30e03
ΔTm3	1.25e00	+/- 5.00e-01
ΔTm3 + Tails	9.32e02	+/- 1.47e02
ΔTm2-3	0.00	+/- 0.00
ΔTm2-3 + Tails	2.28e03	+/- 2.06e02
ΔTm4	1.01e04	+/- 1.52e03
ΔTm4 + Tails	1.58e04	+/- 3.54e03
Δ1-7.5	4.95e01	+/- 9.88e00
Δ1-7.5 + Tails	2.14e04	+/- 5.53e03
Δ1-10	8.25e01	+/- 3.87e00
Δ1-10 + Tails	1.63e04	+/- 2.27e03
Δ24-29	1.01e02	+/- 1.99e01
Δ24-29 + Tails	2.06e04	+/- 4.16e03
Δ1-9+20-29	2.53e02	+/- 3.50e01
Δ1-9+20-29 + Tails	1.40e04	+/- 3.06e03

N/A Not applicable

Erm5 cfu/ml 5 µg/ml erythromycin-resistant colony forming units per millilitre

Supplementary Table 1. Bacterial strains, phages and plasmids used in this study.

Strains, phages and plasmids	Relevant features	Source
Lactococcal strains		
3107	*L. lactis* strain sensitive to infection by phage TP901-1*erm*	
NZ9000-Cro_{t712}-TP901-1*erm*	*L. lactis* NZ9000 derivative with mutation in prophage t712 deduced anti-repressor (Cro) promotor sequence and lysogenized with TP901-1*erm*; contains plasmid pJP005	(Stockdale *et al.*, Chapter 3)
Phages		
TP901-1*erm*	Temperate P335 species phage infecting *L. lactis* strain 3107; contains erythromycin marker	(35, 44, 45)
TP901-TMPΔ1-2	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 and #2 (see Fig. 2)	This study
TP901-TMPΔ1-5	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 to #5; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ1-2	This study
TP901-TMPΔ1-7.5	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 to #7, and 9 out of the 18 amino acids of repeat #8; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ1-5	This study
TP901-TMPΔ1-8	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 to #8; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ1-5	This study
TP901-TMPΔ1-9	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 to #9; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ1-5	This study
TP901-TMPΔ1-10	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 to #10; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ1-8	This study
TP901-TMPΔ24-26	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #24 to #26	This study
TP901-TMPΔ20-26	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #20 to #26; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ24-26	This study
TP901-TMPΔ24-29	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #24 to #29; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ24-26	This study
TP901-TMPΔ1-5+24-26	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 to #5 and #24 to #26; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ1-5	This study
TP901-TMPΔ1-9+24-26	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 to #9 and #24 to #26; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ1-5+24-26	This study
TP901-TMPΔ1-9+20-26	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 to #9 and #20 to #26; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ1-9+24-26	This study
TP901-TMPΔ1-9+24-29	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 to #9 and #24 to #29; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ1-9+24-26	This study
TP901-TMPΔ1-9+20-29	Deletion mutation in $tmp_{TP901\text{-}1}$ removing repeats #1 to #9 and #20 to #29; derived from sequential mutation of $TMP_{TP901\text{--}1}$ mutant Δ1-9+20-26	This study
TP901-TMPΔTm3	Deletion mutation in $tmp_{TP901\text{-}1}$ removing the third annotated transmembrane domain (Fig. 3)	This study
TP901-TMPΔTm2-3	Deletion mutation in $tmp_{TP901\text{-}1}$ removing the second and third transmembrane domains	This study
TP901-TMPΔTm4	Deletion mutation in $tmp_{TP901\text{-}1}$ removing the fourth transmembrane domain	This study
TP901-TMPΔN-term	Deletion mutation in $tmp_{TP901\text{-}1}$ removing 30 amino acids, following the inaugural methionine, from the N-terminal-encoding region of $tmp_{TP901\text{-}1}$	This study
TP901-TMPΔC-term	Deletion mutation in $tmp_{TP901\text{-}1}$ removing 30 amino acids from the C-terminal-encoding region of $tmp_{TP901\text{-}1}$	This study
Plasmids		
pJP005	pNZ8048 derivative expressing RecT protein required for recombineering mutagenesis	(31)

Supplementary Table 2. Recombineering and screening oligonucleotides used in this study to create TP901-1*erm* TMP mutants.

Oligonucleotide name	Prerequisite	Oligonucleotide sequence (5' → 3')
Recombineering primers		
Δ1-2	None	A*A*T*G*T*TGTTCCAGGTATCAGAGAACCATTGGGTAATATCTCCCCAATTTTTATAAACTAAAATTCCTATAGCTATAACAGCCGCTATGGC
Δ1-5	Δ1-2	C*G*T*A*C*CTTTCCAAAGATTAGTGAACCACTCCTTAATGCCATTCCAATTTTTATAAACTAAAATTCCTATAGCTATAACAGCCGCTATGGC
Δ1-7.5	Δ1-5	A*T*C*A*A*GAATTGGTTGAAAGATTGTTTTGATTGTCTCAACAAATGGATTTTTATAAACTAAAATTCCTATAGCTATAACAGCCGCTATGGC
Δ1-8	Δ1-5	G*A*C*T*T*GCCCCCATAATCCGCTAAAGAAATCAAGAATTGGTTGAAAATTTTTATAAACTAAAATTCCTATAGCTATAACAGCCGCTATGGC
Δ1-9	Δ1-5	T*C*T*C*C*CAAGCTGAACCAAAGATAGTTTGGACTTGCCCCCAATTTTTATAAACTAAAATTCCTATAGCTATAACAGCCGCT
Δ1-10	Δ1-8	G*A*G*T*A*ACAAAACTGGTCCCATAACAACCGTCTTAATAATCTCCCAATTTTTATAAACTAAAATTCCTATAGCTATAACAGCCGCTATGGC
Δ24-26	None	T*A*C*C*A*ATATCTAGCAAGTTTATATTTTTTAGATTGTTAAAACCTTGTTTGACTCCATTCACAATAGCATTGGCAGAATCA
Δ20-26	Δ24-26	G*G*C*T*T*TACCAATATCTAGCAAGTTTATATTTTTTAGATTGTTAAAACCATTAACAATTCCGTCAACAATAGACTTGGCCATATTAATGAC
Δ24-29	Δ24-26	T*G*G*C*C*CTTTATGCTTCCGAATCCAATCGCCAATCCCGCTAATGAAACCTTGTTTGACTCCATTCACAATAGCATTGGCAGAATCAACGAC
Δ20-29	Δ20-26	T*G*G*C*C*CTTTATGCTTCCGAATCCAATCGCCAATCCCGCTAATGAAACCATTAACAATTCCGTCAACAATAGACTTGGCCATATTAATGAC
ΔTm3	None	A*T*A*C*T*TCTTTCCCGAATTTAGAAATTTTATCCCAATTTTTTGGTAAAGAAATAGCAGTAAAACTTCCTCCTATTAGAGCA
ΔTm2-3	ΔTm3	T*G*C*C*C*ATACTTCTTTCCCGAATTTAGAAATTTTATCCCAATTTTTTACTCTTTGTAGAGTAACAACAGCTTGCCCAAATATTACTAATAA
ΔTm4	None	G*C*A*T*C*GCAAAATCTTTTTTGAATTGGTTAAAGTCCCCAGTGATAGTTTGGACTTGCCCCCATAATCCGCTAAAGAAATCA
ΔN-term	None	C*T*T*T*C*CCTACTGTAGACATCGTACTATTAGTGTTCTTTTCCATATTTTCCTCCTTTCCTAGTTATTTGCTTTTTTCATGA
ΔC-term	None	G*T*T*T*T*GTCGTATCTCTAAACTTGTACATTTCTACCTCCTAGAATGAAGCCGGCTGTTTTCCTTGGTTAAGATTCACATCA
Screening primers		
Del1-10-Fw		GCAGTTAGGAAAGTTGCTGA
Del1-10-Rv		GTCCCATAACAACCGTCTT
N-term-Del-Fw		CGAAGAAGAATTCATTTATGAACAAGC
N-term-Del-Rv		GCTTGTTGTCTAAGTTCTTCG
C-term-Del-Fw		GCTTGTTGTCTAAGTTCTTCG
C-term-Del-Rv		ACATTAGTGATCACACCTCC
Tm2-3-Del-Fw		AGTTGCTGATTCAATTTCAGGAT
Tm2-3-Del-Rv		ATTTTTGGCACCTTGGACTG
Tm4-Del-Fw		AGAAATTTGCGTCCGATGCA
Tm4-Del-Rv		TGTTTGCCAGAGCATCGCA
Del24-26-Fw		TTCTAATATCTGGACGACAGT
Del24-26-Rv		CCAATATCCACAGAATAGTCA
Del20-29-Fw		TTCTAATATCTGGACGACAGTTGT
Del20-29-Rv		CCTTGGTTAAGATTCACATCATGAG

* Phosphorothioate linkages of recombineering oligos.

Supplementary Table 3. Resultant amino acid sequences of TP901-1*erm* TMP mutants following recombineering mutagenesis.

Mutant name	Number of amino acids deleted	Resultant amino acid sequence of TP901-1 TMP mutants[1, 2]
Δ1-2	29	AIAAVIAIGILVYKN / WGDITQWFSDTWNNI
Δ1-5	69	AIAAVIAIGILVYKN / WNGIKEWFTNLWKGT
Δ1-7.5	100	AIAAVIAIGILVYKN / PFVETIKTIFQPILD
Δ1-8	109	AIAAVIAIGILVYKN / FQPILDFFSGLWGQV
Δ1-9	120	AIAAVIAIGILVYKN / WGQVQTIFGSAWEII
Δ1-10	131	AIAAVIAIGILVYKN / WEIIKTVVMGPVLLL
Δ24-26	40	VVDSANAIVNGVKQG / FNNLKNINLLDIGKA
Δ20-26	84	VINMAKSIVDGIVNG / FNNLKNINLLDIGKA
Δ24-29	73	VVDSANAIVNGVKQG / FISGIGDWIRKHKGP
Δ20-29	117	VINMAKSIVDGIVNG / FISGIGDWIRKHKGP
ΔTm3	19	GLALIGGSFTAISLP / KNWDKISKFGKEVWA
ΔTm2-3	43	LLVIFGQAVVTLQRV / KNWDKISKFGKEVWA
ΔTm4	23	ILDFFSGLWGQVQTI / TGDFNQFKKDFAMLW
ΔN-terminus	30	M / EKNTNSTMSTVGKVA
ΔC-terminus	30	SHDVNLNQGKQPASF / *

1 Site of deletion is indicated by a forward-slash (/)

2 Stop codon represented by an asterisks symbol (*)

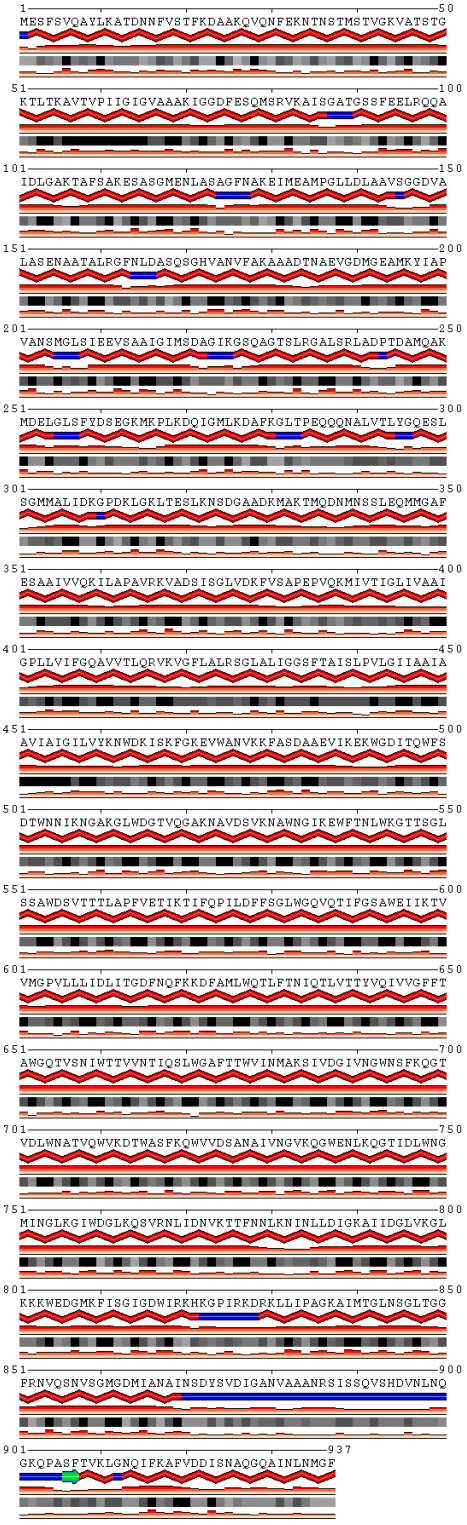
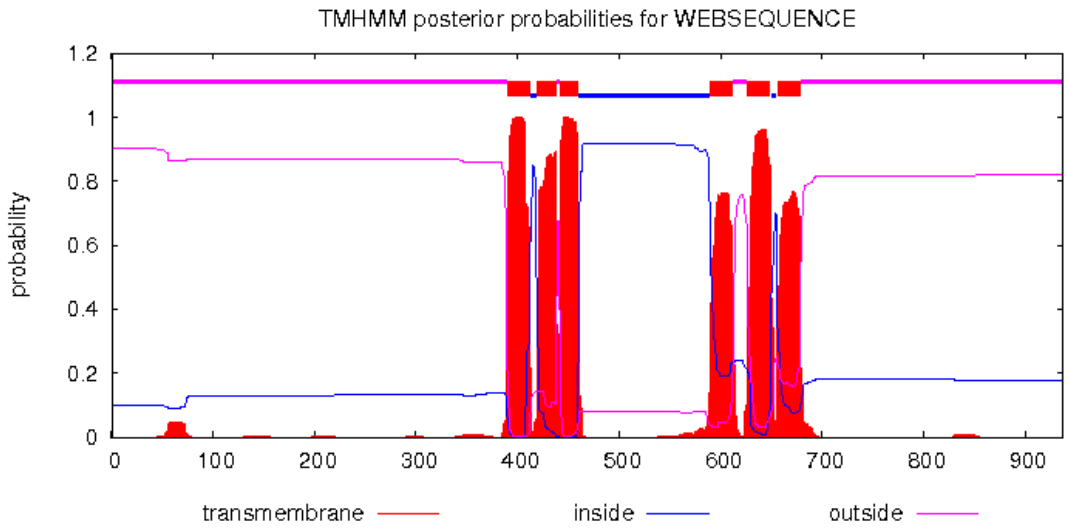


Figure 1. SABLE secondary structure prediction of $TMP_{TP901\text{-}1}$. The protein's secondary structure is predominately α-helical, in red. The C-terminal extremity has a short coiled section, indicated in blue, and a predicted β-sheet, indicated in green, before ending with more α-helices.

>$TMP_{TP901-1}$
MESFSVQAYLKATDNNFVSTFKDAAKQVQNFEKNTNSTMSTVGKVATSTGKTLTKAVTVPIIGIGVAA
AKIGGDFESQMSRVKAISGATGSSFEELRQQAIDLGAKTAFSAKESASGMENLASAGFNAKEIMEAMP
GLLDLAAVSGGDVALASENAATALRGFNLDASQSGHVANVFAKAAADTNAEVGDMGEAMKYIAPVANS
MGLSIEEVSAAIGIMSDAGIKGSQAGTSLRGALSRLADPTDAMQAKMDELGLSFYDSEGKMKPLKDQI
GMLKDAFKGLTPEQQQNALVTLYGQESLSGMMALIDKGPDKLGKLTESLKNSDGAADKMAKTMQDNMN
SSLEQMMGAFESAAIVVQKILAPAVRKVADSISGLVDKFVSAPEPVQKMIVTIGLIVAAIGPLLVIFG
QAVVTLQRVKVGFLALRSGLALIGGSFTAISLPVLGIIAAIAAVIAIGILVYKN
1. WDKISKFGKEV
2. WANVKKFASDAAEVIKEK
3. WGDITQWFSDT
4. WNNIKNGAKGL
5. WDGTVQGAKNAVDSVKNA
6. WNGIKEWFTNL
7. WKGTTSGLSSA
8. WDSVTTTLAPFVETIKTI
9. FQPILDFFSGL
10. WGQVQTIFGSA
11. WEIIKTVVMGP
12. VLLLIDLITGD
13. FNQFKKDFAML
14. WQTLFTNIQTL
15. VTTYVQIVVGF
16. FTAWGQTVSNI
17. WTTVVNTIQSL
18. WGAFTTWVINM
19. AKSIVDGIVNG
20. WNSFKQGTVDL
21. WNATVQWVKDT
22. WASFKQWVVDS
23. ANAIVNGVKQG
24. WENLKQGTIDL
25. WNGMINGLKGI
26. WDGLKQSVRNLIDNVKTT
27. FNNLKNINLLD
28. IGKAIIDGLVK
29. GLKKKWEDGMK
FISGIGDWIRKHKGPIRKDRKLLIPAGKAIMTGLNSGLTGGFRNVQSNVSGMGDMIANAINSDYSVDI
GANVAAANRSISSQVSHDVNLNQGKQPASFTVKLGNQIFKAFVDDISNAQGQAINLNMGF

Figure 2. Analysis of $TMP_{TP901-1}$ amino acid sequence. Twenty-nine aromatic-residue-containing repeat sequences were putatively deduced by manually aligning tryptophan (W) and phenylalanine (F) amino acids with 11 or 18 spaced periodicity. All tryptophan amino acids in TP901-1 TMP sequence are highlighted in yellow, and phenylalanine amino acids within the putative repeat sequences' region are highlighted in red. Transmembrane spanning regions are underlined once and twice, alternatively (see Fig. 3).

(A)



(B)

>$TMP_{TP901-1}$

MESFSVQAYLKATDNNFVSTFKDAAKQVQNFEKNTNSTMSTVGKVATSTGKTLTKAVTVPIIGIGVAA
AKIGGDFESQMSRVKAISGATGSSFEELRQQAIDLGAKTAFSAKESASGMENLASAGFNAKEIMEAMP
GLLDLAAVSGGDVALASENAATALRGFNLDASQSGHVANVFAKAAADTNAEVGDMGEAMKYIAPVANS
MGLSIEEVSAAIGIMSDAGIKGSQAGTSLRGALSRLADPTDAMQAKMDELGLSFYDSEGKMKPLKDQI
GMLKDAFKGLTPEQQQNALVTLYGQESLSGMMALIDKGPDKLGKLTESLKNSDGAADKMAKTMQDNMN
SSLEQMMGAFESAAIVVQKILAPAVRKVADSISGLVDKFVSAPEPVQKM

1. IVTIGLIVAAIGPLLVIFGQAVVTLQRVK
2. VGFLALRSGLALIGGSFTAISLP
3. VLGIIAAIAAVIAIGILVY

KNWDKISKFGKEVWANVKKFASDAAEVIKEKWGDITQWFSDTWNNIKNGAKGLWDGTVQGAKNAVDSV
KNAWNGIKEWFTNLWKGTTSGLSSAWDSVTTTLAPFVETIKTIFQPILDFFSGLWGQVQTI

4. FGSAWEIIKTVVMGPVLLLIDLITGDFNQFKKDFAML
5. WQTLFTNIQTLVTTYVQIVVGFFTAWGQT
6. VSNIWTTVVNTIQSLWGAFTTWV

INMAKSIVDGIVNGWNSFKQGTVDLWNATVQWVKDTWASFKQWVVDSANAIVNGVKQGWENLKQGTID
LWNGMINGLKGIWDGLKQSVRNLIDNVKTTFNNLKNINLLDIGKAIIDGLVKGLKKKWEDGMKFISGI
GDWIRKHKGPIRKDRKLLIPAGKAIMTGLNSGLTGGFRNVQSNVSGMGDMIANAINSDYSVDIGANVA
AANRSISSQVSHDVNLNQGKQPASFTVKLGNQIFKAFVDDISNAQGQAINLNMGF

Figure 3. $TMP_{TP901-1}$ transmembrane (Tm) spanning region predictions. **(A)** TMHMM analysis of TP901-1 TMP putatively predicts six Tm spanning regions. **(B)** Amino acids of the Tm spanning regions predicted inside or outside the bacterial cytoplasm are highlighted yellow and red, respectively.

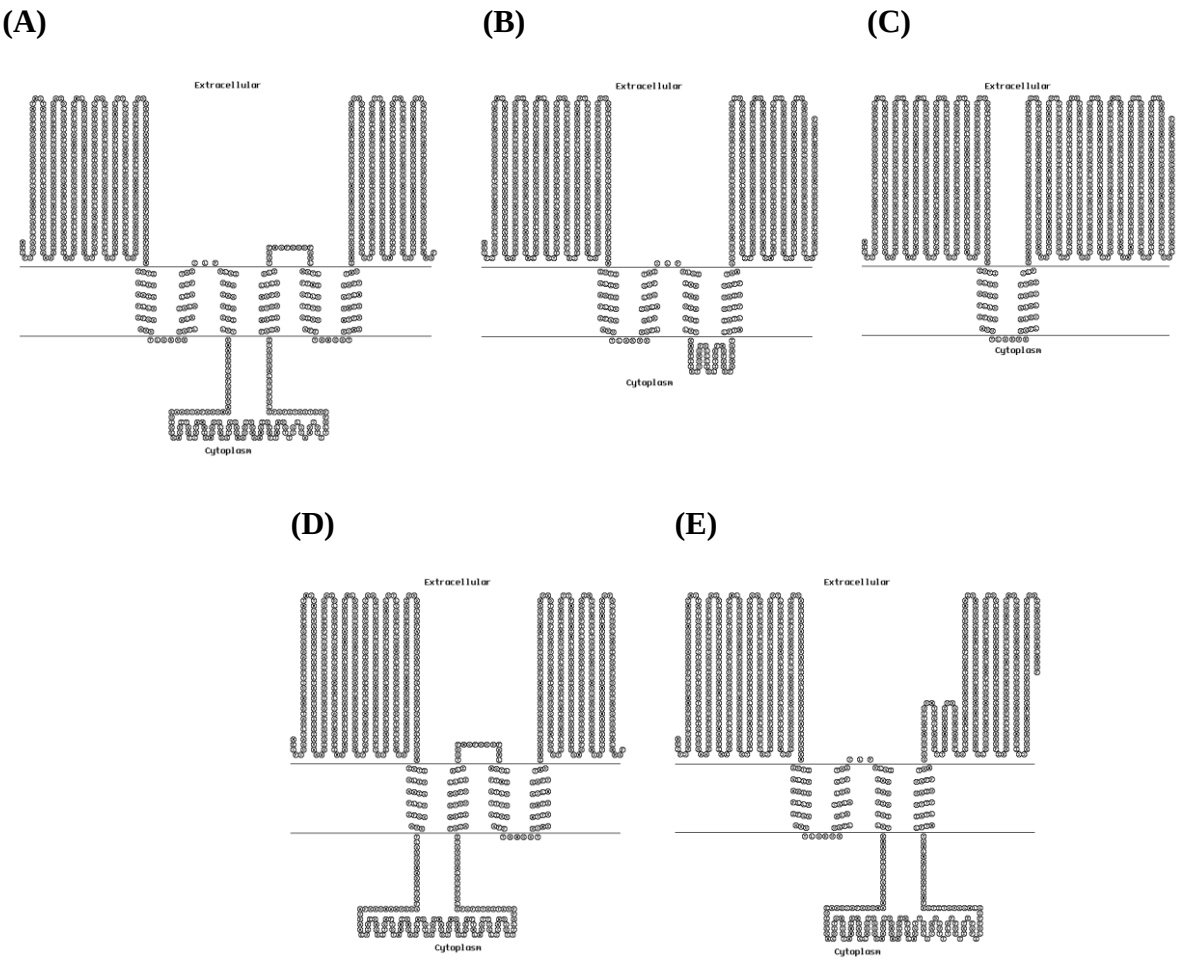


Figure 4. Predicted Transmembrane (Tm) topology of **(A)** $TMP_{TP901-1}$ and **(B-E)** mutant derivatives which were forecast by TOPO2 to adopt an altered Tm topology. **(B)** TMP repeat deletion mutant Δ1-10 displays altered Tm topology, as the end of the tenth annotated repeat sequence overlaps with the start of the forth Tm spanning region (see Fig. 2). **(C)** TMP deletion of the annotated third Tm spanning region, ΔTm3, results in a dramatically altered transmembrane topology. **(D)** TMP deletion ΔTm2-3 was created to restore transmembrane topology of ΔTm3; however, this mutation does not restore the mutant's plaque forming ability. **(E)** Deletion of the forth annotated Tm spanning region, ΔTm4, displays an altered transmembrane topology.



Figure 5. ClustalW alignment of lactococcal phage TP901-1 and Tuc2009 TMP sequences. Tuc2009 TMP, relative to TP901-1, has an addition 88 amino acids which obeys the 11 amino acid periodicity of the aromatic-residue-containing repeat sequences (highlighted yellow); however, two of $TMP_{Tuc2009}$'s additional repeats appear to begin with an alanine. ClustalW conserved amino acids are coloured: red, small/hydrophobic; blue, acidic; magenta, basic; green, hydroxyl/sulfhydryl/amine.

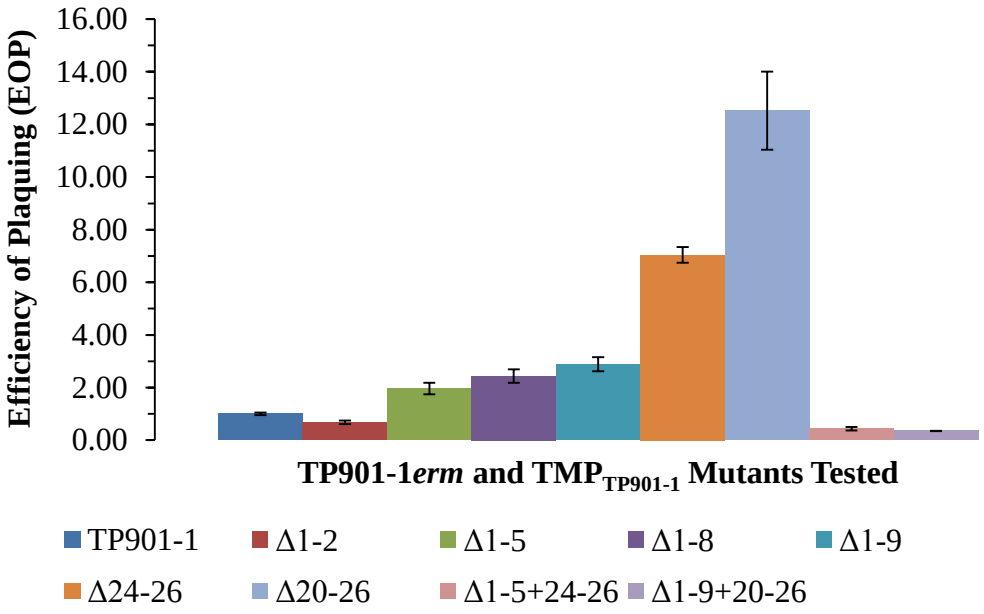


Figure 6. Efficiency of Plaquing (EOP) of TP901-1*erm* and TMP mutant derivatives which form plaques. Plaques for TP901-1*erm* TMP mutant Δ1-9+24-26 were detectable when the bacterial lawn was challenged with a presumed high titre of phages, but plaques in isolation were too small to accurately count. Results are summarized in Table 1.

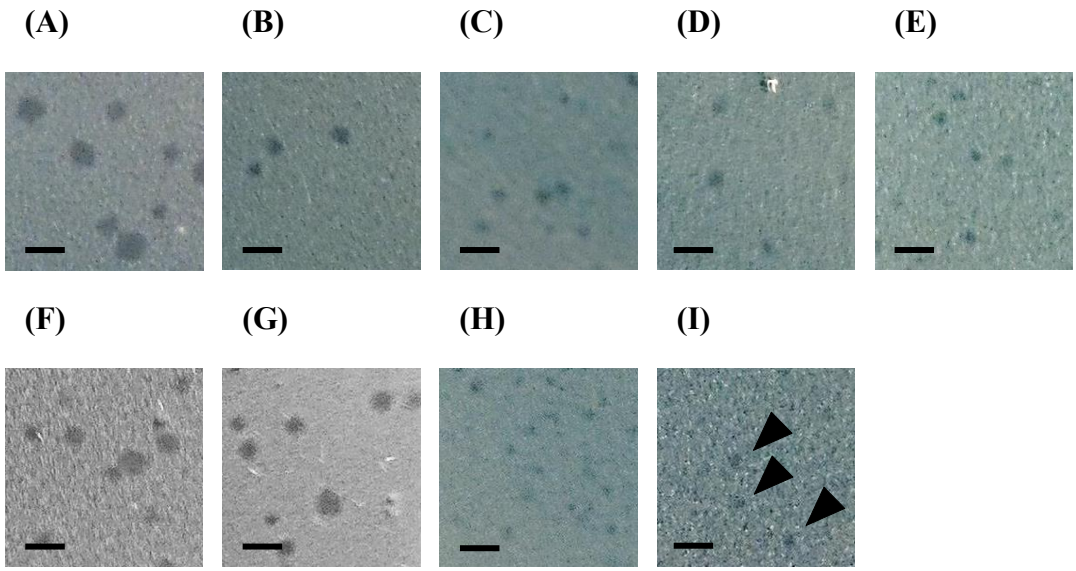


Figure 7. Photographs of plaques formed by **(A)** TP901-1*erm* and **(B-I)** $TMP_{TP901-1}$ mutant derivatives. **(A)** Wild-type phage TP901-1*erm* forms on average the largest plaques, while plaques decrease in size for mutants **(B)** Δ1-2, **(C)** Δ1-5, **(D)** Δ1-8 and **(E)** Δ1-9. Plaque sizes of mutants **(F)** Δ24-26 and **(G)** Δ20-26 are comparable to the wild-type plaque size. However, mutant **(H)** Δ1-5+24-26 forms extremely small plaques, while mutant **(I)** Δ1-9+20-26 are barely visible (examples of several plaques indicated by black-filled triangles). Images were standardised with each scale bar representing 2 mm.

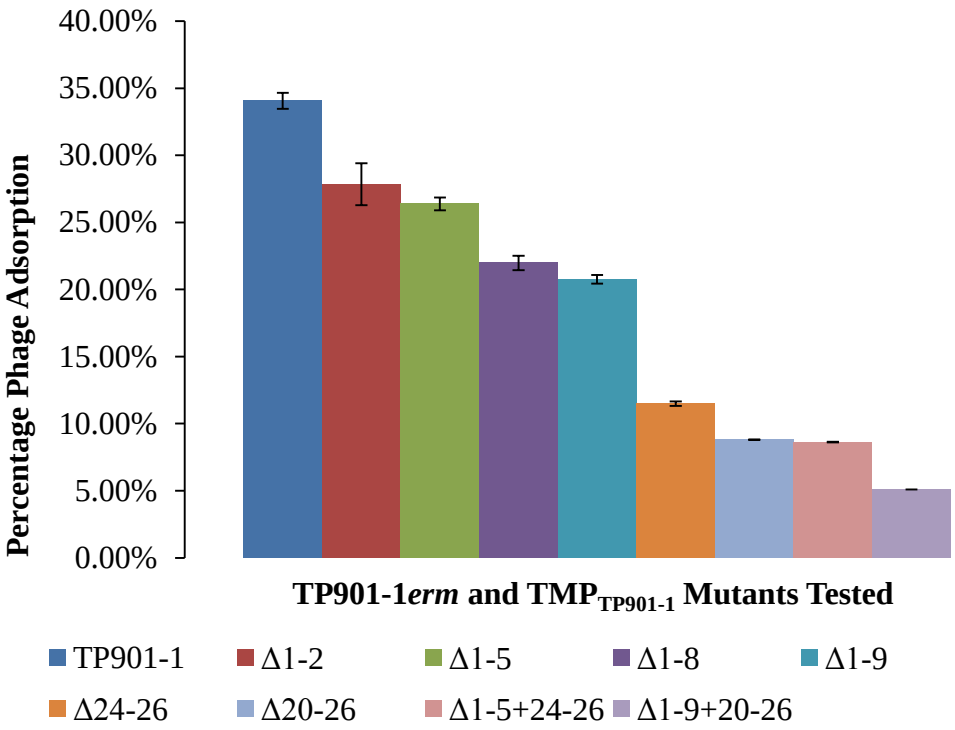
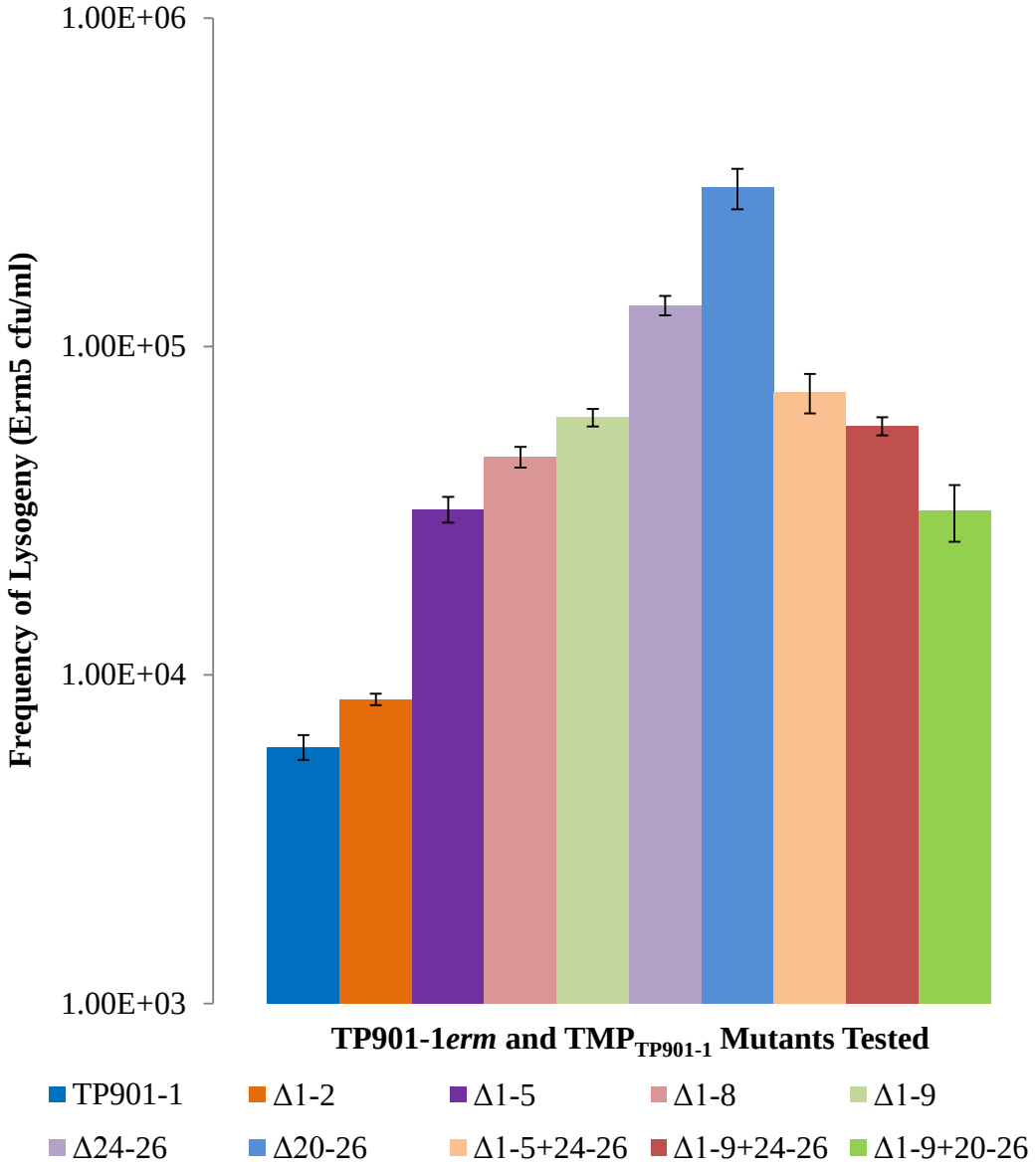


Figure 8. Adsorption assays of TP901-1*erm* and $TMP_{TP901-1}$ mutants which form countable plaques. Results are summarized in Table 1.

(A)


1.00E+06
1.00E+05
1.00E+04
1.00E+03
Frequency of Lysogeny (Erm5 cfu/ml)
TP901-1ermand TMPTP901-1 Mutants Tested
TP901-1
Δ1-2
Δ1-5
Δ1-8
Δ1-9
Δ24-26
Δ20-26
Δ1-5+24-26
Δ1-9+24-26
Δ1-9+20-26

(B)

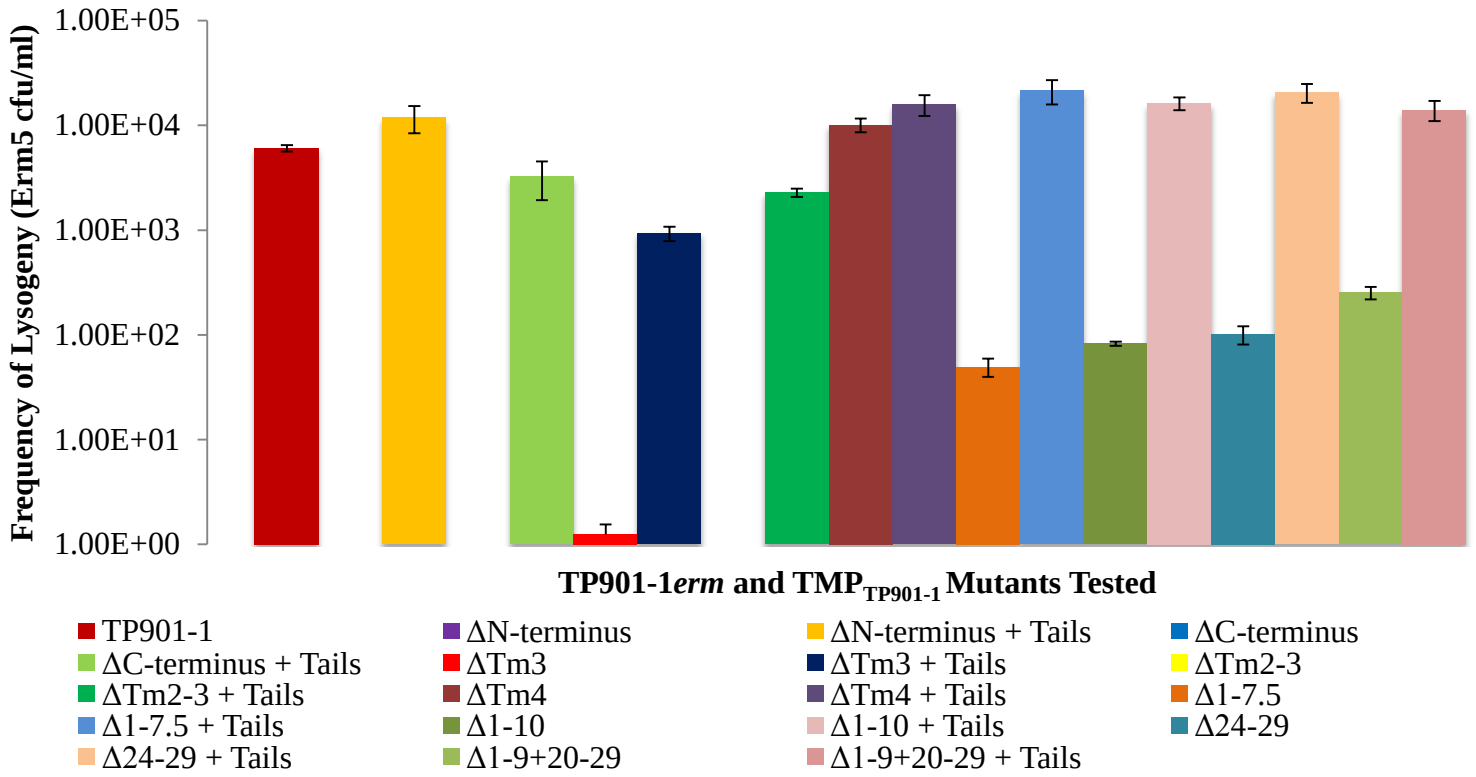


Figure 9. Frequency of lysogeny of TP901-1*erm* and $TMP_{TP901-1}$ mutant derivatives. **(A)** Frequency of lysogeny of TP901-1*erm* and plaque forming $TMP_{TP901-1}$ mutants. **(B)** Frequency of lysogeny of non-plaque forming $TMP_{TP901-1}$ mutants, with and without the donation of phage tails. Results are summarized in Table 1.

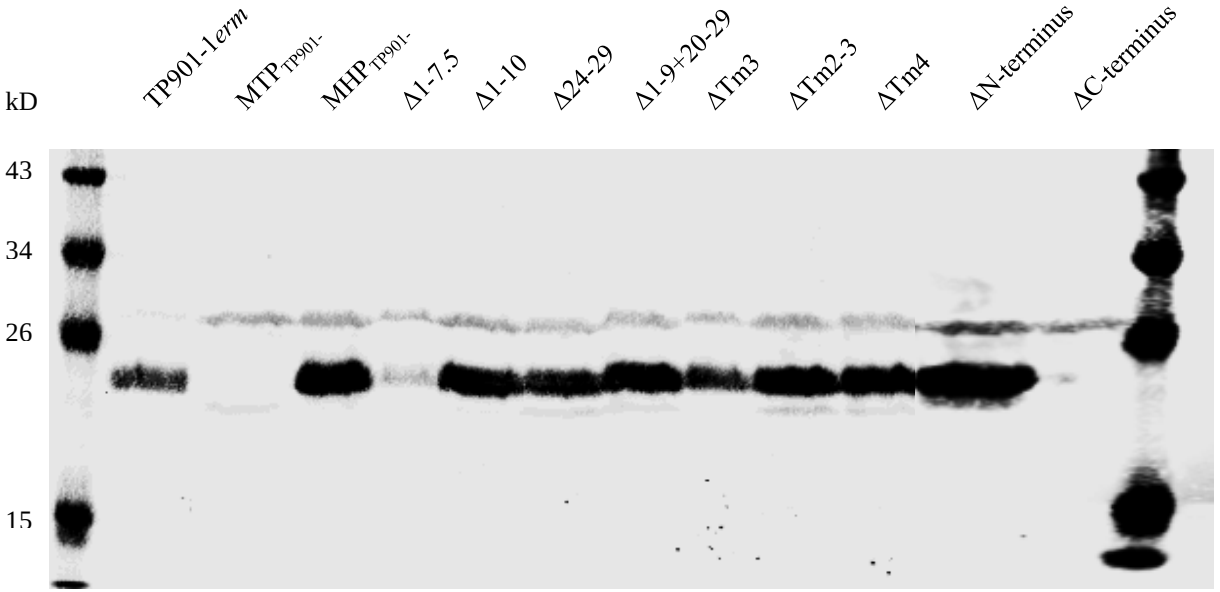


Figure 10. Western blot detection of TP901-1 major tail protein (MTP) for TP901-1*erm* TMP mutants which produced no plaques. MTP$_{TP901-1}$::Ter and MHP$_{TP901-1}$::Ter served as a negative and positive control for the detection of TP901-1 MTP, and *vice versa* for the detection of TP901-1 Major Head Protein (MHP; results not shown). Very little MTP was present in the PEG 8000 precipitate of mutant Δ1-7.5 while no MTP was detected for ΔC-terminus, suggesting aberrant tail production. All other non-plaque forming mutants of TP901-1 TMP showed the presence of MTP, indicative of functional tails produced.

CHAPTER V

General Discussion

The main aim of this PhD study was to investigate *Siphoviridae* structural proteins, and primarily proteins comprising their tail apparatus, in order to characterize their role in commencing infection of their bacterial host. The chapters of thesis describe; (i) the Tail-Associated Lysin (Tal) proteins of lactococcal-infecting TP901-1 and Tuc2009 phages, and their pertinence to degrading the bacterial cell wall for infection; (ii) the structural module-encoded proteins of TP901-1 and their assembly and importance in creating an infectious virion; and (iii) the Tape Measure Protein (TMP) of TP901-1 and its multifaceted role in creating a functional tail and in facilitating DNA injection into a host bacterium.

An important, yet often overlooked, aspect of the phage infection cycle is their ability to overcome the bacterial cell wall. Gram-negative bacteria often possess only a single-layer Peptidoglycan (PG) sacculus (1), yet many of their infecting phages encode Virion-Associated Peptidoglycan Hydrolase (VAPGH) enzymes to overcome this obstacle (2-5). Consequently, one would almost expect all phages infecting Gram-positive bacteria to possess VAPGHs to surmount the thicker Gram-positive cell wall; however, this is not the case.

The characterization of the Tal proteins of TP901-1 and Tuc2009 yielded results consistent with findings by Molineux *et al.* (2001) and Piuri *et al.* (2006) (6, 7): VAPGH enzymes enhance phage infection during environmental and physiological conditions which result in greater cross-linkage of the bacterial cell wall. Thus, TP901-1 and Tuc2009 virions, which possess a lytic domain associated

with their Tal proteins, are primed to infect lactococcal cells with increased cross-linkage of their peptidoglycan as they enter stationary phase growth.

TP901-1 and Tuc2009 may also produce tails with a truncated tail fibre protein lacking the C-terminal M23 peptidase domain which is required for lytic activity. By analysing a mutant of TP901-1 which exclusively incorporates a truncated form of its Tal protein into their virion, and a mutant that produced only full length Tal proteins, it became apparent that under optimal growth conditions the TP901-1 mutant lacking the lytic domain adsorbed to its host more efficiently compared to the TP901-1 Tal mutant with the lytic domain. Similarly, by reversing conditions, the TP901-1 Tal mutant incorporating full length-only Tal proteins into its virion was better at initiating infection of bacteria with increased PG cross-linkage. From these results, we speculate that TP901-1 naturally produces two alternate forms of its tail fibre protein to create a mixture of virions specialized at efficiently infecting its host regardless of the bacterium's cell wall cross-linkage status.

There are several aspects of TP901-1 and Tuc2009 Tal proteins which should be further investigated to generate a more comprehensive molecular model of *Siphoviridae* PG degradation and host penetration at the onset of infection. At approximately the same time Chapter 2 of this thesis was published, a research article by Harada *et al.* was published which detailed a metal-binding capacity for phage Mu tail spike protein. In brief, this aforementioned research demonstrated that calcium ions present at the tip of phage Mu's tail spike, gp45, directed the phage's

cell penetrating needle towards the bacterial cell membrane (8). Despite phage Mu exhibiting *Myoviridae* morphology, HHPred structural analysis of TP901-1 and Tuc2009 Tal protein revealed a putative similarity to the calcium-binding domain of Mu gp45 (structure 3VTO_A; *p*-value 0.0041). Indeed, comparative analysis of Mu gp45 and $Tal_{901\text{-}1}$ and Tal_{2009} at the amino acid sequence-level revealed conserved amino acids which are important in Mu gp45 calcium binding. The region of TP901-1 and Tuc2009 Tal protein similar to the calcium binding domain of Mu gp45 immediately precedes the proteolysis site of $Tal_{901\text{-}1}$ and Tal_{2009} that is responsible for generating the truncated form of their Tal proteins (9). This is interesting as regardless of TP901-1 and Tuc2009's strategy to produce full-length and truncated tail fibre proteins, calcium ions immediately preceding the proteolysis site may help to direct the phage's tail tip perpendicular to the bacterial membrane. Therefore, mutations in $Tal_{901\text{-}1}$ amino acid residues Asp600 and Ser601, utilising previous TP901-1 mutants which incorporate into their virions the full length only and the truncated form of the Tal protein, may further our understanding of the host-penetration mechanism utilised by lactococcal-infecting phages as a prerequisite to efficient genome ejection.

Virions of lactococcal *Siphoviridae* phages represent complex quaternary structures assembled from multiple proteins and subunits thereof. The assembly of infectious viral particles occurs within the bacterial cell and, like a machine, phage progeny are designed and programmed to adsorb to, and initiate infection of a new host following lysis. The complexity of phage virions is indicative of a common evolutionary origin; and in agreement with this, the genomic organisation and

predicted structures of TP901-1 proteins constituting its virion were found to be conserved with phages infecting both Gram-positive and Gram-negative bacteria.

Chapter 3 of this thesis takes a broad experimental look at the presumed structural module of phages TP901-1 and Tuc2009. However, there is a lot more to learn by mutational probing the structural module-encoded proteins at the amino acid level. In addition, now that the exact boundaries of the capsid and tail-encoding regions of TP901-1 and Tuc2009 have been determined, and each of these found to be < 10 kb in length, it may be possible to separately clone these regions into a plasmid and overexpress these structures, as has been done for phage λ (10, 11). For example, the overexpression of complete TP901-1 tail structures, without the added physiological stress on the host bacterium to replicate the phage's genome and capsid components, may facilitate increased production and subsequent purification of tails for further analyses. An alternative strategy to produce and consequently purify TP901-1 capsid and tail structures, without the physiological stress of replicating a complete infectious phage, is achievable through recombineering mutagenesis of TP901-1 genome. Via such a recombineering methodology, it should be possible to introduce the necessary genetic sequence, at the N- or C-terminus of phage encoded-proteins, to tag proteins with 6 histidine residues for Ni-NTA column-affinity purification. For example, by introducing a non-sense mutation into TP901-1 Major Capsid Protein (MCP), and adding a hexa-histidine tag to the Major Tail Protein (MTP), this may offer a novel mechanism to generate pure and functional tail structures.

Analysis of the TP901-1 and Tuc2009 genomes revealed that almost half of their genomes encode proteins required for the assembly of their respective virion structures. Mutational analysis of the TP901-1 genome highlighted that virtually all targeted genes within the presumed structural module are required to form a fully functional infectious particle. As TP901-1 and Tuc2009 genomes appear to contain very little 'junk' DNA, combined with the knowledge that the genomic order of functionally analogous structural genes is conserved despite sequence divergence, this makes TP901-1 and Tuc2009 ideal models to annotate and assign functions to genes in newly sequenced phage genomes which do not necessarily infect lactococci.

Enterobacterial phage λ continues to serve as a prototypical model to enhance our understanding of *Siphoviridae*. The efficient recombineering mutagenesis strategy, as applicable to TP901-1 (12, 13), has demonstrated that many features of phage λ are conserved between this Gram-negative-infecting phage and TP901-1. For instance, the programmed ribosomal translational frame-shift required to produce chaperone proteins gpG and gpGT is now experimentally proven as conserved between these phages (14). In addition, the independent production of, and subsequent ability to assembly *in vitro*, *Siphoviridae* phage capsids and tails, previously reported for phage λ (15), has now been documented for a Gram-positive infecting phage.

There are three familial classifications of *Caudovirales*, yet the *Siphoviridae* family is the only morphology for which a detailed molecular model for DNA injection into a host bacterium is lacking (16, 17). Several studies have indicated a

role of the TMP in creating a channel from the phage tail-tube tip to the bacterial cytoplasm (18-21); however, this model does not appear to be widely accepted or referenced in the literature. Previous analyses of *Siphoviridae* TMP proteins (including TP901-1) have shown how the TMP dictates the length of phage tails and acts as a scaffold for the assembly of the MTP rings which form the tail tube. However, these studies did not appear to focus on maintaining fully functional tails. Finally, the TMPs of *Siphoviridae* phages are a central component of the tail structures, and in all studied instances, essential for creating functional tails. Chapter 4 of this thesis tries to dissect the various roles of the TMP, including; (i) its roles in the assembly process of the phage tail structure, (ii) its role in dictating the length of the *Siphoviridae* phage tail, and (iii) the hypothesized role of TMP in creating a channel that covers the distance from the phage tail tip and across the membrane to reach the host's cytoplasm.

The TMP of phage TP901-1 was shown as one of three essential proteins required to assemble tail structures, the other two proteins being the Distal Tail protein (Dit) and the N-terminus of the Tal protein (22). During mutational analysis of TP901-1 TMP carried out in Chapter 4 of this thesis, we identified the orientation of the TMP in the complete tail tube structure. To the best of our knowledge, this has previously not been shown for a *Siphoviridae* phage, likely proving difficult as the TMP is not an exposed protein within complete virions. However, mutagenesis of TP901-1 TMP N-terminus produced free phage tails structures, presumably due to the lack of an interaction between the phage TMP and head-tail connector proteins associated with the phage capsid. A deletion in the C-terminus of TP901-1 TMP did not yield assembled phage tail structures, indicating the lack of an initiator complex

formed of TMP-Dit-Tal, which, as mentioned, is a key requisite for phage tail assembly.

The ground-breaking research by Katsura (1984), who demonstrated that the TMP of lambda dictates the length of a *Siphoviridae* tail, also showed shortening of lambda tails from 29 Major Tail Protein (MTP) rings to 13 MTP rings (23). However, these phage mutants were incapable of infecting their host. Successive mutations in TP901-1 TMP protein, outlined in Chapter 4 of this thesis, demonstrated how specific aromatic-residue-containing repeating units of 11 and 18 amino acids dictate the final length of the tail, and that the distance between repeating aromatic residues is important for tail function. Methods to generate the shortest possible functional tail of TP901-1 has so far indicated that 16 aromatic-residue-containing repeating units from TP901-1's TMP, out of 29 annotated repeats, can be removed without drastically altering the efficiency of plaquing of TP901-1 TMP mutants.

Potentially the tail of TP901-1 may be shortened further by additional deletions; however, additional work is required to determine if this is possible. Specifically, should this work be revisited, the annotated 29th repeat of the TP901-1 TMP should be subjected to investigation. During the manual annotation of repeat sequences with either a tryptophan or phenylalanine spaced 11 or 18 amino acids apart, the 29th repeat of TP901-1 TMP was annotated as an 11 amino acid repeat. This was performed as 11 amino acid repeats seem to occur far more frequently than the 18 amino acids repeats, the reason for which is currently unknown, and a

phenylalanine residue occurred exactly 11 amino acids from the start of the annotated 29th repeat, potentially marking the end-boundary of repeating units. However, results for mutants with deletions in repeats 27-29 were dramatically affected in their efficiency of plaquing. The 29th TP901-1 TMP aromatic-containing repeat sequence therefore may be a longer repeating unit than initially annotated, as a tryptophan is 18 amino acids following the beginning of the 29th repeating unit of TP901-1 TMP. If this is indeed the case, this raises an interesting question as to why the 7 extra amino acids of an 18 residue repeat cannot join to an 11 amino acid repeating unit and once again form an 18 amino acid repeating unit. This suggests the sequence of specific individual repeats, or the exact sequence of 11 and 18 amino acid repeats, differs and are important for functional tails.

The numerical superiority of *Siphoviridae* in environmental samples analysed is indicative of their successful replication strategy, which involves their long non-contractile tail appendages. However, there is sparse evidence to support the role of *Siphoviridae* TMP proteins tails in creating a channel for DNA ejection. The strongest argument in support of TMP proteins forming a channel for DNA injection is based on experiments conducted on phage T5. Böhm and colleagues (2001) demonstrated phage T5 could inject DNA into proteoliposomes containing the phage's proteinaceous receptor, FhuA (19). Combined with analyses showing T5's TMP, Pb2, contains transmembrane helices and inserts into black-lipid bilayer membranes, the TMP was suggested by Böhm *et al.* as the candidate protein responsible for creating a channel into bacterial membranes. Additional support for the TMP of *Siphoviridae* in forming a channel for DNA entry into a host bacterium is provided by the *Podoviridae* phage T7. The recent study by Hu *et al.* (2013)

demonstrated how phage T7 ejected proteins from its virion, like the TMP of *Siphoviridae*, and that the ejected proteins underwent extensive structural rearrangements to form a channel in the bacterial membrane, directing and protecting the phage genome from the phage tail-tip into the bacterial cytosol (16). In Chapter 4 of this thesis, our results support the TMP of *Siphoviridae* in creating a channel for DNA injection into host bacterium. Mutants deleting specific transmembrane helices, possessed MTP which is indicative of tail tubes forming; however, they were incapable of injecting their DNA into host bacteria. In order to verify the normal production of phages tails for TP901-1 TMP transmembrane mutants, and confirm the predicted effects of numerous TP901-1 structural mutants, electron micrograph (EM) images are required. Purification of TP901-1, and its mutant derivatives, for EM analysis is currently still underway as a prerequisite to publishing Chapters 3 and 4 of this thesis as research articles.

Further experiments are required to prove the TMP of *Siphoviridae* is responsible, beyond any doubt, for creating a channel in the bacterial cell membrane during phage DNA injection. Ideally, structural studies of TMP would provide the most comprehensive proof of this concept. However, as this large protein requires chaperones for its correct folding and incorporation into phage virions, as it is presumed a largely hydrophobic protein, alternative methods may be required. Using antibodies against the N-terminal region of TP901-1 TMP, which are currently under production (Collins *et al.*, unpublished results), it may be possible to detect the TMP in cell wall and membranous fractions of TP901-1's host bacterium following phage infection. A detailed molecular model of all the various steps of a *Siphoviridae* initiating infection, from adsorption to penetration and DNA injection, is currently

not available, and many of these important questions can potentially be answer using TP901-1 and Tuc2009 as prototypical models.

1. **Gan L, Chen S, Jensen GJ.** 2008. Molecular organization of Gram-negative peptidoglycan. Proc Natl Acad Sci U S A **105:**18953-18957.
2. **Coburn DL.** 2011. The Role of Bacteriophage Lambda gpK in Tail Assembly and Host Cell Entry. Ph.D. Thesis. University of Toronto.
3. **Boulanger P, Jacquot P, Plancon L, Chami M, Engel A, Parquet C, Herbeuval C, Letellier L.** 2008. Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. J Biol Chem **283:**13556-13564.
4. **Briers Y, Miroshnikov K, Chertkov O, Nekrasov A, Mesyanzhinov V, Volckaert G, Lavigne R.** 2008. The structural peptidoglycan hydrolase gp181 of bacteriophage phiKZ. Biochem Biophys Res Commun **374:**747-751.
5. **Kanamaru S, Leiman PG, Kostyuchenko VA, Chipman PR, Mesyanzhinov VV, Arisaka F, Rossmann MG.** 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature **415:**553-557.
6. **Molineux IJ.** 2001. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. Mol Microbiol **40:**1-8.
7. **Piuri M, Hatfull GF.** 2006. A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. Mol Microbiol **62:**1569-1585.
8. **Harada K, Yamashita E, Nakagawa A, Miyafusa T, Tsumoto K, Ueno T, Toyama Y, Takeda S.** 2013. Crystal structure of the C-terminal domain of

Mu phage central spike and functions of bound calcium ion. Biochim Biophys Acta **1834:**284-291.

9. **Kenny JG, McGrath S, Fitzgerald GF, van Sinderen D.** 2004. Bacteriophage Tuc2009 encodes a tail-associated cell wall-degrading activity. J Bacteriol **186:**3480-3491.

10. **Xu J, Hendrix RW, Duda RL.** 2013. Chaperone-Protein Interactions That Mediate Assembly of the Bacteriophage Lambda Tail to the Correct Length. J Mol Biol **426**:1004-1018.

11. **Xu J, Hendrix RW, Duda RL.** 2013. A Balanced Ratio of Proteins from Gene G and Frameshift-Extended Gene GT Is Required for Phage Lambda Tail Assembly. J Mol Biol **425:**3476-3487.

12. **van Pijkeren JP, Britton RA.** 2012. High efficiency recombineering in lactic acid bacteria. Nucleic Acids Res **40:**e76.

13. **van Pijkeren JP, Neoh KM, Sirias D, Findley AS, Britton RA.** 2012. Exploring optimization parameters to increase ssDNA recombineering in *Lactococcus lactis* and *Lactobacillus reuteri*. Bioengineered **3:**209-217.

14. **Xu J, Hendrix RW, Duda RL.** 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol Cell **16:**11-21.

15. **Weigle J.** 1966. Assembly of phage lambda *in vitro*. Proc Natl Acad Sci U S A **55:**1462-1466.

16. **Hu B, Margolin W, Molineux IJ, Liu J.** 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science **339:**576-579.

17. **Kostyuchenko VA, Chipman PR, Leiman PG, Arisaka F, Mesyanzhinov VV, Rossmann MG.** 2005. The tail structure of bacteriophage T4 and its mechanism of contraction. Nat Struct Mol Biol **12:**810-813.

18. **Marinelli LJ.** 2008. Characterization of the role of the Rpf motif in mycobacteriophage tape measure proteins. Ph.D. Thesis. University of Pittsburgh.

19. **Bohm J, Lambert O, Frangakis AS, Letellier L, Baumeister W, Rigaud JL.** 2001. FhuA-mediated phage genome transfer into liposomes: a cryo-electron tomography study. Curr Biol **11:**1168-1175.

20. **Roessner CA, Ihler GM.** 1986. Formation of transmembrane channels in liposomes during injection of lambda DNA. J Biol Chem **261:**386-390.

21. **Feucht A, Schmid A, Benz R, Schwarz H, Heller KJ.** 1990. Pore formation associated with the tail-tip protein pb2 of bacteriophage T5. J Biol Chem **265:**18561-18567.

22. **Vegge CS, Brondsted L, Neve H, Mc Grath S, van Sinderen D, Vogensen FK.** 2005. Structural characterization and assembly of the distal tail structure of the temperate lactococcal bacteriophage TP901-1. J Bacteriol **187:**4187-4197.

23. **Katsura I.** 1987. Determination of bacteriophage lambda tail length by a protein ruler. Nature **327:**73-75.

ACKNOWLEDGEMENTS

Simply put, without the help and support of a great number of people this thesis would have been near impossible.

First of all, I would like to thank Douwe. Thank you for the opportunity to conduct a PhD in phage biology; thank you for all your help, patience, and guidance through-out; and thank you for responding to all the emails I bombarded you with towards the end. You will be glad to know, due to a lack of recent publications, 'The Journal of Inexplicable Results' has been discontinued.

Next, I would like to thank all the postdocs and staff around the department who answered numerous – and often repetitive – questions; without this help, the more difficult techniques during the PhD would still be undone. Although there are too many postdocs/staff to name (and I would likely forget a few), I would just like to thank Jenn in particular. Your help in getting me started in the lab, and continually sparing time to help with everything and anything, has been amazing.

One of the highlights of the PhD has been the friends made throughout. The hours necessary to finish the PhD, 'half-days' included, would have been so much more difficult without your support. There are just so many friends, associated with so many great memories, that I would like to go into detail to thank; but as I want to get this thesis written, I cannot list them all. Therefore, I would just like to say thank you – I am sure you know who you are – for being great friends throughout the PhD, and hopefully long into the future.

Very briefly, I would like to thank some friends in particular. Thanks to Stu, despite spending more time in the lab with you over the last four years than I have spent time with my own fiancée, it was always good craic (although I don't think I could tolerate getting hit with that Disney pen one more time!). Thanks to James, for some of the funniest moments and more than a few phrases that I must now unlearn – 'could you face that'. Thanks to Phil, despite being the newest lab member, you settled into the general lab humour very quickly and always made it a good laugh. Amy wasn't too pleased with 'scone breaks', but they were needed. Finally, thanks to Annaleen and Joanna, and many others, for the general banter which made the days enjoyable.

To my parents, for your unwavering support in everything I do, thank you so much. Despite endless rants, on topics I am sure you have no clue, you always listened and that was often all that was needed. I have no idea how I will ever repay everything you have done, but when the time comes, I promise I will pick you a nice retirement home.

Finally, to Amy, how can I even thank and praise you enough for everything? Never once have you complained of the long hours, my general lack of contribution to housework, and many, MANY other things which I have neglected in order to finish the PhD. Luckily, I know there are many years ahead during which I can return the favour – which I cannot wait for. I simply could not have asked for more from someone – 'I mo do'.