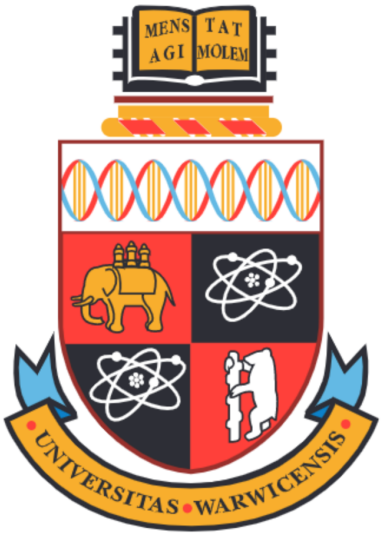


Engineering Bacteriophage K1F to Develop a Novel Point-of-Care *Escherichia coli* K1 Detection System

by

Joseph Wheatley

Thesis

Submitted to the University of Warwick in partial fulfilment of

the requirements

for the degree of

Doctor of Philosophy

School of Engineering

June 2022

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank my supervisors, Dr. Vishwesh Kulkarni and Dr. Antonia Sagona, for first believing in me and providing me with this fantastic opportunity. Also, to Dr. Peter Gammon, whose kind pieces of advice and support over the years have been hugely appreciated, and to Dr. Tamás Fehér, for always being available to discuss any technical obstacle, no matter how big or small, with such inspiring enthusiasm and knowledge. I would also like to thank all of the students, lab members and post-docs I have had the pleasure of working with, both at the School of Engineering and School of Life Sciences. In particular, to (soon to be Dr.) Sahan Liyanagedera, for all of the long hours shared in the lab together and for your continual, ever-present support throughout this PhD.

This thesis is dedicated to my amazing family. To my parents and sisters, for always being there for me and making any goal or ambition seem possible. To my beautiful children: Owen, for being the most hilarious, intelligent and caring little boy, you are always eager to dive into an in-depth science-based chat - you must be up there as one of the only 5-year old bacteriophage experts around, and Florence, your recent arrival has brought such happiness to our lives, you really are a little bundle of pure joy and have provided the perfect last ounce of motivation for me to finish this PhD. Finally, to Loren, without you none of this would be possible, you keep our lovely family going and have done so much for me over the past few years, I am forever indebted to you.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented was carried out solely by the author except in the cases outlined below:

- The cell-free crude extract optimisation work presented in Chapter 3 was carried out in collaboration with Sahan Liyanagedera, University of Warwick.
- The mass spectrometry protein analysis work presented in Chapter 4 was carried out in collaboration with Dr Cleidiane Zampronio, University of Warwick.
- The electron microscopy phage analysis work presented in Chapter 5 was carried out in collaboration with Ian Hands-Portman, University of Warwick.

Parts of this thesis have been published by the author:

J. Wheatley, S. Liyanagedera, R. Amaee, A. Sagona, and V. Kulkarni. Synthetic Biology for the Rapid, Precise and Compliant Detection of Microbes. In: V. Singh (eds) *Advances in Synthetic Biology*, Springer, 2020.

S. Liyanagedera, J. Williams, J. Wheatley, A. Biketova, M. Hasan, A. Sagona, K. Purdy, R. Puxty, T. Fehér, and V. Kulkarni. SpyPhage: A Cell-Free TXTL Platform for Rapid Engineering of Targeted Phage Therapies. *ACS Synthetic Biology*, 2022.

Abstract

The internal capsid proteins that reside within the phage K1F capsid structure hold high potential for being used as sensitive and reliable diagnostic tools. The concealed nature of the capsid interior ensures that any encapsulated signal or signal-generating enzyme, e.g. fused to an internal capsid protein, is suppressed whilst the phage is unaccompanied by its host. Furthermore, the only naturally-occurring mechanism for releasing the internal capsid proteins and therefore exposing their amalgamated signal/enzyme, is for them to be passed through the tail and subsequently ejected out of the phage - a post-adsorption phenomenon which exclusively occurs when the host is present – thus presenting a precise model for signal/enzyme release only upon pathogen presence. Here, a small N-terminal subunit of the NanoLuc luciferase is fused and incorporated into the K1F internal capsid structure using a simple, non-genomic method. This internalised subunit is exposed to the test solution containing its C-terminal counterpart (spontaneous complementation immediately forms the full NanoLuc enzyme) and substrate (furimazine) only when the K1F host, *E. coli* K1 or lab strain EV36, is present – thereby presenting a novel method for rapidly detecting this disease-causing pathogen. Finally, it is expected that by building upon this internal capsid protein engineering approach, which completely bypasses the time-consuming processes of intracellular nucleic acid transcription and translation, an unprecedentedly rapid detection device can be developed for an array of bacterial pathogens.

Abbreviations

AMR	Antimicrobial resistance
DNA	Deoxyribonucleic acid
CcP	Cytochrome c Peroxidase
GFP	Green fluorescent protein
PCR	Polymerase chain reaction
ATP	Adenosine triphosphate
SoD	Speed of detection
CRISPR	Clustered regularly interspaced short palindromic repeats
RNA	Ribonucleic acid
Cas	CRISPR associated protein
g	Gene
gp	Gene protein
TXTL	Transcription-translation
LPS	Lipopolysaccharide
EPS	Exopolysaccharide
CPS	Capsular polysaccharide
TLR4	Toll-like receptor 4
N-CAM	Neural cell adhesion molecule
ds	Double stranded
ICP	Internal capsid protein
WT	Wild type
PFU	Plaque forming units
gp16-N	N-terminal gp16
LTase	Lytic transglycosylase
gp16-C	C-terminal gp16
NanoLuc	Nanoluciferase

CtNL C-terminal NanoLuc

NtNL N-terminal NanoLuc

HR Homologous recombination

PAM Protospacer adjacent motif

OD Optical density

RFU Relative fluorescence units

Mg-glu Magnesium glutamate

K-glu Potassium glutamate

DTT Dithiothreitol

PEG 8k Polyethylene glycol 8000

RLU Relative light units

MOI Multiplicity of infection

ExRec Expression-assisted recombination

MS Mass spectrometry

ns Not significant

EoP Efficiency of plating

ELISA Enzyme linked immunosorbent assay

AuNP Gold nanoparticle

NASBA Nucleic acid sequence-based amplification

LAMP Loop-mediated isothermal amplification

RPA Recombinase polymerase amplification

GP General practice

LFT Lateral flow test

RT Room temperature

EM Electron microscopy

PAP Procapsid accumulation phenomenon

LoD Limit of detection

SN Sequence number

Chapter 1

Introduction

1.1 The bacteriophage

A bacteriophage, or phage, is a type of virus that infects bacteria. The relationship between phage and their cognate bacterial prey constitutes the oldest predator-prey interaction on Earth, having existed for at least 1 billion years [1]. During this time, phage have evolved extreme specificity and sensitivity towards their hosts [2]. Today, phage are ubiquitous in all natural environments and represent the most abundant living entities on the Earth with an estimated 10^{31} in total [3]. In many cases it is the phage tail fibres, extending from their baseplate, that constitute the primary determinants of bacterial host specificity [4]. Furthermore, specific molecular motifs expressed on the bacterium surface are recognised and adsorption is initiated prior to the phage injecting their genetic material and beginning the infection process [5, 6]. Phage predominantly overlook all life except their specific bacterial hosts, upon which they are dependent for propagation on account of lacking any metabolic machinery [7]. As a group, they can infect every known type of bacterium in every known environment on Earth [8–11].

The two primary phage types are virulent and temperate. These two types of phage experience either one or both of the life cycles displayed in Figure 1.1. Virulent phage exclusively use the lytic cycle, which is a cytoplasmic replication process whereby the genetic material of the phage is injected into the host and subsequently used as a template to rapidly generate membrane/cell wall-targeting enzymes and an abundance of new, identical phage particles within the host. During this process, the transcriptional and translational machineries within the host cell are diverted from their usual activities towards the sole purpose of generating and assembling phage proteins. Shortly after the commencement of this cycle, the host cell is fatally breached and the phage progeny are released into the external environment for the perpetual cycle to continue [12]. Furthermore, whilst temperate phage do utilise the lytic cycle, infection by this phage type can also lead to lysogeny [13]. The lysogenic cycle begins in the same manner as its lytic counterpart, whereby the infecting phage injects its genetic material into the host cytoplasm. However, following on from this and instead of rapidly generating new phage particles, the genetic material of the phage is incorporated into the host chromosome - forming the prophage - and subsequently becomes part of the host cell life cycle and is passed on from one generation to the next [14] - a process that can continue indefinitely [12]. Furthermore, upon entering certain ecological conditions or receiving a molecular stimulus (this differs from one phage to the next), the lysogenic cell can be induced to re-enter the lytic cycle - thereby generating the phage progeny, bursting the host cell and restarting the lytic or lysogenic process [15].

The discovery of phage, over a century ago now, was a two-fold achievement by two scientists working independently of each other - Frederick Twort and Felix d'Herelle. Publishing his

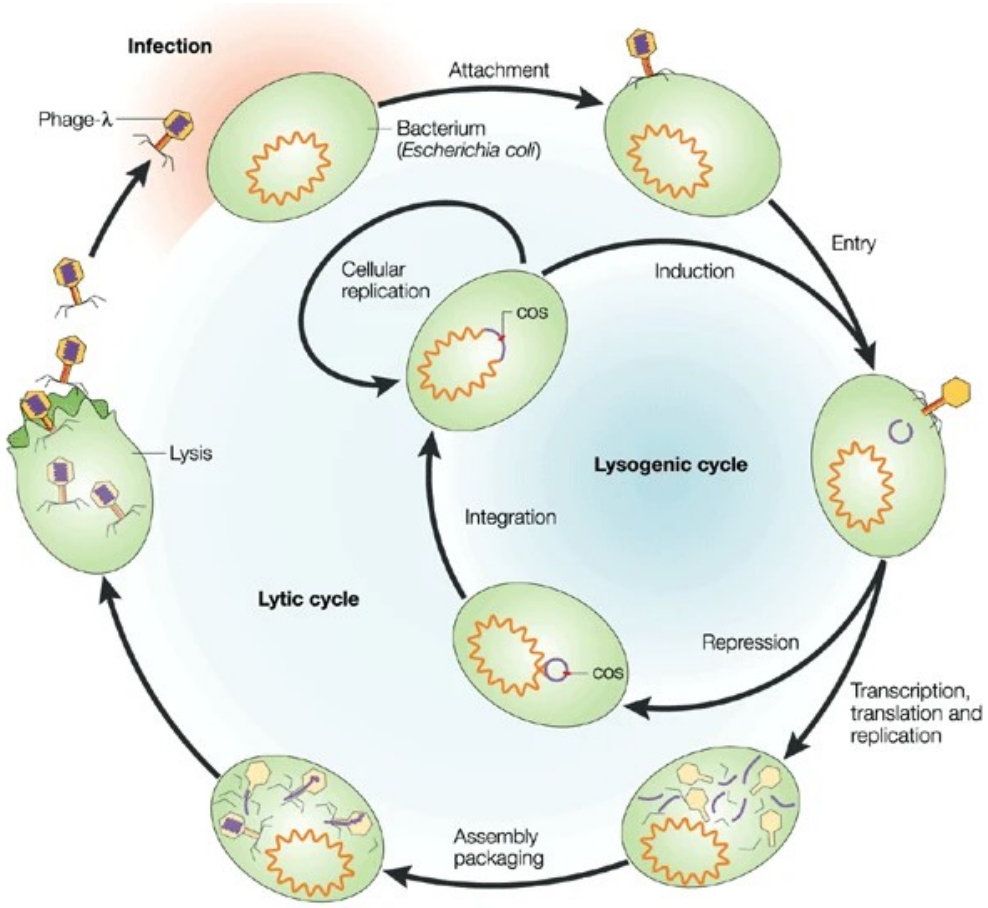


Figure 1.1: **Schematic representation of the lytic and lysogenic phage life cycles.** The life cycle of the typical temperate phage ($coliphage - \lambda$) is shown above, displaying both the lytic and lysogenic stages. This image was taken directly from Campbell (2003) [12].

initial observations in 1915, Twort was the first to notice a phenomenon whereby spots of plated bacterial cultures would periodically become “glassy and transparent”, for which he suggested may be attributed to an “ultra-microscopic virus” [16]. In a piece of independently investigated research that was published two years later, d’Herelle reported similar observations and first coined the term "bacteriophage" [17], literally meaning "bacteria eater". However, despite the initial enthusiasm and potential of this newly unveiled ultra-microscopic bacteria-eating virus, the discovery of penicillin in the 1920s by Alexander Fleming [18] and the subsequent dawn of the antibiotic age in medicine resulted in phage research attracting little attention on a global scale throughout the 20th century. Subsequently, progress in this field was somewhat thwarted. Perhaps, if the antibiotic age continued to flourish without any complications, the impending burst of phage research (including this thesis project and the research group it materialised from) would continue remain latent. However, the emergence of antimicrobial resistance (AMR) and

its identification as a severe global threat has prompted a recent revival in phage research - the phage renaissance - which can now be found at the forefront of biological research in abundance throughout academic and commercial establishments across the world.

1.2 Antimicrobial resistance

Recent statistics state that antimicrobial-resistant "superbugs" are the cause of approximately 700,000 deaths each year [19]. Furthermore, it is predicted that if the overwhelming surge of AMR is not seriously tackled over the next 30 years, then it will overtake cancer in the number of fatalities caused - with potential death tolls reaching ten million per annum [20]. AMR is a natural process, however, the misuse of antibiotics in humans and animals is dramatically accelerating the process at a dangerous rate [21]. One approach for reducing AMR is to not disburse antibiotics in the first instance unless they are completely necessary. The problem with this approach, however, is that it is impossible to know for certain whether or not antibiotics are necessary unless the microbes causing the infection are identified. Currently, apart from a limited number of exceptions, there is a lack of truly rapid (<10 mins detection) diagnostic tools available for knowing precisely what microbe has caused an infection (although the COVID-19 outbreak has resulted in some significant strides forward in terms of user acceptance of such tools). Consequently, it is common that antibiotics are prescribed in situations where they will be completely ineffective (e.g. viral infections). This perennial problem could be overcome if medical professionals would be able to rapidly diagnose which, if any, microbe has caused an infection. Armed with that information, a targeted, patient/microbe-specific (not symptom-specific) therapeutic could be administered. This would be desirable not only for the individual patient but also, in a more general sense, in the fight against AMR. In some cases, this therapeutic may well be in the form of antibiotics, however, practising within these means is likely to lead to a reduction overall in antibiotic use by eradicating their misuse [22]. Subsequently, this will have a considerably positive effect on reducing AMR worldwide.

Delays to the early, accurate diagnosis of many bacterial infections carry a significant personal and economic cost, both nosocomially and within the community. Hospital-acquired pathogenic bacterial infections are extremely common, with an annual occurrence of 4.1 million in Europe [23] and an annual mortality rate of 100,000 in the US [24]. Aside from the humanitarian toll, these infections incur a huge financial burden estimated at €7.5B and $5B in Europe and the US respectively [25]. Furthermore, it is highly desirable within all affiliated fields to have the capability to differentiate between different bacterial infections so that specific and effective treatments can be administered.

1.3 Phage-based diagnostics

Phage offer a naturally occurring chassis that can, with the tools that synthetic biology offers, be modified and optimised for use in highly specific bacterial detection devices. Even prior to any modifications, they offer many unique benefits over traditional diagnostic approaches [26–28], including:

- High specificity and sensitivity towards their cognate host
- Capacity to detect even traces of host presence
- Capacity to function with impure samples under diverse or even harsh conditions
- Discrimination between viable and incapacitated target pathogenic cells (i.e. removes false positive)
- Signal amplification capacity alongside signal transduction
- Low cost, easy propagation and purification

Furthermore, the ability to express and assemble phage from their isolated genome in a cell-free system [29–31] potentially allows for extensive control over when and how they are deployed. This ground-breaking technique also gifts researchers the ability to extensively analyse phage assembly and offers a unique premise for identifying genes completely necessary for assembly and other purposes through iterative gene knockouts. This knowledge is likely to supply ammunition for future phage-based detection systems and is something that is explored in this thesis.

Since phage are highly specific regarding their host, the phage:host interaction in itself is enough to elicit detection when coupled with either a natural or engineered diagnostic mechanism. Furthermore, regardless of the chosen diagnostic mechanism, the positive response is always limited to the target host - this is a unique advantage. Detection via lysis exposes cytoplasmic enzymes to the external medium only upon lysis. Detection via reporter genes involves engineering exogenous reporter deoxyribonucleic acid (DNA) into the phage genome and thus exclusively into the cells of a specified, target pathogen bacterial species. Non-lytic infection methods are limited to expression of phage-borne chromogenic, fluorescent or luminescent proteins. Before considering these approaches in more detail, it should be noted that the unique ability of phage to discriminate live from dead cells can be exploited to reveal antibiotic resistance profiles of target host by comparing their detection responses with and without antibiotics [32].

1.3.1 Phage genome-borne exogenous reporter genes

Phage genomes have been modified with several exogenous genes as a means of observing infection, thus confirming presence of the cognate host. Luciferase, through the engineering of luxAB genes, is perhaps the most common phage reporter system [33–35]. Furthermore, a recent innovation is the development of the NanoLuc variant, which achieves far greater signal capacity from a smaller protein [36,37]. Another similar system makes use of Cytochrome c Peroxidase (CcP). CcP catalyses the conversion of reduced cytochrome c, which is red, to oxidised cytochrome c, which is colourless. Integrating CcP into a phage genome therefore can confirm presence of cognate host through a loss of red colour [38]. The methodology of this is that, once infection has begun, the recombinant phage will express CcP via its host's transcriptional and translational machinery and upon the host being lysed by the phage progeny, the CcP will be exposed to its substrate in the extracellular medium.

Unsurprisingly, the ubiquitous reporter and marker protein, green fluorescent protein (GFP), has found utility as a phage-based diagnostic tool for many bacterial species including *E. coli* K12 [39], pathogenic O157:H7 [40], and *M. tuberculosis*. Aside from its robustness and in-depth characterisation, the main advantage of GFP over other reporter gene systems is that it uses light instead of a substrate to generate signal, thereby making it a much simpler model. Moreover, it is also stable and well tolerated within living cells [41]. In addition to the use of reporter genes carried by modified phages, it is possible to achieve detection by following increases in the concentration of phage genome via real-time quantitative polymerase chain reaction (PCR) as recently shown for *B. abortus* samples in mixed culture and spiked blood [42].

1.3.2 Detecting host enzymes in phage-induced lysate

Another diagnostic strategy relies upon detection of enzymes that are present in the extracellular medium only after lysis. The presence of such enzymes therefore can act as indication of phage activity which in turn indicates presence of the cognate host. Several enzymatic phage-based systems have been previously exhibited, including: alkaline phosphatase [43], Adenosine triphosphate (ATP) + luciferase [26, 35] and β-glycosidase [44]. Either the enzyme or substrate is present only within bacterial cells, therefore, its extracellular presence indicates cell lysis due to phage activity, allowing these engineered phage to act as a bacterial diagnostic.

Finally, it is noted that all phage-based diagnostic offerings to date rely on some form of bacterial processing - whether that be in the form of cytoplasmic expression of phage-borne reporter genes, the propagation of phage particles to incur bacterial lysis, a combination of the two or a similar approach. This bacterial processing (i.e. re-purposing the endogenous transcriptional

and translational machinery for phage protein expression) is somewhat time consuming and is the primary limiting factor impacting the speed of phage-based diagnostics. Moreover, the sensitivity and specificity of such phage-based systems have been well explored and validated - it would now be appropriate for some innovation to be applied to the speed of detection (SoD) in order to achieve a truly rapid and industry leading diagnostic assay.

1.4 Synthetic biology - bridging the gap between engineering and life sciences

Synthetic biology provides a platform that enables the rewiring of natural biological parts, thereby giving rise to optimised and predictable entities capable of performing tasks to a degree that is unmatchable in the natural world. The field makes use of engineering-like principles in order to view biological systems as computable devices that can efficiently complete complex tasks while allowing a high level of control over all parameters. The comparison of biological systems to classical engineering circuitry was first made in the early 1960s, where a suggestion was made that 'Teleonomic Mechanisms' within cells operated in a more linear fashion than had first been perceived [45]. This novel idea that biological organisms employed a command-and-control-like regulatory system, and therefore had the capacity to be externally controlled and modified much like a computational device, weaved the first philosophies that formed the foundations of synthetic biology. Some of the highlights from the synthetic biology field will be summarised below.

1.4.1 Genetic engineering with CRISPR

The clustered regularly interspaced short palindromic repeats (CRISPR) system is based on a ribonucleic acid (RNA)-directed DNA endonuclease mechanism [46] that allows bacteria to adaptively immunise themselves against harmful, invading nucleic acids (e.g. from phage or plasmids) [47]. This is achieved by storing - on the host chromosome - sequences corresponding to past encounters with harmful foreign DNA [48]. These "signatures", called spacers, are then used by CRISPR associated (Cas) proteins to detect and destroy any DNA or RNA in the cell that harbour the same sequences [49]. In this way, spacers act as memory that can be updated, hence the immunity is acquired and not merely innate [50, 51]. Spacers can target any nucleic acid sequence and thus can even cause bacterial auto-immunity [52] – although in some cases mechanisms exist to prevent self-digestion [53].

Functionally, CRISPR/Cas systems are composed of chromosomal CRISPR DNA sequences and Cas proteins [54]. Traditionally, CRISPR DNA sequences carry distinctive, repeating segments

23-55 base pairs long that are regularly positioned, interspersed with non-repeating sequences called spacers that are 26-72 base pairs long [55]. Spacers are of non-host origin and carry the memory of what sequences to destroy [50]. The repeating parts often contain palindromic sequences and are postulated to form stem and loop structures when transcribed [56]. The number of spacers in a host can be extremely variable, but no host carries more than one copy of a given spacer [57]. Flanking these interspersed sequences, there is also always a leader sequence at the 5' end and occasionally at the 3' end too. Leader sequences carry the promoters for transcription of the whole interspersed sequence as a single block [58] that is later spliced to remove most of each repeated motif, leaving small RNA fragments corresponding to the spacer sequences flanked by part of the repeat sequence [59]. Cas genes occur only in CRISPR-competent hosts and always lie adjacent to CRISPR sequences [60]. They serve various functions, including adding new spacer sequences [61], cutting DNA matching spacer RNA fragments [62] and editing target DNA [63].

1.4.2 Cell-free transcription-translation systems

The decades-old concept of a cell-free transcription-translation (TXTL) system has recently been reinvigorated and subsequently, these next-generation TXTL systems have been engineered for use in multiple applications across the field of synthetic biology (Figure 1.2). These applications include: facilitating and accelerating the design-build-test-learn cycle for genetic circuits - ranging from simple DNA circuits to complex gene networks [64–66], assembling whole phage infectious particles in a cell-free environment [29–31], metabolic engineering [67], biomanufacturing [68, 69] and medical research [70, 71]. Cell-free TXTL harnesses the endogenous transcriptional and translational machinery extracted from bacterial cells, commonly *E. coli*, and combines this cellular hardware with an energy solution and amino acid mix, allowing for the expression of DNA in a single cell-free reaction [72, 73].

Furthermore, the design and implementation of nucleic acid-based circuits is becoming one of the most rapidly growing areas within cell-free synthetic biology [74]. Genetic circuits generally comprise two domains: a sensor domain - which senses and responds to an input, and an actuator domain - which generates an output in the presence of a compatible input [75]. The output of one circuit is often designed to become the input of the next, thus, complex circuitry networks can be engineered to have a vast array of outcomes. Although a cell-free system is not completely necessary for the mere implementation of gene circuits, by utilising TXTL, circuit cascades can be designed, tested, debugged and redesigned in a matter of days – a luxury that would not be easy to achieve in the absence of a cell-free environment. This approach has immensely accelerated both the magnitude and fidelity of research in this field. Moreover, TXTL significantly expands and helps comprehend the real-world applications that are imminent for genetic circuits

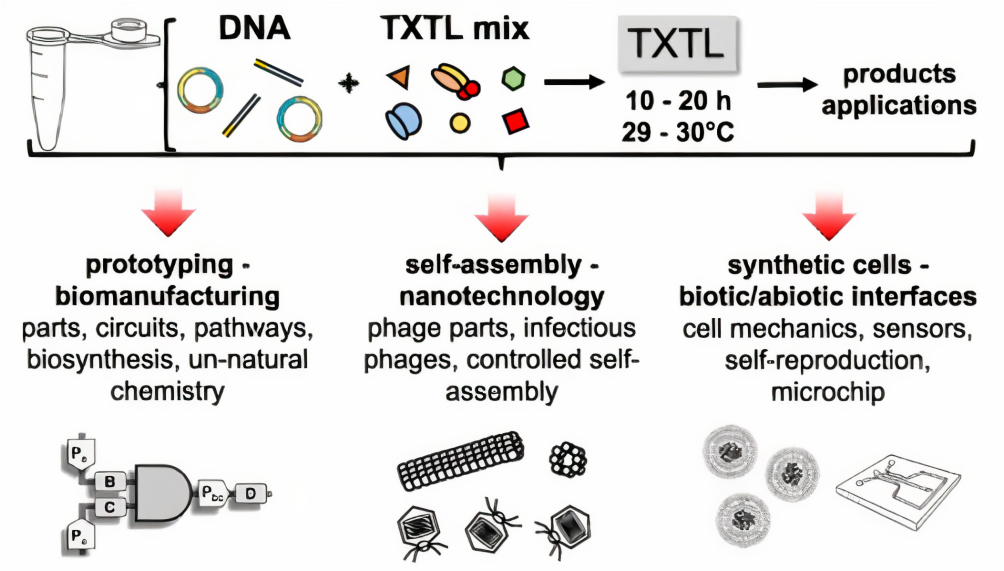


Figure 1.2: **Cell-free TXTL overview and its many applications within synthetic biology.** This image was taken directly from Garenne (2021) [74].

– with a particular relevance to microbial diagnostics. Indeed, TXTL-mediated circuits have inevitably played a role in pathogen detection in recent years, with one specific example being the detection of the Zika virus [76]. The virulence of a virus rests on its ability to transfer its genetic material into a host, then use the hosts' nucleic acid-expressing hardware to run its viral software in order to assemble its progeny, which proceed to kill and escape the host in search for a new victim to restart the cycle. One detectable constituent of a virion can be its genetic material - DNA or RNA depending on the species. As aforementioned, of the most impressive recent innovations to rise from the field of microbial detection is the lyophilised, paper-based, cell-free Zika virus RNA genome sensor [76]. This genetically programmed device, based on a toehold riboswitch, is instigated by the addition of isothermally amplified viral RNA. The riboswitch detects the presence of Zika virus RNA and through a structural change, it subsequently triggers the downstream transcription of LacZ - an enzyme that catalyses the colorimetric conversion of a yellow substrate (chlorophenol red-β-D-galactopyranoside) to a purple product (chlorophenol red) - which indicates whether Zika virus is present or not. This technology has recently been rewired to detect norovirus [77] and also γ-hydroxybutyrate - a substance used as a date-rape drug [78]. These applications using state-of-the-art synthetic biology research display the true versatility of TXTL-based microbial detection programmes.

1.4.3 Smart biomaterials

The construction of a synthetic biology-based smart material for the purpose of microbial detection requires an appreciation of the current trends and goals of the field itself. The ultimate aim of synthetic biology is to create and modify synthetic cells to understand the evolution and functionality of all life, as well as to engineer and exploit new life forms. Through the encapsulation and utilisation of TXTL systems, genetic circuits and/or engineered phage particles/genomes within biomaterials, it is conceivable that a smart biomaterial that emulates many of the functionalities of a living cell could be constructed. These synthetic entities could subsequently be engineered to perform tasks such as detecting microbes and it is the physical materialisation of these processes that provides the potential to take them from the research laboratories to being applied in the real world. The current state-of-the-art displays a two-fold approach for these purposes. The top-down approach targets the synthesis of a minimal genetic network that is able to sustain life, as well as the modulation of metabolic pathways for biotechnological applications [79]. The bottom-up approach is focused on the *de novo* design of cells, through creation of minimal structures that are capable of mimicking complex cellular functions [80, 81]. The recent advances in the field and the emergence of hydrogels as a highly biocompatible and tunable substrate has paved the way for the creation of programmable soft materials [82]. The rate of the continual expansion of the synthetic biology tool box and innovations for the incorporation of such modules into materials of interest will have a significant impact in determining the growth of this nascent field.

1.5 Bacteriophage K1F and its cognate host *Escherichia coli* K1

1.5.1 Polysaccharide-mediated pathogenicity

Similarly to many other Gram-negative bacterial species, *E. coli* cells are cocooned by cell surface glycoconjugates [83] - which are defined as biopolymers that contain one or more carbohydrate units [84]. The three core categories of *E. coli*-encapsulating glycoconjugates are: lipopolysaccharide (LPS), exopolysaccharide (EPS) and capsular polysaccharide (CPS). Together, these three glycoconjugate types serve bacterial cells in many of their key survival activities including: escaping from a host immune response, providing adhesion for bacterial colonisation and offering resistance from antibiotics [85]. LPS, which is also commonly referred to as bacterial endotoxin, is a key component of the outer membrane of Gram-negative bacterial cells [86]. The part of LPS that interacts with mammalian immunity is Lipid A - a lipid component which anchors LPS to the bacterial membrane and also binds to mammalian cell Toll-like receptor 4 (TLR4), initiating signalling pathways leading to an inflammatory immune response [87]. EPS units are loosely associated with the bacterial membrane and are often

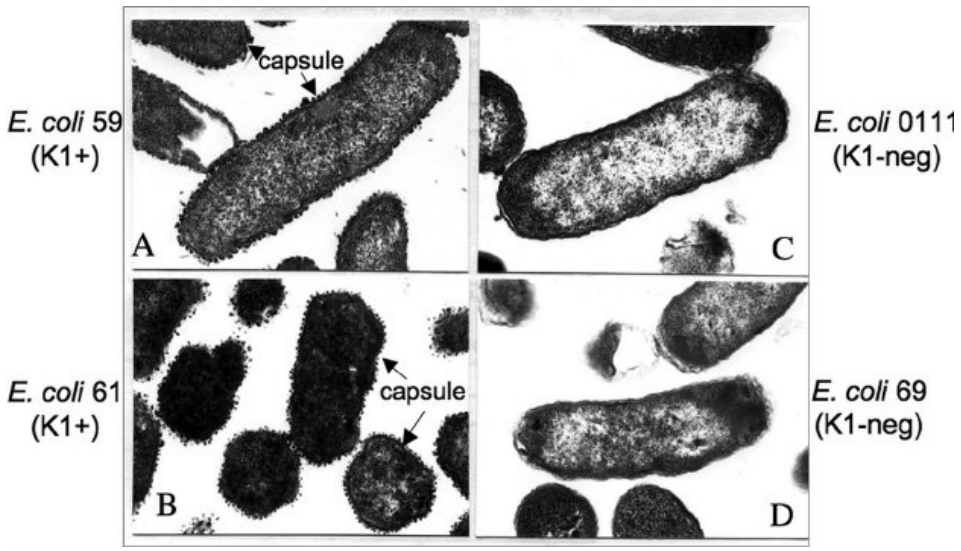


Figure 1.3: ***E. coli* cells with and without the K1 capsule.** This image was taken directly from Metkar (2007) [96].

secreted into the extracellular environment to aid in biofilm formation [88]. CPS, also known as the K-antigen, form a tight capsule which envelopes the bacterial cell and is often responsible for allowing it to roam its environment incognito [89]. This is the category that the *E. coli* K1 capsule falls into, which will be explored further below.

1.5.2 *E. coli* K1

E. coli K1 (*E. coli* O18:K1:H7) is responsible for an array of serious diseases (e.g. sepsis, neonatal meningitis, urinary tract infections and inflammatory bowel syndrome [90–93]) and much of its pathogenicity is attributed to its K1 CPS exterior - an α-2-8-linked homopolymer of sialic acid [94]. The K1 capsule, visualised in Figure 1.3 in comparison with *E. coli* cells that do not display the capsule, is hydrophilic and therefore causes the cell to have a negative charge - a well known anti-phagocytic characteristic [95]. Additionally, structural similarities between certain components of the K1 capsule and human polysaccharides enhance the strains' ability to evade an immune response. For example, the neural cell adhesion molecule (N-CAM) displays many similarities to K1 CPS [85]. These K1 capsule characteristics are the key contributors to the pathogenicity of (*E. coli* O18:K1:H7) and the subsequent diseases that it causes in humans. However, whilst this *E. coli* strain executes a sophisticated operation against human cells, its primary threat comes in the form of a much smaller entity - its evolutionary agonist, phage K1F.

1.5.3 Bacteriophage K1F

The virulent phage K1F is a member of the *Teseptimavirus* genus (also known as the T7-like phage genus) which belongs to the *Podoviridae* family of viruses in the *Caudovirales* order - "the tailed bacteriophage" [98]. Two key characteristics of phage belonging to the *Podoviridae* family are their encapsulated, linear double stranded (ds) DNA genomes and their short, non-contractile tails [99]. K1F specifically infects *E. coli* strains that display K1 CPS - the mechanism of action for this is by incorporating endosialidases within the tail structure of the phage which cleave the *alpha*-2,8 bond of the host capsule, allowing the recognition and degradation of the capsule [100]. In fact, the K1F name actually originates from the name given to the gene encoding the endosialidase enzyme for the phage [101]. Ironically, though, the exact mechanism utilised by *E. coli* K1 for human immune response evasion is also the precise instigator for phage K1F adsorption which consequently leads to the bacterial cells' own death.

The 39,704bp, dsDNA K1F genome is extremely similar to the genome of phage T7 [100]. T7 is a well studied phage and has been used as a model organism in bioengineering research for decades. Some of its key attributes are: its rapid propagation cycle, ease in displaying proteins on its capsid, high stability, strong cloning efficiency for genetic engineering and convenient storage [102]. It is assumed that, due to its similarity, phage K1F also embodies many of these

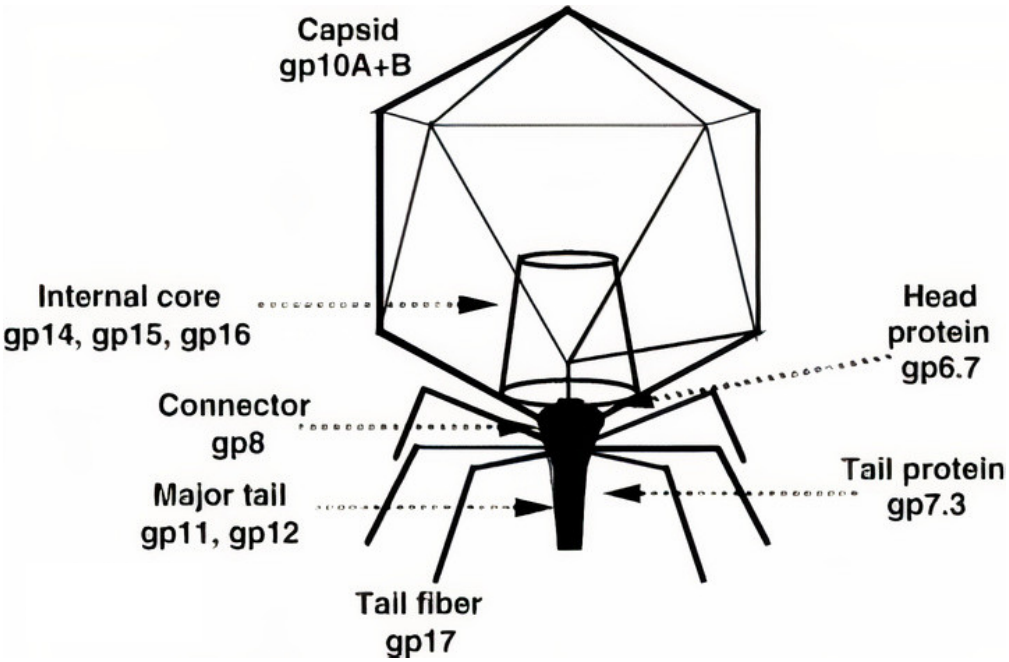


Figure 1.4: **The structure of phage T7/K1F with annotated genes of interest.** This image was taken directly from Kemp (2005) [97].

attributes - to a certain extent, this has already been demonstrated in the literature (e.g. with successful K1F engineering projects [94]). Some of the important T7 (and K1F) genes and their structural positions are displayed in Figure 1.4.

As aforementioned, the key differentiator between K1F and T7 is the inclusion of an endosialidase in the K1F tail structure. Interestingly, the N-terminal, head-binding domain of the K1F tail fiber has a great deal of similarity with the T7 tail fiber. Furthermore, it is the C-terminal region of the K1F tail fiber that differentiates and is replaced by an endosialidase domain, allowing for propagation on *E. coli* K1 cells [100]. Moreover, an overview of the K1F genomic highlights is listed below:

- g1.0 - **K1F RNA polymerase** - recognises K1F-specific promoter sequences and conducts the cytoplasmic transcription of genomic phage DNA into RNA fragments capable of translating the ensuing genes into phage proteins. It is also involved in translocating (or "pulling") the rearmost region of the K1F genome into the host cell [103], as well forming a complex with lysozyme to regulate DNA replication, maturation, and packaging [104].
- g2.0 - **Host RNA polymerase inhibitor** - binds to and inhibits the host RNA polymerase, obstructing TXTL of the host DNA and therefore making all endogenous TXTL machinery available for phage synthesis [105]
- g5.0 - **K1F DNA polymerase** - facilitates the concatemeric synthesis of the K1F genome within the host cytoplasm [100]
- g6.7 - **K1F small internal capsid protein** - ejected out of the phage alongside genomic DNA upon host adsorption. Also potentially improves structural stability of the phage and prevents premature genome ejection [97]
- g10a and g10b - **K1F major and minor capsid proteins** - assemble together to form the K1F capsid [94]
- g14, g15 and g16 - **K1F internal core proteins** - act as the internal core in mature K1F virions. Also ejected out upon host recognition and form the ejectosome complex in the host membranes to facilitate DNA translocation [106]
- g17 - **K1F tail fiber protein** - incorporates a K1 capsule-degrading endosialidase as aforementioned [100]

This compact and highly efficient K1F genome enables the phage to be a truly effective agonist to *E. coli* K1. Furthermore, when combined with intricate and well executed engineering, the potential therapeutic and diagnostic applications for phage K1F are vast. The aim of this thesis is to tap into this potential, whilst presenting novel approaches for doing so.

1.6 Introduction to Research Aims

Whilst the research aims for each chapter will be presented in further detail within the Introduction sections for each individual chapter, there is a prevailing theme that runs throughout this thesis which shall be introduced here.

The original idea that instigated the commencement of the research presented in this thesis was to develop a rapid, phage-based diagnostic that could function without the necessity of bacterial processing and only relied on bacterial presence to emit a signal. Due to ongoing K1F/*E. coli* K1 research occurring at the University of Warwick, this phage:host relationship was not only extremely attractive due to each of the organisms' characteristics (e.g. K1F is T7-like and therefore many of its genes and actions are very well understood; K1 is responsible for many human fatalities on an annual basis and therefore a rapid diagnostic for this pathogen would be highly valuable for humanity), but it also made logistical and practical sense, alongside the advantage of being able to lean on existing expertise in the laboratory. Therefore, phage K1F and *E. coli* K1 were selected as model organisms for this diagnostic system.

As aforementioned, prior phage-based therapeutic approaches have validated the sensitivity and specificity of such systems, so the key contribution of this thesis work is to attempt to improve on the SoD. To achieve this, the K1F internal capsid proteins (ICPs) were identified as interesting engineering targets due to the natural phenomenon of the ICPs being ejected, alongside the genomic DNA, out of the phage capsid and into the bacterial host cell upon infection. If phage engineering could be used to fuse a detectable signal to the K1F ICPs, then it is hypothesised here that this signal would only be released (and therefore detectable) upon the presence of the bacterial host (i.e. *E. coli* K1). Furthermore, in the absence of the host, this ICP fusion would be shielded from the external environment (and therefore undetectable) due to its encapsulation - stationed in the structural position of the wild type (WT) K1F ICPs - within the phage capsid. If the protein that is fused to the K1F ICPs is capable of emitting a signal as soon as it is exposed to the external environment, then as long as it is packaged in a sufficient abundance throughout the phage stock, this phage-based diagnostic assay will be capable of detecting *E. coli* K1 in a truly rapid manner (i.e. minutes rather than hours). It may then be possible, in future work, to replicate this mechanism of action to create diagnostic assays for an an array of host pathogens and develop a commercially viable rapid bacterial detection test.

1.7 Outline of Thesis Structure

Following on from this initial Introduction (Chapter 1) and the subsequent chapter which describes the General methods used throughout the thesis work (Chapter 2), the three research chapters (Chapters 3-5) each have their own specific Introduction, Methods and Results and discussion sections. The thesis will conclude with the final chapter, Conclusions and future directions (Chapter 6), followed by the Appendices and Bibliography.

Presented in Chapter 3 are the preliminary results and data collected that were necessary to verify/set up prior to commencing the phage engineering work. This includes:

- Calculations for verifying the rationale of the ICP fusion phage engineering concept
- Construction of plasmids for phage engineering and fusion protein expression
- Establishment and optimisation of an in-house cell-free TXTL system
- Testing the activity of the ICP fusion proteins

Chapter 4 comprises all of the K1F phage engineering experiments and the confirmation assays that indicate whether each approach had successfully generated the desired engineered phage or not. The three engineering approaches deployed were:

- Homologous recombination (genomic engineering)
- CRISPR selection (genomic engineering)
- *in vivo* fusion packaging (non-genomic engineering)

Finally, in Chapter 5, cell-free TXTL is used to optimise the non-genomic engineering approach and the diagnostic capabilities of the engineered phage are tested. Included in this chapter is:

- TXTL synthesis of phage K1F
- TXTL and electron microscopy-mediated visual analysis of K1F synthesis
- *in vitro* TXTL fusion packaging (non-genomic engineering)
- Host-induced signal release and other diagnostic assay investigations

Chapter 2

General methods

2.1 General methods

This chapter presents a set of methods that were used repeatedly over the course of the entire project and are referred to throughout this thesis. More specific methods are referred to within the relevant chapter for the sake of coherence and clarity.

2.1.1 Strains and plasmids

The bacterial strains used during this project are shown below in Table 2.1.

Table 2.1: **Bacterial strains used in this study**

Strain	**Relevant genetic markers or characteristics**	**Comments**	**Source**
E. coli NEB® 5-alpha Competent (DH5α derivative)	fhuA2, Δ(argF-lacZ)U169, phoA, glnV44, Φ80, Δ(lacZ)M15, gyrA96, recA1, relA1, endA1, thi-1, hsdR17	Cloning strain	NEB
E. coli EV36 (K-12 derivative)	argA$^+$, kps$^+$, rha$^+$	K-12/K1 hybrid harboring the chromosomal K1 gene cluster. Used as an *E. coli* O18:K1:H17 analogue	Kind gift of Dr Eric R. Vimr [107]
E. coli Rosetta™ 2 (DE3) (BL21 derivative)	F$^-$, ompT, $hsdS_B(r_B{}^- \ m_B{}^-)$, *gal*, *dcm*, DE3, pRARE2, Cam^R	Carries the DE3 lysogen encoding the T7 RNA polymerase under the control of the *lac*UV5 promoter. Carries pRARE2 to enhance the expression of eukaryotic proteins that contain 7 codons rarely used in *E. coli*. Used for cell-free TXTL crude extract production	Novagen
...	...	...	**Continued on next page**

E. coli R2-pAD-LyseR	F^-, ompT, $hsdS_B(r_B{}^- \ m_B{}^-)$, *gal*, *dcm*, DE3, pRARE2, Cam^R	Rosetta 2 cells carrying the pAD-LyseR plasmid for autolysis-mediated crude cell extract production	This work

Many *E. coli* EV36 strains were also made by transforming the various plasmids listed in Table 2.3 into competent EV36-wt. The full list of these EV36 derivatives is: EV36-pBEST-g6.7::NtNL, EV36-pBEST-g14::NtNL, EV36-pBEST-CtNL, EV36-pUC19-g6.7::NtNL, EV36-pUC19-g14::NtNL, EV36-pCas9-g6.7 and EV36-pCas9-g14.

Listed below in Table 2.2 are the phage that were used during this project.

Table 2.2: **Phage used in this study**

Phage	**Relevant genetic markers or characteristics**	**Comments**	**Source**
K1F	WT	Incorporates an endosialidase enzyme within its tail structure which enables specific infection of *E. coli* strains that produce the K1 polysaccharide capsule	Kind gift of Dr Dean Scholl [100]
K1F-gp6.7::NtNL	non-genomically incorporates the structural fusion protein gp6.7::NtNL	Includes a packaged fusion protein within its capsid structure comprising K1F gp6.7 fused to an N-terminal NanoLuc sub-unit	This work
K1F-gp14::NtNL	non-genomically incorporates the structural fusion protein gp14::NtNL	Includes a packaged fusion protein within its capsid structure comprising K1F gp14 fused to an N-terminal NanoLuc sub-unit	This work

The plasmids that were used during this project are shown below in Table 2.3.

Table 2.3: **Plasmids used in this study**

Plasmid	**Relevant genetic markers or characteristics**	**Comments**	**Source**
pBEST-OR2-OR1-Pr-UTR1-deGFP-T500	ColE1 *ori*, OR2-OR1-Pr, UTR1, T500 Amp^R	Cell-free TXTL expression of GFP and pBEST cloning vector template	Kind gift of Dr Vincent Noireaux [108]
pBEST-OR2-OR1-Pr-UTR1-gp6.7::NtNL-T500	"	*in vivo* and *in vitro* expression of the gp6.7::NtNL fusion	This work
pBEST-OR2-OR1-Pr-UTR1-gp14::NtNL-T500	"	*in vivo* and *in vitro* expression of the gp14::NtNL fusion	This work
pBEST-OR2-OR1-Pr-UTR1-CtNL-T500	"	*in vivo* and *in vitro* expression of the C-terminal NanoLuc sub-unit	This work
pBEST-OR2-OR1-Pr-UTR1-NanoLuc-T500	"	Cell-free TXTL expression of the full NanoLuc enzyme	This work
pUC19	pMB1 *ori*, *lacZ*α, Amp^R	Cloning vector	NEB
pUC19-g6.7::NtNL	"	Incorporates g6.7::NtNL homologous recombination cassette for attempting to generate K1F-g6.7::NtNL via genomic engineering	This work
pUC19-g14::NtNL	"	Incorporates g14::NtNL homologous recombination cassette for attempting to generate K1F-g14::NtNL via genomic engineering	This work
...	...	...	**Continued on next page**

pCas9	p15A *ori*, tracrRNA, crRNA, Cam^R	pCas9 cloning vector	Kind gift of Dr Luciano Marraffini [109]
pCas9-g6.7	"	Expression of Cas9 and targeted double-stranded digestion of K1F-wt (g6.7 protospacer) for CRISPR enrichment of K1F-g6.7::NtNL	This work
pCas9-g14	"	Expression of Cas9 and targeted double-stranded digestion of K1F-wt (g14 protospacer) for CRISPR enrichment of K1F-g14::NtNL	This work
pAD-LyseR	f1 *ori*, pBR322 *ori*, rrnB T1, rrnB T2, Amp^R	Expression of phage lambda endolysin (gene R) for autolysis-mediated crude extract production	Kind gift of Dr Jeff Hasty [73]

2.1.2 Stock solutions

In Table 2.4, all of the media stock solutions used throughout this project are listed.

Table 2.4: **Media stock solutions**

Solution	**Component**	**Amount**	**Comments**
LB medium	Tryptone	10 g	
	Yeast extract	5 g	
	Sodium chloride	10 g	
	Milli-Q water	up to 1 L	Autoclave
LB agar	Tryptone	10 g	
	Yeast extract	5 g	
	Sodium chloride	10 g	
	Agar	15 g	
	Milli-Q water	up to 1 L	Autoclave
...	...	...	**Continued on next page**

LB top agar	Tryptone Yeast extract Sodium chloride Agar Milli-Q water	10 g 5 g 10 g 6 g up to 1 L	 Autoclave and store at 50 °C
2xYTP medium	Tryptone Yeast extract Sodium chloride *1 M potassium phosphate dibasic *1 M potassium phosphate monobasic Milli-Q water	16 g 10 g 5 g 40 mL 22 mL up to 1 L	 *174.18 g potassium phosphate dibasic in 1 L milli-Q water *136.09 g potassium phosphate monobasic in 1 L milli-Q water Autoclave
2xYTP agar	Tryptone Yeast extract Sodium chloride Agar 1 M potassium phosphate dibasic 1 M 1M potassium phosphate monobasic Milli-Q water	16 g 10 g 5 g 15 g 40 mL 22 mL up to 1 L	 Autoclave

All autoclaving was done at 121°C for 20 minutes. Following on from this, media was stored at room temperature. If antibiotics were added to the media it would be used immediately or stored in the fridge until required. Antibiotics were used at the following final concentrations:

- 100 μg/mL ampicillin
- 25 μg/mL chloramphenicol

Table 2.5 comprises the stock solutions used for DNA gel electrophoresis experiments.

Table 2.5: **DNA gel electrophoresis stock solutions**

Solution	Component	Amount	Comments
10x Tris-borate (TBE)	Tris	43.2 g	
	Orthoboric acid	22 g	
	*0.5 M EDTA (pH 8.0)	16 mL	*186.1 g EDTA disodium salt dihydrate in 800 mL milli-Q water. Adjust to pH 8.0 with sodium hydroxide
	Milli-Q water	up to 400 mL	Autoclave
1% Agarose gel	Agarose	1 g	For one large gel or two small gels
	1x TBE	100 mL	Heat in the microwave at full power until the agarose has dissolved
	SYBR™ Safe DNA gel stain	5 μL	

Table 2.6 displays the SM buffer stock solution, which was used to maintain stable phage suspensions throughout the project. Phage were stored at 4 °C.

Table 2.6: **Phage suspension stock solutions**

Solution	Component	Amount	Comments
SM buffer	Sodium chloride	5.84 g	
	Magnesium sulfate heptahydrate	1.97 g	
	*1 M Tris-HCl (pH 7.5)	25 mL	*121.1 g Tris base in 1 L milli-Q water. Adjust to pH 7.5 with hydrochloric acid
	Milli-Q water	up to 400 mL	Autoclave

Finally, Table 2.7 comprises the stock solutions used for preparing the in-house cell-free TXTL system.

Table 2.7: **Cell-free TXTL stock solutions**

Solution	Component	Amount	Comments
S30A buffer	Magnesium glutamate	5.44 g	
	Potassium glutamate	12.2 g	
	*2 M Tris	25 mL	*242.2 g tris base in 1 L milli-Q water
	Milli-Q water	up to 1 L	Adjust to pH 7.7 with acetic acid and autoclave
	DTT	308.5 mg	Add just before use
S30B buffer	Magnesium glutamate	5.44 g	
	Potassium glutamate	30.5 g	
	2 M Tris	2.5 mL	
	Milli-Q water	up to 1 L	Adjust to pH 8.2 with 2 M tris and autoclave
	DTT	154.3 mg	Add just before use
10x Lysozyme Buffer	*0.5 M Tris-HCl (pH 8.0)	40 μL	*60.57 g Tris base in 1 L milli-Q water. Adjust to pH 8.0 with hydrochloric acid
	Sodium chloride	29 mg	
	0.5 M EDTA (pH 8.0)	4 μL	
	Triton X-100	50 μL	
	Milli-Q water	up to 1 mL	Pass through a filtration unit to sterilise
	Lysozyme from chicken egg white	10 mg	Vortex and store at -20°C. Add protease inhibitor cocktail just before use
4x Amino acid mix	Molecular grade water	12 mL	Flash freeze aliquots in liquid N^2 and store at -80 °C
	*Ala, Arg, Asn, Asp, Gln, Glu, Gly, His, Ile, Lys, Met, Phe, Pro, Ser, Thr, Val, Trp, Tyr, Leu and Cys	1.5 mL	*RTS Amino Acid Sampler (Biotechrabbit, Volmerstr, Germany). The final concentration for each amino acid is 6 mM, except for leucine which is 5 mM
...	...	...	**Continued on next page**

14x Energy solution			Flash freeze aliquots in liquid N^2 and store at -80 °C
	*2 M HEPES	3.6 mL	*1.91 g HEPES in 4 mL milli-Q water. Adjust to pH 8.0 with potassium hydroxide
	Molecular grade water	144 μL	
	*Nucleotide mix	1.39 mL	*145 mg ATP, 133 mg GTP, 79.4 mg CTP and 82.6 mg UTP in 1.5 mL milli-Q water
	*50 mg/mL tRNA	576 μL	*30 mg tRNA in 600 μL milli-Q water
	*65 mM Coenzyme A	576 μL	*30 mg CoA in 600 μL milli-Q water
	*175 mM Nicotinamide adenine dinucleotide	276 μL	*34.83 mg NAD in 300 μL milli-Q water. Adjust to pH 7.5 with 2 M tris
	*650 mM Cyclic adenosine monophosphate	170 μL	*42.8 mg cAMP in 200 μL milli-Q water. Adjust to pH 8 with 2 M tris
	*33.9 mM Folinic acid	288 μL	*20 mg solid folinic acid calcium salt in 1.15 mL milli-Q water
	*1 M Spermidine	144 μL	*23.55 μL spermidine in 150 μL milli-Q water
	*1.4 M 3-Phosphoglyceric acid	3.09 mL	*1.03 g 3-PGA in 3.2 mL milli-Q water. Adjust to pH 7.5 with 2 M tris

2.1.3 Protocols

2.1.3.1 Bacterial cultures

Unless otherwise specified, *E. coli* cultures were grown at 37 °C using LB agar for solid cultures and LB media, shaking at 225 rpm, for liquid cultures. Appropriate antibiotics were also added to the cultures at concentrations previously specified.

2.1.3.2 Polymerase chain reaction (PCR)

PCR was used for screening purposes and also for gene of interest amplification whilst adding restriction sites to the 3' and 5' ends if necessary. Primer sequences can be found in Table A.2. The PCR reaction setup used is displayed in Table 2.8 and the cycle protocol is shown in Table 2.9. Unless otherwise specified, the Phusion® High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, USA) was used due to its well established high speed and high performance in PCR.

Table 2.8: **PCR reaction setup**

Component	**Amount**
100 μ M Forward primer	0.5 μ L
100 μ M Reverse primer	0.5 μ L
Template DNA	0.5 μ L
DMSO	1.5 μ L
2x NEB high fidelity Phusion® master mix	25 μ L
Molecular grade water	up to 50 μ L

Table 2.9: **PCR cycle protocol**

Cycle/s	**Denature**	**Anneal**	**Extension**	**Hold**
1	98 °C (1 min)	-	-	-
2-31	98 °C (20s)	62 °C (15s)	72 °C (30s/kb + 30s)	-
32	-	-	72 °C (5 min)	-
33	-	-	-	4 °C (infinite)

If the PCR reaction was unsuccessful, then the protocol was optimised by varying one or more of these parameters: the volume of DMSO, the annealing temperature or the concentration of the template. Unless otherwise specified, this PCR protocol was used throughout the project.

2.1.3.3 DNA gel electrophoresis

The approximate sizes of DNA fragments were checked via gel electrophoresis by comparing them to a standard DNA ladder on a 1% agarose gel. The gels were made using the stock solutions specified in Table 2.5. Electrophoresis experiments were run at 100 V for 30-40 minutes before being observed under UV light.

2.1.3.4 Gel extraction of DNA

Following on from running a DNA gel electrophoresis experiment, if the samples were needed for downstream processing, the QIAquick gel extraction kit and associated protocol were used (Qiagen, Venlo, Netherlands). Changes to this protocol are as follows; 30 μL molecular grade water heated to 55 °C was used instead of the elution buffer in the final step, followed by incubation at room temperature for 10 minutes before the sample was centrifuged for two minutes to collect the purified DNA.

2.1.3.5 Plasmid construction

Plasmids were constructed by cutting the gene insert and backbone vector with the same restriction enzymes, then ligating the two parts via the homology of sticky ends created by the restriction digest. The NEBuffer (New England Biolabs, Ipswich, USA) restriction digest reaction setup is displayed in Table 2.10. The inserts were either a PCR product or a synthetically produced gBlock (IDT, Coralville, USA).

Table 2.10: **Restriction digest reaction setup**

Component	**Amount**
Restriction enzyme/s	1 μL (10 units)
DNA	1 μ g
10x NEBuffer	5 μ L
Molecular grade water	up to 50 μ L

Restriction digest reactions were incubated at 37 °C for 15 minutes before being separated via gel electrophoresis and then continuing with the next stage of the cloning workflow - plasmid ligation. The ligation reactions were carried out using the Anza™ T4 DNA Ligase Master Mix (Thermo Fisher Scientific, Waltham, USA). Table 2.11 displays the plasmid ligation reaction setup.

Table 2.11: **Plasmid ligation reaction setup**

Component	**Amount**
Linearised vector	10-100 ng
DNA insert	3:1 molar excess over vector DNA
Anza™ T4 DNA Ligase master mix	5 μ L
Molecular grade water	up to 20 μ L

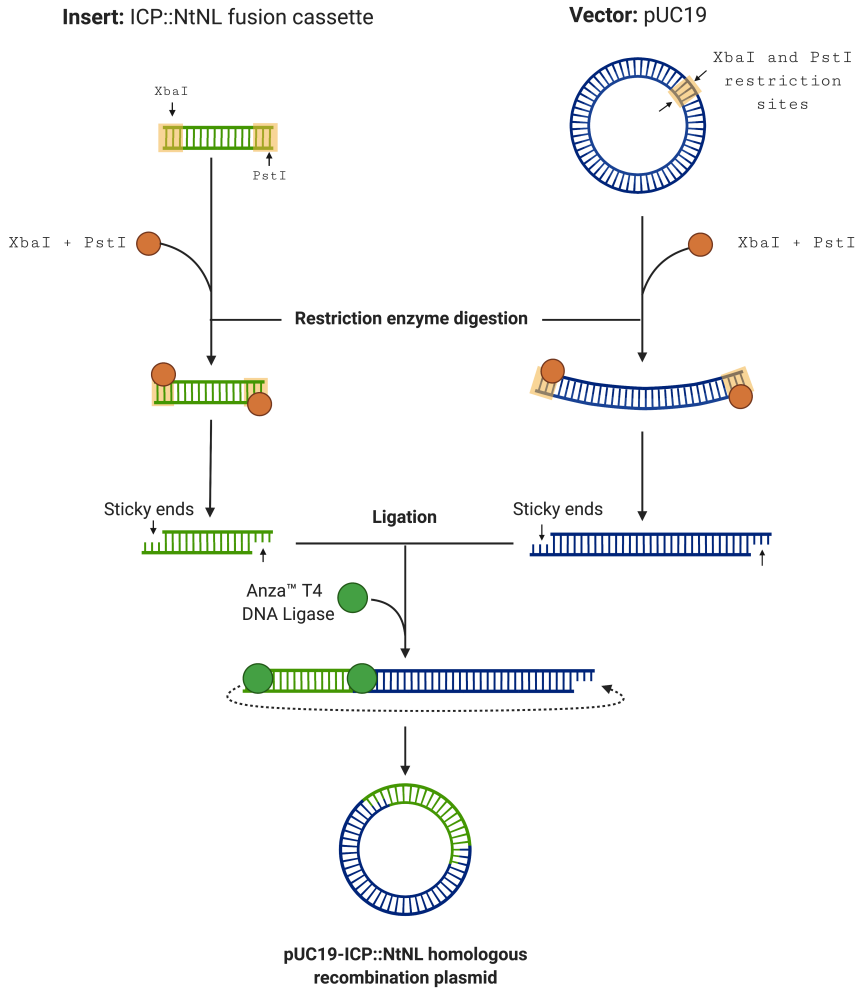


Figure 2.1: **Workflow for plasmid construction via restriction cloning.** The homologous recombination plasmid construction is shown here as an example, with the pUC19 vector and ICP::NtNL fusion cassette insert.

Figure 2.1 visually demonstrates the typical plasmid construction workflow, using the assembly of the homologous recombination pUC19 plasmid as an example.

2.1.3.6 Plasmid purification

Plasmid purification was performed prior to screening to confirm successful plasmid construction, prior to transformation and prior to pBEST plasmids being used in TXTL reactions. The purification was carried out using the QIAprep Miniprep kit and associated protocol (Qiagen, Venlo, Netherlands). Changes to this protocol are as follows; 30 μL molecular grade water heated to 55 °C was used instead of the elution buffer in the final step, followed by incubation at room temperature for 10 minutes before the sample was centrifuged for two minutes to collect the purified DNA.

2.1.3.7 Screening and sequencing

Once a new plasmid had been successfully constructed, it was subjected to a three-stage screening process to verify it had the correct configuration and sequence:

1. **PCR Screening** - primers were designed to bind either side of the site of insertion. This subsequently produced PCR products of different sizes if the gene of interest had been successfully inserted or not. The PCR screening reaction as per the instructions in Tables 2.8 and 2.9.

2. **Restriction digest screening** - a diagnostic restriction digest was also run to verify the product sizes derived from a successfully constructed vector. This was compared with a control vector and the products of these reactions were analysed via gel electrophoresis. The restriction digest screening reactions were carried out by following the protocol described in Table 2.10.

3. **Sequencing** - sequencing was carried out using the GATC LIGHTrun™ service (GATC Biotech, Ebersberg, Germany). The samples were sent with primers to amplify the region of interest. Sequencing results were then analysed using Benchling's Alignment tool to verify correct insertion orientation and/or identify any mutations that may have occurred during the plasmid construction process.

2.1.3.8 Preparation of chemically competent *E. coli*

10 mL of LB medium was inoculated with a single colony of the relevant *E. coli* strain and grown overnight at 37 °C. If necessary, the relevant antibiotic was added for this overnight growth. 1 mL of this culture was used to inoculate 100 ml of LB and grown at 37 °C until

an OD_{600} value of 0.35-0.40 was reached. The cells were then transferred into Falcon tubes and chilled on ice for 20 minutes. Next, the cells were centrifuged at 3220 g for 5 minutes at 4 °C in a Rotanta 46R centrifuge (Hettich, Tuttlingen, Germany). The supernatant was gently removed and the pellet was resuspended in 10 mL ice-cold 0.1 M $CaCl_2$ and incubated on ice for 30 minutes. The cells were centrifuged again as before and then re-suspended in 1 mL 0.1 M $CaCl_2$/15% (v/v) glycerol. 50 μL of cells were aliquoted into microcentrifuge tubes, flash frozen in liquid N_2 and stored at -80 °C until use.

2.1.3.9 Chemical transformation into competent *E. coli*

Competent cells were thawed on ice. Approximately $1-5$ μL of purified plasmid DNA was gently mixed with the cells and incubated on ice for 20 minutes. The mixture was then heat-shocked at 42 °C for 30 seconds followed by incubation on ice for a further 5 minutes. 500 μL of S.O.C. medium (Thermo Fisher Scientific, Waltham, USA) was added to the cells, which were then incubated at 37 °C for 1 hour whilst shaking at 225 rpm. 100 μL of the culture was then spread on an LB agar plate with the relevant antibiotic. Plates were incubated at 37 °C overnight until colonies had formed.

2.1.3.10 Measurement of bioluminescence

Bioluminescence was measured using a FLUOstar Omega microplate reader (BMG Labtech, Aylesbury, England) and the assay was carried out using the Nano-Glo® Luciferase assay system and associated protocol (Promega, Madison, USA). Samples were mixed with an equal volume of the Nano-Glo® assay reagent and 20 μL of this mix was subsequently pipetted into the wells of a Nunc™ 384-well microplate (Thermo Fisher Scientific, Waltham, USA). The relative luminescence was measured within one hour of mixing the assay reagent and samples together. If combining the N- and C- terminal NanoLuc sub-units prior to the luminescence assay, unless otherwise specified the two sub-units were mixed and incubated at room temperature for one hour before adding the assay reagent to allow for spontaneous complementation.

2.1.3.11 Cell-free TXTL reaction setup

Whilst optimal reaction conditions varied depending on which template DNA was being used (this will be elaborated on in subsequent chapters), the TXTL setup does have a consistent core: 10 mg/mL crude extract, 1.5 mM amino acid mix (Table 2.7) and 1x energy solution (Table 2.7) alongside an optimised concentration of additives (PEG 8000, magnesium glutamate, potassium glutamate and dithiothreitol). Each single TXTL reaction was assembled on ice and completed with the addition of molecular grade water up to 20 μL. Following this, the reaction was gently mixed via pipetting and incubated for 16 hours at 29 °C.

2.1.3.12 Bacteriophage propagation

A high K1F phage titer was achieved by propagating the phage over a series of infection cycles in cultures with increasing *E. coli* EV36 concentration - starting with a culture with an OD_{600} value of 0.2, followed by 0.8 and finally 1.2. Immediately after bacterial clearance was observed, each propagation culture was centrifuged at 3220 g for 15 minutes at 4 °C and the supernatent was obtained. After the clearance and centrifugation of the final culture (OD_{600} 1.2), the supernatent was passed through a filtration unit to remove the majority of any bacterial remains still present. At this stage, 1 μg/mL of both DNase I and RNase I (New England Biolabs, Ipswich, USA) was added and incubated with the phage at room temperature for 1 hour to digest any non-phage nucleic acid present in the mix. Next, 0.2 M NaCl was added and incubated on ice for 1 hour to facilitate the release of phage particles from bacterial membranes. Following on from this was a centrifugation at 5000 g for 45 minutes at 4 °C. The supernatant was recovered and incubated with 10% w/v PEG8000 overnight at 4 °C to precipitate phage particles. The following day, the PEG-phage solution was centrifuged for 25,000 g for 1 hour at 4 °C and the pellet was resuspended in SM buffer.

2.1.3.13 Bacteriophage purification

A CsCl density gradient was prepared by mixing varying amounts of CsCl with water to produce three densities of 1.7, 1.5 and 1.3 g/ml. CsCl was then added to the SM buffer-phage solution to achieve a density of 1.3 g/ml. Using a manual pipette, the CsCl solutions were slowly added to a centrifuge tube in equal volumes starting with the heaviest. Once the CsCl-phage tube was prepared, it was placed in a SW28.1 rotor and centrifuged in a Beckman L-90K ultracentrifuge (Beckman Coulter, Pasadena, U.S.A) at 125,000 g for 20 hours at 4 °C. The resulting blue/grey band was extracted and transferred into a Slide-A-Lyzer™ 10k MWCO cassette (Thermo Fisher Scientific, Waltham, USA) which was then dialysed overnight (approximately for 16 hours) in SM buffer. Following on from dialysis, the purified phage were stored at 4 °C.

2.1.3.14 Plaque assay

LB agar plates were overlain with 4 mL of top agar containing 400 μL of *E. coli* EV36 cells (OD_{600} = 0.2) and 100 μL of serially diluted phage. After solidification, plates were incubated overnight at 37 °C. The following day, the plaque numbers were counted and subsequently used to calculate phage titers in PFU/mL. The titer calculation is as follows:

$$\frac{plaques \times\ dilution}{volume} = PFU \tag{2.1}$$

2.1.3.15 Bacteriophage DNA isolation

CsCl purified phage were used for DNA isolation. An equal volume of Tris-saturated phenol (pH 8.0) was added to the phage suspension, vortexed and left at room temperature for 10 minutes. The tube was spun down at 13,000 g for 10 minutes at 4 °C in a 38 Rotanta 46R centrifuge (Hettich, Tuttlingen, Germany). The aqueous phase containing the phage at the top was extracted and added to a new tube. An equal volume of phenol:chloroform (1:1) was added to the extracted phage and vortexed. The tube was then centrifuged again at 13,000 g for 10 minutes at 4 °C. The top layer was extracted again and an equal volume of phenol:chloroform:isoamylalcohol (25:24:1) was added to the tube, vortexed and centrifuged again. 1/10th volume 7.5 M ammonium acetate (pH 8.0) was added together with 2 volumes of ice-cold isopropanol and left overnight at -20 °C for precipitation. The next day, the tubes were centrifuged at 3220 g for 15 minutes at 4 °C (Avanti J-25/JA 25.5 rotor). The supernatant was removed and the pellet was washed with ice-cold ethanol. Finally, the pellet was air-dried and then resuspended in molecular grade water.

2.1.4 Figures, graphs and statistics

All original schematic figures displayed throughout this thesis were made on BioRender. All of the graphs and statistics were completed on GraphPad. The statistics were calculated by carrying out a student's t-test and subsequently generating P values and determining the statistical significance of data generated by paired or unpaired groups. The following symbols were assigned to represent the P values:

- $P < 0.001$ (***)
- $P < 0.01$ (**)
- $P < 0.05$ (*)
- $P > 0.05$ (ns)

Chapter 3

K1F internal capsid protein engineering: experimental design and preliminary investigation

3.1 Abstract

When coupled with a detectable signal, the internal capsid proteins of K1F and other T7-like phage hold high potential for being harnessed as specific and sensitive tools within a rapid diagnostic system. The concealed nature of the phage capsid ensures that any encapsulated signal (i.e. fused to an internal capsid protein) is suppressed. The only naturally-occurring mechanism for releasing and exposing the fused signal is for the internal capsid proteins to be ejected out of the phage - a phenomenon which exclusively occurs when the host is present. Previously, reporter protein sequences have been inserted into non-structural regions of the phage genome, thereby having no effect on the phage phenotype. This approach necessitates the post-infection expression of the protein inside the host - a prolonged process which increases the wait-time on a diagnostic test. In this chapter, it is hypothesised that by fusing a reporter protein sequence with a K1F internal capsid protein sequence (thereby directly modifying the phage phenotype), the process of host-TXTL is side-stepped so that detection can occur within minutes rather than hours. Furthermore, the internal capsid proteins are presented as ideal candidates for use in constructing a rapid detection system that does not rely on host protein expression or the phage propagation process. The work that is presented within this chapter includes a series of novel internal capsid fusion protein designs, the conception of the engineering strategy, the optimisation of an in-house cell-free system and the utilisation of the cell-free system to investigate the signalling capabilities of the fusion proteins.

3.2 Introduction

In this chapter, the proposition of modifying the ICPs of a phage from the *Teseptimavirus* genus is presented for the first time. Furthermore, the various systems and processes needed for such a process are established. In fact, a deep literature search has produced no indication that the modification of ICPs belonging to any type of phage has ever been attempted, confirming the novelty (and perhaps limiting the probability of success) of this concept. In this case, the motivation for engineering the K1F ICPs is to develop a phage-based diagnostic model for rapidly detecting pathogenic bacteria. In this first instance, the pathogen of interest is K1Fs cognate host - *E. coli K1*. The strategy deployed here is potentially beneficial for future diagnostic applications due to the reasons outlined below:

1. The reporter protein fused to the phage ICPs is protected and silenced by the encapsulating proteins that make up the phage capsid, therefore, there is no signal whilst the fusion remains within the capsid

2. The ICPs are ejected out of the phage upon host adsorption, therefore, the fused reporter protein is also propelled out and becomes exposed to the external environment for signal generation (n.b. *only* when the host is present)

3. Previously published phage-based diagnostic models are entirely reliant on phage synthesis and accumulation inside the host and/or host-mediated TXTL of a reporter gene inserted into the phage genome [26, 33–35, 37–40, 43, 44, 110] - a prolonged processes that does not allow for truly rapid detection. Here, the proposed model is based on pre-translated, "ready-to-go" fusion proteins that are packaged inside the phage capsid. This does not rely on intracellular host processing and only requires host presence, thereby enabling a truly rapid diagnostic procedure

3.2.1 The *Caudovirales* order: DNA translocation

One of the most common methods of phage-to-host genome translocation for phage belonging to the *Caudovirales* order (i.e. tailed phage) is via the tail traversing the full outer membrane-periplasm-inner membrane host casing (also known as the host envelope) and injecting the genetic material directly into the cytoplasm [111]. This, along with all other phage DNA translocation methods, is an energy dependent process. To begin with, the energy required for packaging DNA into the phage capsid is derived from the hydrolysis of ATP by the phage packaging motor, terminase (g19 in phage T7 and K1F [100]) - a non-structural phage protein which interacts with the phage portal (g8 in phage T7 and K1F [100]) to package the genomic DNA. Once DNA packaging is complete (i.e. one full genome copy has entered the capsid), terminase

cleaves the concatemeric genome at specific regions, ensuring the correct amount of DNA has entered the capsid [112]. Following on from this, terminase also provides the energy for reverse osmosis to occur - removing water so that the usually hydrated DNA can be further condensed in order to fit inside the capsid. Much of the energy that later propels the genome out of the phage upon infection does in fact come from this hydrodynamic origin that is derived from the terminase-induced osmotic pressure [113].

The three main families that comprise the *Caudovirales* order are: *Siphoviridae*, *Myoviridae* and *Podoviridae*. Phage belonging to the *Siphoviridae* family possess a long, non-contractile tail which is long enough in its natural state to penetrate and traverse the host envelope and deliver the genomic DNA to the host (e.g. phage SPP1, which infects *B. subtilis* [114]). Moreover, *Myoviridae* family phage particles (e.g. phage T4, a coliphage - i.e. a phage that infects *E. coli*) have a more complex tail structure, comprising a contracting sheath which, when contracted, uncovers the internal tube which is then able to penetrate the host envelope and facilitate DNA translocation [115]. However, for phage members of the *Podoviridae* family (including phage K1F and T7), their short, non-contractile tails are not capable of penetrating the host envelope [99], therefore, an alternative mechanism - involving the ICPs [106] - must be used to facilitate the delivery of the genome into the intracellular environment.

3.2.2 The K1F ICPs

In recent years, a great deal of research has contributed towards solving the enigma of *Podoviridae* family phage DNA translocation. Furthermore, for some time it has been postulated that the proteins residing within the phage capsid (i.e. the ICPs) play a role in transporting the genome across the envelope and into the host cytoplasm [116, 117]. Fortunately, within the last decade, significant strides have been made to map out the mechanism of this phenomenon. This *Podoviridae* family DNA translocation research commonly uses phage T7 as a model organism for characterisation [106, 113, 117–121] and as aforementioned, K1F shares a significant proportion of its DNA and compositional structure with T7 (including the ICPs [100]). Therefore, it is proposed here that this body of T7 ICP research can be used to make informed assumptions about the propensity of K1F.

The four primary ICPs residing within the K1F capsid are: gp6.7 (9.3 kDa), gp14 (20.8 kDa), gp15 (84.2 kDa) and gp16 (144 kDa) [122]. All four of these ICPs, along with gp7.3 (10 kDa), are ejected out of the phage and into the host cell upon adsorption. gp6.7 is the smallest of the "ejection proteins" and alongside gp7.3, is known to be the first one to leave the phage, initiating the DNA translocation process [97]. The remaining three ICPs (gp14-16) form the cylindrical "internal core" within a mature K1F particle and until host adsorption occurs they

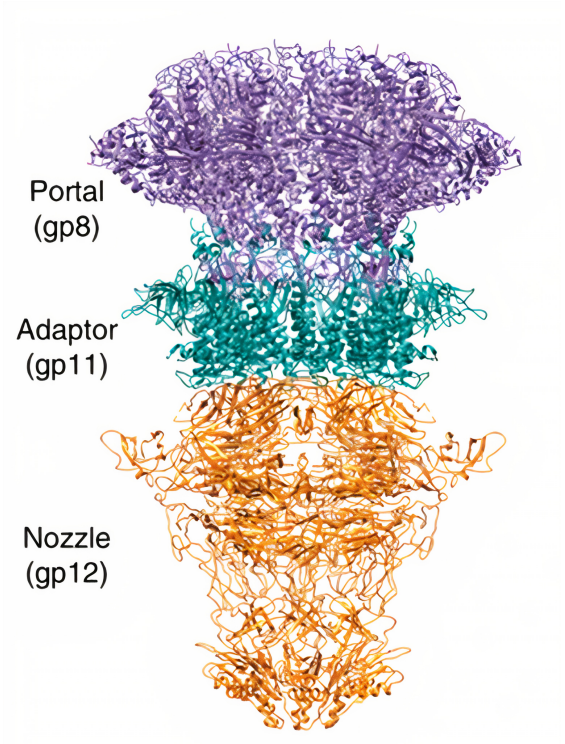


Figure 3.1: **The structure of the T7/K1F tail.** The ICPs and genomic DNA must pass through this compact tail structure prior to being injected into the host. This image is taken directly from Cuervo (2019) [120].

remain solely as an internalised structural element of the phage particle [119]. Following on from the identification of the host via tail fiber interactions, the K1F nozzle (gp12) undergoes a conformational change which results in the opening of the channel required for genomic release [120]. This channel, comprising the portal (gp8), adaptor (gp11) and nozzle (gp12), is visualised in Figure 3.1.

After the genomic translocation process is initiated by the opening of the nozzle and ejection of gp6.7 and gp7.3 into the host cytoplasm (this potentially forms the first pore in the host envelope which the subsequent ejection proteins use as a tunnel to assemble their DNA translocating machinery within), the three ICPs that comprise the K1F internal core (gp14-16) must undergo a profound conformational change (partially unfolding) in order to pass through the tail channel and reconstitute within the host envelope to form what is known as the "ejectosome" [106]. This complex process is presented in Figure 3.2

The first of the internal core proteins to pass through the tail is gp14, which forms a channel across the outer membrane. As this is by far the smallest internal core protein, it is likely that the

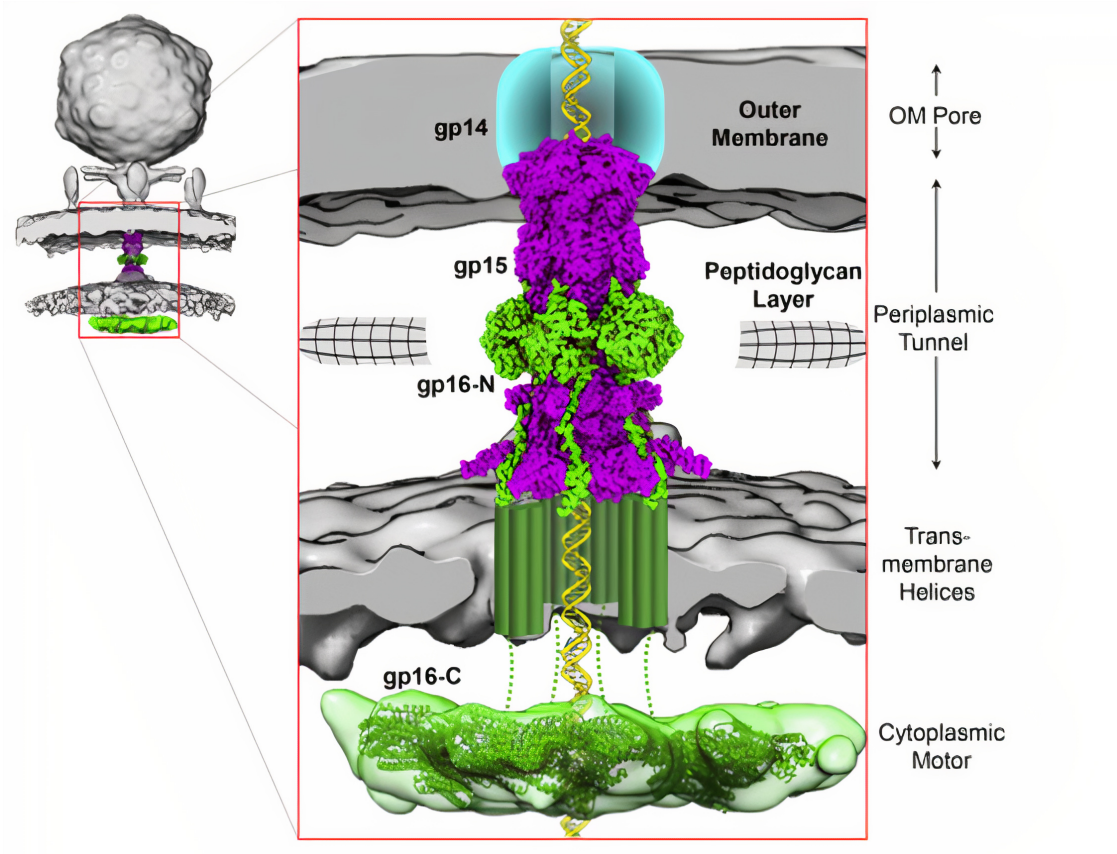


Figure 3.2: **The T7/K1F ejectosome.** Upon host adsorption, the internal core proteins are ejected through the phage tail structure (mediated by partial unfolding) and subsequently refold to form the ejectosome within the host envelope. This image is taken directly from Swanson (2021) [121].

conformational change while passing through the tail is less dramatic for gp14 compared to the other two proteins which are much larger in size. Next, gp15 and gp16 are propelled through the gp14 channel and into the periplasmic space, where the N-terminal gp16 (gp16-N) complexes with gp15 to form the periplasmic tunnel [121]. The degradation of the bacterial peptidoglycan layer is facilitated by gp16 which harbors a lytic transglycosylase (LTase) sub-unit, allowing the gp16-N:gp15 tunnel to digest and penetrate the periplasm [119]. It is also hypothesised that the C-terminal gp16 (gp16-C) forms a pore in the inner membrane to complete the ejectosome channel - spanning the entire envelope. Furthermore, it has been found that gp16-C has DNA-binding activity and it is subsequently hypothesised that it forms a large cytoplasmic hub which binds to the incoming phage DNA and is involved in the motor-like "pulling" of the remaining genome into the cytoplasm [121] - perhaps in collaboration with the endogenous host RNA polymerase (responsible for the transcription of the initial phage genes) and subsequently, once translated, the phage-specific RNA polymerase (K1F RNA polymerase).

In conclusion, whilst there are still aspects of the DNA translocation process and ejectosome mechanism which remain hypothetical, the significant amount of research over the past few years in this area can confidently deduce the following facts regarding phage T7 (and therefore also phage K1F) DNA translocation:

- gp6.7 and gp7.3 are the first ejection proteins to be propelled from the phage particle and subsequently enter the host cell
- gp14, gp15 and gp16 partially unfold, exit the phage tail and refold within the host envelope to form the ejectosome
- gp14 forms the outer channel within the host outer membrane whilst gp15 and gp16 form a complex to extend the ejectosome tunnel throughout the entire the envelope
- the phage genome is delivered through the ejectosome in order to reach the host cytoplasm for phage propagation to commence

Furthermore, whilst members of the *Siphoviridae* and *Myoviridae* phage families use their tail to span the envelope and deliver the genomic DNA, members of the *Podoviridae* family reach the same outcome by utilising their ICPs to form an envelope-spanning tunnel - compensating for their short tails. A remarkable feat of viral evolution which forms the basis of this thesis.

3.2.3 The reporter protein

Following on from the conception of the 'rapid phage ICP-based diagnostic system' idea, the first key component to be considered was the choice of reporter protein to fuse to the K1F ICPs. The ideal candidate must be of appropriate size so that it can physically fit inside the phage capsid without having a significantly detrimental impact on the phage subsistence; it must also be be inducible, suppressible or substrate-dependent so that no detectable signal is generated whilst it resides within the capsid; it must have a certain degree of structural pliability so that it can be fused to a phage ICP without inhibiting its own function; it must produce a specific signal that is not be susceptible to background noise in a bacterial environment; and it must have a significantly strong signal so that even in small quantities it can be reliably detected.

When taking into account all of these considerations, the luciferase enzyme group is presented as an attractive option due to their high specificity and their reliability on a substrate being provided in order to emit light. This bioluminescence-generating group of enzymes can be found across a plethora of life forms including bacteria, fungi, insects and marine organisms [123]. Bioluminescence occurs when a luciferin (i.e. a photon-emitting substrate) is oxidized

by a luciferase enzyme. When considering the K1F ICP fusion, one particular luciferase-based bioluminescence platform, NanoLuc^{TM} - developed a decade ago [124], is particularly attractive and is explored further below.

NanoLuc is a small (19.1 kDa) luciferase which catalyses the conversion of a novel coelenterazine analog (2-furanylmethyl-deoxy-coelenterazine or Furimazine) into Furimamide - a reaction that emits high intensity, glow-type luminescence. In comparison, the two previously most commonly used luciferases, Firefly and Renilla, are 61 kDa and 36 kDa respectively in size and additionally, the NanoLuc system has enhanced thermal and acidic stability, increased brightness and prolonged glow kinetics [124]. The ability for NanoLuc to be detected at very low quantities (as low as 0.01 pM) is also favourable [125] - a comparison of the relative luminescence performance of NanoLuc compared to its two luciferase rivals is displayed in Figure 3.3. Furthermore, previous demonstrations that NanoLuc can be split into two spontaneously complementing sub-units that are not sensitive to unfolding/refolding cycles [126] (referred to here as C-terminal NanoLuc (CtNL) and N-terminal NanoLuc (NtNL)), enhances its position as the ideal candidate for the ICP fusion system. This is due to the fact that it further decreases the size of the component which must be packaged inside the K1F phage capsid and also has to be

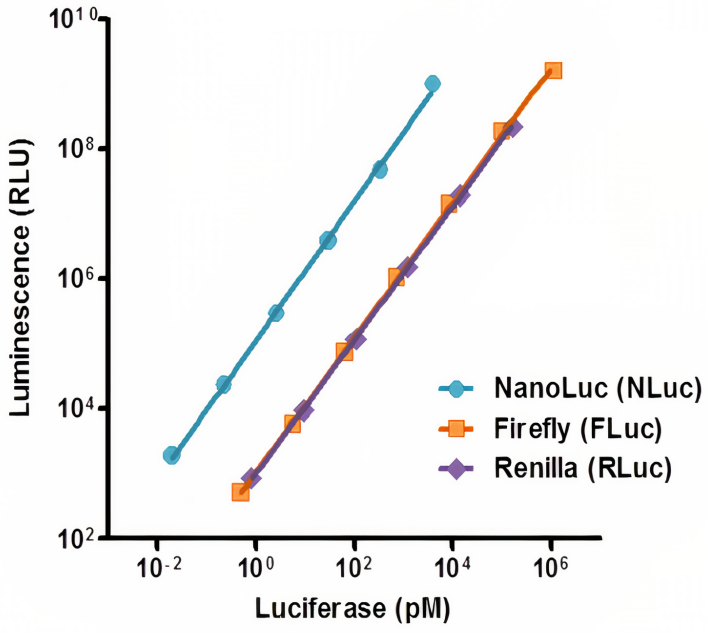


Figure 3.3: **The brightness of NanoLuc at different concentrations compared to Firefly and Renilla luciferases.** This image is taken directly from Riss (2010) [125].

passed through the phage tail upon infection. The smaller and more stable this component is, the higher the probability of success is.

3.2.4 Research aims

The aims of this chapter are as follows:

1. To design, construct and test (if applicable) all DNA constructs and plasmids needed for generating the ICP::NtNL engineered phage K1F
2. To validate the ICP::NtNL engineered phage K1F diagnostic model that has been proposed. In particular, calculate the minimum titer of engineered phage (and subsequently the minimum encapsulated ICP::NtNL fusion protein concentration) needed for a diagnostic assay that is capable of generating detectable luminescence
3. To determine which crude extract lysis method performs the best, explore whether having multiple lysis steps can increase performance yield and optimise an entirely in-house cell-free TXTL system
4. To validate that, by fusing NtNL to the K1F ICPs, the ability of NtNL to compliment with CtNL and subsequently generate luminescence is not inhibited

3.3 Methods

3.3.1 Quadrilysis: *E. coli* crude extract preparation

E. coli crude extract was prepared using Rosetta 2 cells transformed with the pAD-LyseR plasmid, which codes for- and constitutively expresses the bacteriophage lambda endolysin gene R in order to increase cell lysis and subsequently improve the crude extract yield for optimal cell-free TXTL reactions. The protocols described in sections 2.3.6, 2.3.8 and 2.3.9 were followed to create the *E. coli* R2-pAD-LyseR strain.

E. coli R2-pAD-LyseR cells were grown overnight on 2xYTP agar with 100 μg/mL ampicillin and 25 μg/mL chloramphenicol at 37 °C. The following day, 40 mL of 2xYTP media with relevant antibiotics was inoculated with a single colony and cultured overnight at 37 °C, shaking at 300 rpm.

Main growth and washing steps

660 mL of 2xYTP media with 100 μg/mL ampicillin and 25 μg/mL chloramphenicol was inoculated with 6.6 mL of overnight R2-pAD-LyseR culture and grown at 37 °C, shaking at 300

rpm until OD_{600} 1.8-2.0 was reached. The cells were then incubated on ice for 10 minutes and centrifuged at 5000 g for 15 minutes at 4 °C. The pellet was then resuspended in 200 mL of chilled S30A buffer. Once resuspended, the cells were spun again at 5000 g for 15 minutes at 4 °C. The resuspension and centrifugation steps were repeated, then the pellet was resuspended a final time in 40 mL of S30A buffer and spun down at 2000 g for 8 minutes at 4 °C. The pellet was then centrifuged for a further 2 minutes to remove any residual supernatent. Finally, the pellet was weighed and resuspended in 1 x volume S30A buffer, before being flash frozen in liquid N^2 and stored at -80 °C overnight or until the protocol was continued.

Cell lysis

The frozen pellet was thawed in a room temperature water bath and then vortexed vigorously for 3 minutes to begin the cell lysis process. Following this, 1 mg/mL of lysozyme buffer was added to the cells and followed by an incubation on ice for 90 minutes. Cells were then sonicated at 70%, 10 seconds on, 10 seconds off and repeated 12 times. Finally, the cells were vortexed for a further 3 minutes to complete lysis. Following on from the lysis steps, the cells were incubated at 37 °C for 80 minutes at 250 rpm in order to digest any remaining nucleic acids using endogenous exonucleases released during the lysis process.

Final extract preparation

After the exonuclease incubation step, the extract was centrifuged at 30,000 g for 60 minutes at 4 °C. The supernatant was collected and spun again at 21,000 g for 10 minutes 4 °C in the table top microcentrifuge. The extract was then transferred into a new tube and the centrifugation step was repeated using the same settings until the extract was clarified. The clarified extract was inserted into 10k MWCO cassettes (Thermo Fisher Scientific, Waltham, USA) and dialysis was then performed overnight at 4 °C in 900 mL of S30B buffer. Following on from overnight dialysis, the completed extract was aliquoted, flash frozen in liquid N^2 and stored at -80 °C until use.

3.3.2 Measurement of fluorescence

Fluorescence was measured using a FLUOstar Omega microplate reader (BMG Labtech, Aylesbury, England). 20 μL of the samples were pipetted into the wells of a Nunc™ 384-well microplate (Thermo Fisher Scientific, Waltham, USA) and the relative fluorescence was measured.

3.4 Results and discussion

It was decided that two simultaneous phage engineering attempts would be made, focusing on the two most suitable K1F ICP candidates. Moreover, these two engineering targets were identified as gp6.7 and gp14. Firstly, gp6.7 was selected as it is the only encapsulated ejection protein that is propelled into the host cytoplasm upon host adsorption (n.b. gp7.3 is also ejected into the host cytoplasm, however, this protein is a component of the phage tail structure and is therefore continuously available for interactions with the external environment - i.e. it is not internalised and protected within the K1F capsid - so it is not appropriate for this diagnostic system). Subsequently, as opposed to the other membrane-bound ICPs which could become confined within the membrane components and unable to compliment with CtNL, it may be easier to obtain any ejected gp6.7::NtNL fusion protein by simply lysing the host cells and spilling the cytoplasmic fusion into the external, CtNL-rich environment. Secondly, not only because it is the smallest of the ejectosome proteins (and therefore the easiest to work with), gp14 was also selected as a NtNL fusion candidate because it forms the outer domain of the ejectosome within the hosts' outer membrane. Therefore, it is possible that the flexibly fused NtNL protein will be facing outwards, towards the external environment (see Figure 3.2 to help visualise this). As a result of this, upon ejection and settlement within the outer membrane, gp14::NtNL may be instantly available (whilst remaining lodged in the membrane) for complimentation with the CtNL sub-units that reside in the extracellular diagnostic assay solution.

The results that are presented and discussed in this chapter encompass the foundational work and preliminary data that were necessary to establish prior to exploring the efficacy of subsequent phage engineering and *E. coli* detection assays (for which the results are displayed and discussed in the following two chapters). Furthermore, the intention was that following the conclusion of this chapter, the rationale, strategy and preliminary laboratory workflows would be established - providing a strong foundation for facilitating the completion of the remainder of the thesis.

3.4.1 Rationale

Before attempting the novel enterprise of engineering the ICPs of a phage belonging to the *Podoviridae* family, it was important to establish a strong rationale for doing so. Questions such as "What is the minimum quantity of NanoLuc protein required for detectable luminescence emissions?" and "Will there be a sufficient quantity of ICP::NtNL fusion within the engineered phage stock to achieve detectable luminescence emissions?" would need to be satisfactorily answered before proceeding to undergo any wet-lab experiments.

Furthermore, the fundamental aim of establishing a rationale that could justify the further inves-

tigation of utilising the K1F ICPs as NtNL fusion vessels for highly specific and rapid bacterial detection was initiated. The two K1F ICPs that are focused on in this study are gp6.7 and gp14, and the enzyme selected for ICP fusion is NanoLuc - specifically, a 65 residue N-terminus sub-unit of NanoLuc (i.e. NtNL). Assuming the K1F virion structure is similar to T7 (their comparable respective genomes would suggest so), then in each K1F phage particle there are approximately 18 copies of gp6.7 and 10 copies of gp14 [97]. Therefore, in a standard phage stock of 10^{10} plaque forming units (PFU)/mL (10 billion phage particles per mL) it is expected that there would be 180 billion copies of gp6.7 and 100 billion copies of gp14 present within the stock. In an engineered phage stock of the same titer, where either gp6.7 or gp14 is fused to NtNL at a modestly calculated success rate of 10% (1 billion successfully engineered phage particles per mL), it can be estimated that there would be 18 billion copies of gp6.7::NtNL or 10 billion copies of gp14::NtNL present within the 10^{10} PFU/mL engineered phage stock.

Continuing with this rationale, the next calculation to consider is the minimum known number of NanoLuc copies required for detectable light emission. Furthermore, previously published data has shown that as little as 0.01 pM of NanoLuc is sufficient for generating detectable light [125]. The molecular mass of NanoLuc is 19.1 kDa (19100 g/mol), which can be used to calculate its mass in a volume of 1 mL (0.001 L) at a molar concentration of 0.01 pM (10^{-13} mol/L):

$$\text{Mass (g)} = \text{Concentration } (\text{mol/L}) \times \text{ Volume } (\text{L}) \times \text{ Molecular Weight } (\text{g/mol}), \tag{3.1}$$

$$10^{-13} \ (\text{mol/L}) \times \ 10^{-3} \ (\text{L}) \times \ 1.91 \times 10^{4} \ (\text{g/mol}) = 0.00000000000191 \ (\text{g}) \tag{3.2}$$

So, now that the minimum concentration of NanoLuc needed to emit detectable light (1.91×10^{-12} g/mL) is known, Avogadro's constant ($6.02214076 \times 10^{23}$) can be used to calculate the number of NanoLuc copies in this concentration:

$$\frac{((6.02214076 \times \ 10^{23}) \times \ 1.91 \times 10^{-12})}{1.91 \times 10^{4}} = 6.02214076 \times 10^{7}. \tag{3.3}$$

In conclusion, from these calculations it can be deduced that 60,221,407 NanoLuc copies per mL is the minimum known number of copies per mL needed for the emission of detectable light. If the number of gp6.7::NtNL and gp14::NtNL copies per mL that were conservatively calcu-

lated above (18 billion and 10 billion copies respectively) are considered, and a final cautious estimation of 10% for the number of fusion proteins that successfully complement with their C-terminal counterpart (i.e. CtNL) is applied, this results in a final copy number of 1.8 billion for gp6.7::NtNL::CtNL (gp6.7::NanoLuc) or 1 billion for gp14::NtNL:CtNL (gp14::NanoLuc) per mL. Armed with these estimates, it can rationally be concluded that there would be a sufficient concentration of NanoLuc (>60 million copies/mL) in the hypothesised engineered phage-based pathogen detection assay.

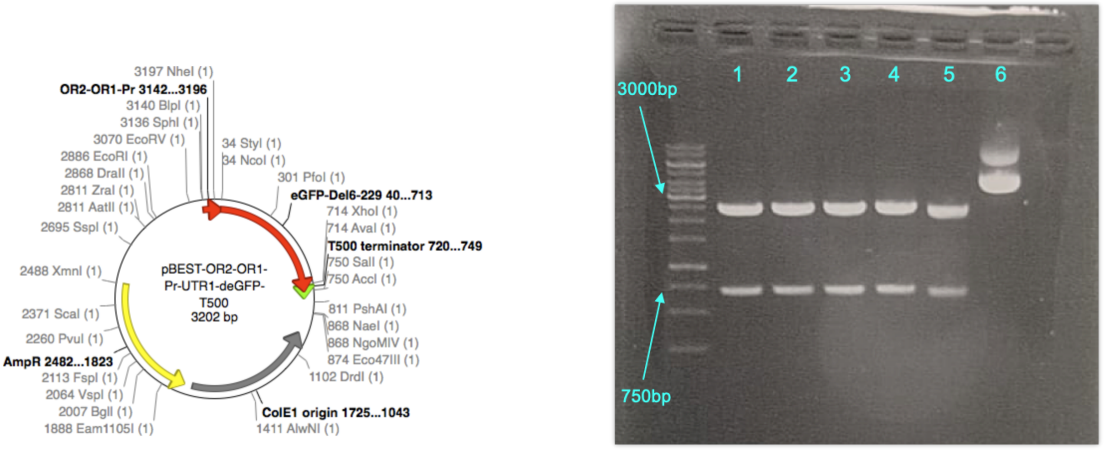
3.4.2 Experimental design

The design and construction work that is implemented on the plasmids presented in this chapter bestow a pivotal contribution in generating the desired engineered phage (K1F*g6.7::NtNL* and K1F*g14::NtNL*). Furthermore, this subsection can be split into three distinct segments: Gene expression, homologous recombination and CRISPR-Cas9. The first, displaying the fusion design and its subsequent use in rapid activity testing and protein spiking experiments, whereas the remaining two segments lay the foundations for the genome engineering experiments. The traditional restriction cloning method was used to assemble all of the plasmids. The detailed protocol for this method can be found in Chapter 2.

Gene expression

As the entirety of the gene expression in this project was to be carried out in a cell-free TXTL system, it was apt that a well-established cell-free vector was chosen as the template to clone the fusion proteins into. Vincent Noireaux's group have published a plethora of cell-free data with pBEST as the gene expression vector of choice [64,66,72,108], so this was the obvious choice for the cell-free work carried out in this thesis. Furthermore, pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 (pBEST-GFP) was a kind gift from Vincent Noireaux (Addgene 40019) and Figure 3.4 shows the plasmid map alongside a gel electrophoresis image displaying the plasmid, cut with NcoI and XhoI (2494bp pBEST backbone + 708bp deGFP gene) and uncut (3202bp full pBEST-GFP).

The insert was designed in an identical way for both g6.7 and g14 (Figure 3.5). Flanked with the NcoI restriction site on the 5' end, the full K1F ICP gene sequence then follows, excluding the stop coding at 3' end of the ICP gene (to allow for one continues fusion protein). A (Gly-Gly-Gly-Gly-Ser)2 flexible linker immediately succeeds the ICP gene, followed by the 65 residue NtNL sequence (ending with a stop codon) flanked by the XhoI restriction site on the 3' end.
Two of the most commonly used linker structures are the GGGGS flexible unit and the EAAAK rigid unit - (GGGGS)n structures comprise a coiled conformation which have been demonstrated



(a) pBEST-GFP map

(b) pBEST-GFP digested with NcoI and XhoI

Figure 3.4: **pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 plasmid map and agarose gel electrophoresis.** The O'GeneRuler 1kb DNA Ladder (Thermo Fisher Scientific, Waltham, USA) was used to estimate band sizes of **1-5:** a pBEST-GFP double digest with restriction enzymes NcoI and XhoI and **6:** pBEST-GFP uncut. Figure 3.4a is taken from Shin (2010) [108].

to be flexible, whilst (EAAAK)n structures form helixes with hydrogen bonds, therefore exhibiting a rigid structure [127]. Furthermore, (GGGGS)2 is a suitable choice of linker as, hypothetically, it should allow for the bifunctionality of the fusion protein - enabling the NtNL region of the fusion protein to have optimal flexibility in order to compliment with CtNL, whilst not impeding rigidly on the ICP function within the phage capsid.

The two NanoLuc sub-units, NtNL (residues 1–65) and CtNL (residues 66–171), have been demonstrated to have negligible light-emitting activity prior to spontaneous reassembly and have also been identified as the brightest pair of NanoLuc sub-units when screened against several other spontaneously reassembling candidates [128]. These characteristics are ideal for the proposed ICP fusion phage-based detection system as minimum background noise will be generated and when the ejected ICP::NtNL fusion is exposed to the externally provided CtNL (which will be a component of the detection assay solution), a bright luminescent signal will be generated. Furthermore, the ICP was fused to the N-terminus of NtNL, rather than the C-terminus of NtNL or the N-/C-terminus of CtNL, in line with previously published data which demonstrates that this is the best performing configuration for a split NanoLuc fusion protein [126].

As previously mentioned, once the inserts were designed and ordered as gBlocks from IDT, restriction cloning was used to cut and ligate them into the plasmid. Figure 3.6 shows the three pBEST plasmids that were constructed - 2974bp pBEST-g6.7::NtNL (well 1), 3337bp pBEST-g14::NtNL (well 3) and 2851bp pBEST-CtNL (well 5). The latter was made so that CtNL could

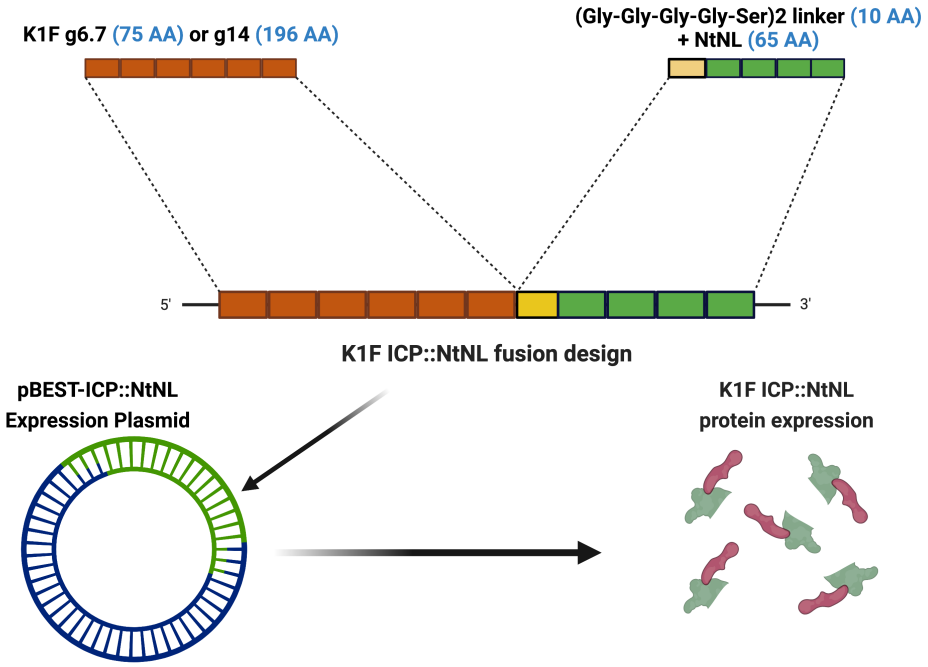


Figure 3.5: **ICP::NtNL fusion protein design schema.** The N-terminus of NanoLuc is fused to the C-terminus of the phage internal capsid protein with a (Gly-Gly-Gly-Gly-Ser)2 linker. The fusion sequences were inserted into the pBEST expression plasmids using restriction cloning, enabling expression of the fusion protein.

be expressed in cell-free prior to ICP::NtNL:CtNL complementation when performing luminescence assays. Wells 2, 4 and 6 show the plasmids cut with the two restriction cloning enzymes, NcoI and XhoI, and therefore also displaying the 2524bp pBEST backbone and their respective inserts - 450bp g6.7::NtNL (well 2), 813bp g14::NtNL (well 4) and 327bp CtNL (well 6).

Using the Benchling 'Analyze As Translation' tool [129], respective sizes of 15.5 kDa and 28.2 kDa were calculated for gp6.7::NtNL and gp14::NtNL. This isn't a huge increase from the WT ICP sizes - 9.3 kDa for gp6.7 and 20.8 kDa for gp14 - so, whilst confirmation was not possible until the phage engineering experiments were complete, there was a degree of optimism that, regarding the process of ICP ejection, little inhibitory activity would be incurred by this size increase. Moreover, the flexible (GGGGS)2 linker allows for the distinct bifunctionality of the two fused proteins, therefore, whilst the ICP may need to partially unfold in order to pass through the phage tail, the NtNL portion of the fusion may not. The two fused components can be viewed as separately functioning entities which do not necessarily need to act in unison. In fact, it has previously been demonstrated that the fusion of two protein domains with different stabilities does not alter the structural changes involved in folding (i.e. one protein domain can unfold

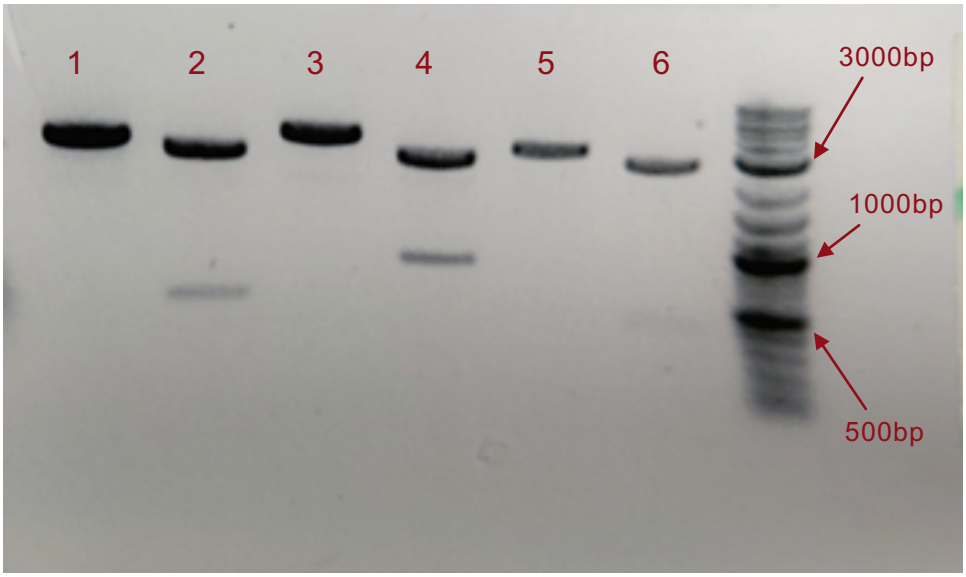


Figure 3.6: **Agarose gel electrophoresis of the pBEST expression plasmid variants.** The 1kb plus DNA Ladder (NEB) was used to estimate band sizes of **1:** pBEST-g6.7::NtNL digested with NcoI, **2:** pBEST-g6.7::NtNL digested with NcoI and XhoI, **3:** pBEST-g14::NtNL digested with NcoI, **4:** pBEST-g14::NtNL digested with NcoI and XhoI, **5:** pBEST-CtNL digested with NcoI and **6:** pBEST-CtNL digested with NcoI and XhoI.

whilst the other remains unchanged) [130]. Furthermore, when using an online protein size calculator [131], the diametric size of NtNL was calculated to be approximately 2.7 nm. With this considered, should the NtNL portion of the fusion protein be small enough to pass through the K1F portal and tail (unlike its larger ICP counterpart), then it is conceivable that it could remain in its unchanged, folded conformation whilst the fused ICP domain makes the necessary conformational changes in preparation for ejection. Moreover, the phage portal and tail are approximately 4 nm wide at their maximum proximity [122], so it is possible that the 2.7 nm NtNL domain of the fusion protein could pass through without necessitating a conformational change. However, even if it was necessary for NtNL to undergo partial unfolding in unison with its fused ICP domain, this could still be a viable outcome as NanoLuc has been shown to have strong stability throughout folding/unfolding cycles [126], so the conformational changes it may make when passing through the K1F tail might not impact its ability to compliment with CtNL and subsequently emit luminescence. In conclusion, regardless of the exact ICP::NtNL ejection mechanism, it is tenable that the act of fusing NtNL to K1F ICPs will not inhibit the ICP ejection or CtNL complimentation processes.

Homologous recombination

The natural process of Homologous Recombination (HR), referring to the exchange of genetic material between two DNA fragments which share similar or identical sequences, has been harnessed as an established technique used in genome engineering research for some time. *E. coli* HR involves many components, including various exonucleases and DNA polymerase, Ligase and Helicase. Once homologous fragments have been identified, the HR process comprises a ssDNA generation step, genetic exchange step and finally ligation and extension steps to finalise the nascent DNA heteroduplexes [132]. Here, the desired use of HR is to insert the ICP::NtNL fusion gene into the relevant regions of the K1F phage genome. A schematic of this approach is displayed in Figure 3.7.

The commonly used and well established *E. coli* cloning vector, pUC19, was the obvious vector-of-choice for the HR experiments due to its success in the same role in a recent publication from the Sagona group [94]. The key features of pUC19 (Figure 3.8a) are its high-copy origin of replication (ori), Ampicillin resistance gene and the N-term β-galactosidase (lacZ) gene. The pUC19 multiple cloning site (MCS) is located within the lacZ gene so that when the genetic material of interest in inserted into the vector, it causes lacZ to become inactive. Post-transformation, the cells that have been transformed with a mixed population of pUC19-empty and pUC19-insert are

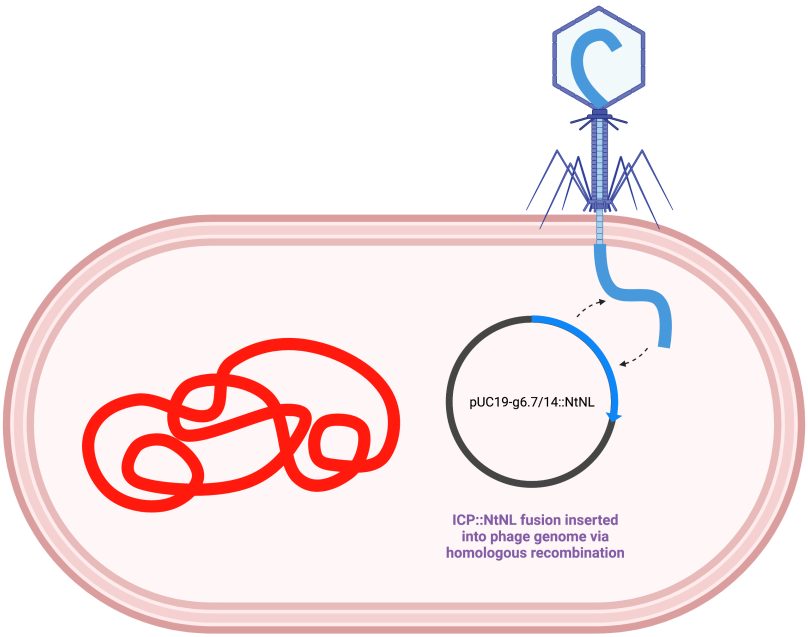
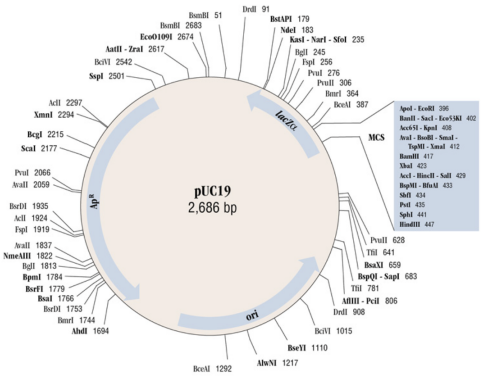


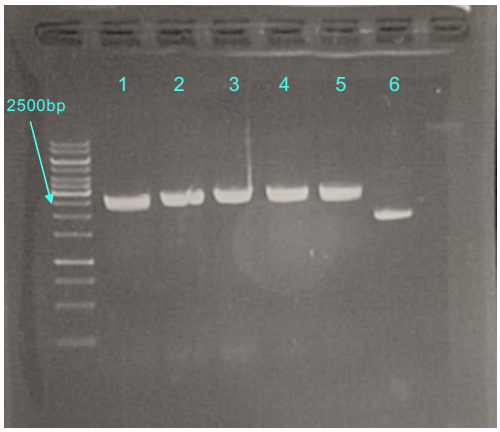
Figure 3.7: **Homologous recombination schema.** DNA fragments are exchanged between interchangeable regions of the pUC19 donor plasmid cassette and the K1F phage genome.

grown on agar plates containing X-Gal and IPTG. Cells with intact lacZ (i.e. the cells that have failed to be transformed with the gene of interest) will undergo intra-allelic complementation with a defective form of β-galactosidase enzyme endogenously encoded by the *E. coli*. The two β-galactosidase fragments are subsequently capable of hydrolysing X-Gal and therefore form blue colonies on the agar plate. However, cells that have been successfully transformed with the pUC19-insert DNA will not express lacZ and therefore will not hydrolyse X-Gal, resulting in the colonies appearing white on the agar plate. This streamlines the restriction cloning process and allows for easy identification of successful transformants. Furthermore, the high-copy trait of pUC19 makes it ideal for HR, as this will result in a higher frequency of pUC19:phage genome interactions, increasing the likelihood of HR occurring.

Figure 3.8 shows the pUC19 plasmid map alongside a gel electrophoresis image displaying the empty vector (2686bp) digested with XbaI and PstI in preparation for restriction cloning (wells 1-5) and uncut in plasmid form (well 6), hence the smaller appearance in size due to supercoiled DNA (which is more compact) running faster on the gel. The restriction enzymes XbaI and PstI were used to cut both the pUC19 vector and the inserts in order to create sticky ends for successful cloning. As in previous K1F phage engineering work [94], the HR fusion cassette inserts for g6.7 and g14 were designed to begin with a 150bp upstream homology arm - up to the last residue prior to the stop codon of the respective gene - followed by the (Gly-Gly-Gly-Gly-Ser)2 linker + NtNL insert, and finally, positioned at the 3' end of the cassette, a 150bp downstream homology arm beginning with the stop codon of the respective gene (Figure 3.9).



(a) pUC19 map



(b) pUC19 digested with XbaI and PstI

Figure 3.8: **pUC19 vector map and the agarose gel electrophoresis.** The O'GeneRuler 1kb DNA Ladder (Thermo Fisher) was used to estimate band sizes of **1-5:** a pUC19 double digest with restriction enzymes XbaI and PstI and **6:** pUC19 uncut. Figure 3.8a is taken from the New England Biolabs website [133].

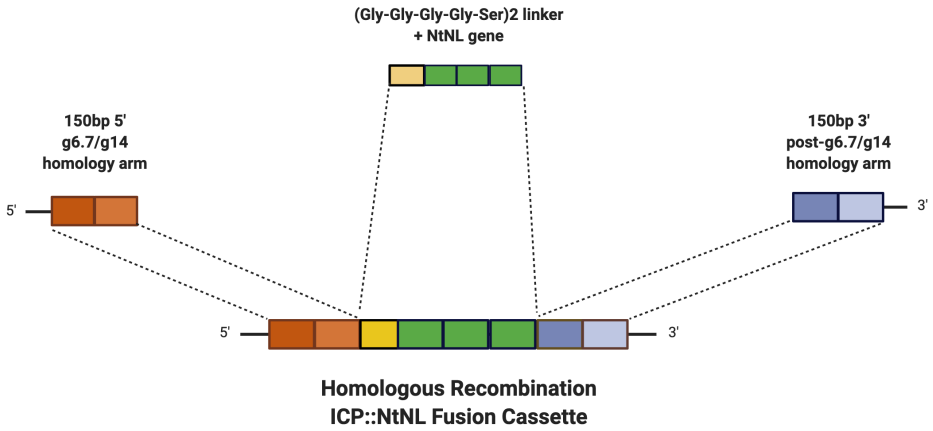


Figure 3.9: **Homologous recombination cassette design for incorporating g6.7 or g14 fused to NtNL into the K1F genome.** The cassettes were inserted into the pUC19 homologous recombination plasmids via restriction cloning.

The 5' homology arm and central component of the HR cassette (i.e. the ICP::NtNL fusion) is entirely based on and in accordance with the design and rationale presented in Figure 3.5. The function of the two 150bp homology arms is to allow for HR to occur between the HR cassette and WT phage genome - with the desired end result being the genome swapping WT g6.7/14 for g6.7/14::NtNL, and the intracellular transcription and translation of this altered genome resulting in the synthesis of an engineered K1F phage population incorporating the internalised ICP::NtNL fusion.

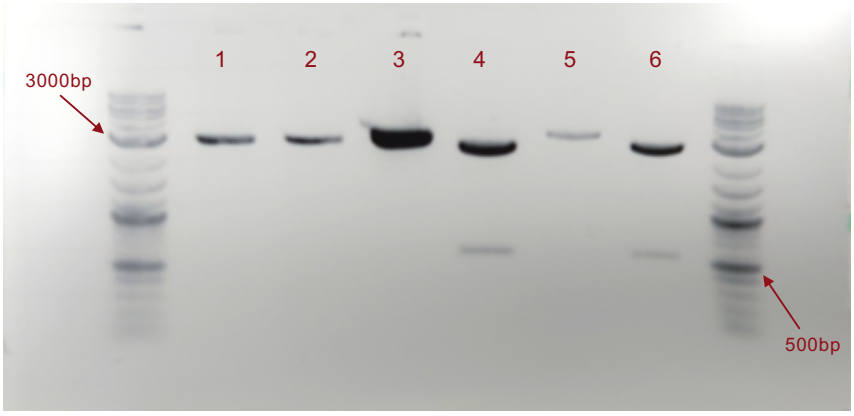
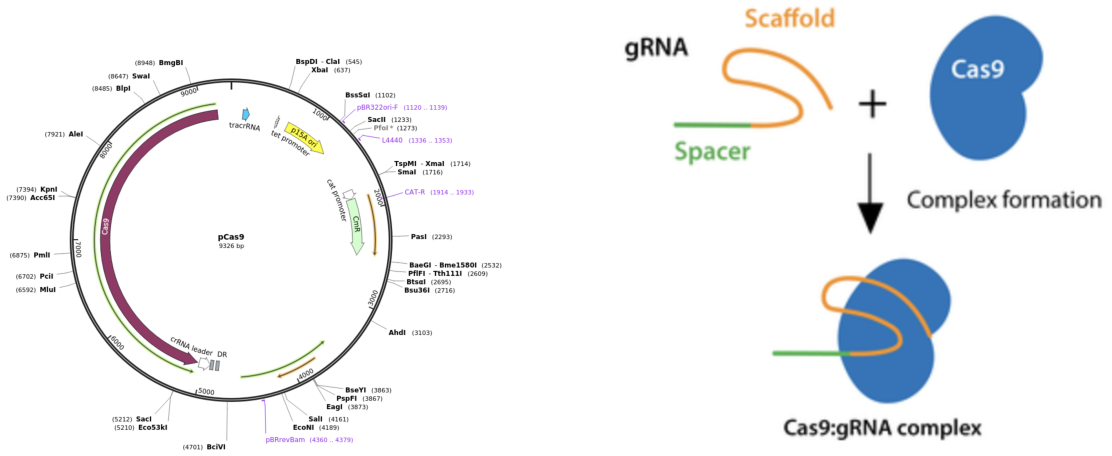


Figure 3.10: **Agarose gel electrophoresis of the pUC19 homologous recombination plasmid variants.** The 1kb plus DNA Ladder (NEB) was used to estimate band sizes of **1:** pUC19 empty vector digested with XbaI, **2:** pUC19 empty vector digested with XbaI and PstI, **3:** pUC19-g6.7::NtNL digested with XbaI, **4:** pUC19-g6.7::NtNL digested with XbaI and PstI, **5:** pUC19-g14::NtNL digested with XbaI and **6:** pUC19-g14::NtNL digested with XbaI and PstI.

The final step of preparing the HR plasmids was to insert the two fusion cassettes into a pUC19 vector via restriction cloning with XbaI and PstI. As shown in wells 1 and 2 in Figure 3.10, there is no observable difference in band size when digesting the empty vector with either one or two of the enzymes - confirming the pUC19 multiple cloning site (MCS) was vacant. Wells 3 and 5 display bands for the linearised pUC19-g6.7::NtNL and pUC19-g14::NtNL plasmids, both 3218bp in size due to having identically sized cassettes. When digested with both enzymes, the HR plasmids produce two fragments, revealing the 522bp ICP::NtNL fusion cassettes (wells 4 and 6).

CRISPR-Cas9 selection

Informed by the aforementioned HR-based phage engineering publications, it was anticipated that multiple rounds of CRISPR selection would be necessary to enrich the engineered phage population following on from HR and prior to attempting to isolate a homogenous engineered phage stock via a plaque assay. Figure 3.11a displays the pCas9 plasmid map, which identifies the key components of the vector, including the parts that were lifted from the *S. pyogenes* genome, thereby enabling their use in *E. coli*. These components include: the Cas9 gene sequence - facilitates expression of the Cas9 endonuclease, the tracrRNA sequence - the fixed portion of gRNA which serves as a binding scaffold for Cas9, and the crRNA leader sequence - pairs with tracrRNA to form functional gRNA and leads the Cas9:gRNA complex to the desired protospacer sequence for which the crRNA is complimentary, a spacer sequence is inserted into the crRNA between BsaI sites using annealed oligonucleotides to allow for specific DNA



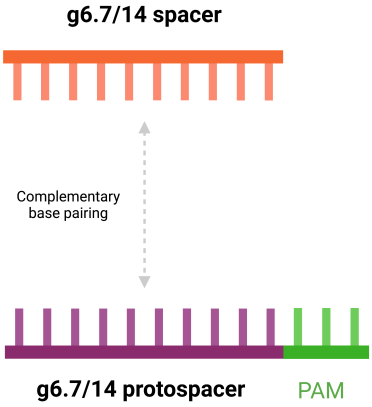
(a) pCas9 map

(b) pCas9 mechanism

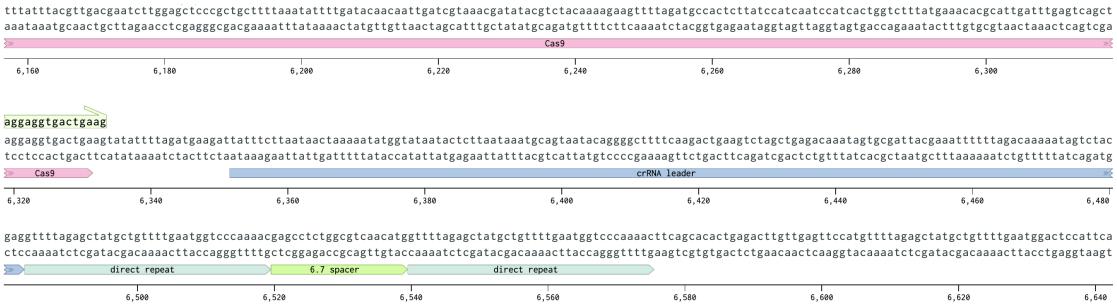
Figure 3.11: **pCas9 plasmid map and mechanism.** The pCas9 plasmid codes for the Cas9 enzyme and its RNA constituents, which subsequently form the Cas9:gRNA complex. Figure 3.11a is taken from Jiang (2013) [109] and Figure 3.11b is taken from the Addgene website [134].

cleavage via the Cas9:gRNA complex. These components are displayed in Figure 3.11b, which shows how they combine to form the CRISPR machinery.

The aim of the CRISPR selection process during phage engineering is to cleave the WT genome whilst sparing the engineered genome, enabling the engineered phage population to become enriched. For this to occur, the protospacer sequence (20bp in length) must be present in the WT genome but not the engineered genome. To facilitate this, the simplest solution is to design the crRNA spacer to complement with a protospacer sequence that spans the intended site of insertion (i.e. the site of ICP::NtNL fusion insertion) so that once HR has occurred, the protospacer sequence is disrupted in the engineered genome and therefore the Cas9:gRNA complex cannot target it.



(a) pCas9 spacer sequence identification



(b) pCas9-g6.7 sequence extract from Benchling

Figure 3.12: **pCas9 spacer insert design and schema.** The spacer sequences must be complimentary to a 20bp stretch in the desired cleavage region of the genome, which in turn must be preceded by a 3bp PAM site. Two complimentary oligos were phosphorylated and annealed to create the spacer sequences, which were then inserted into the pCas9 CRISPR plasmids betwixt two identical direct repeats with a single enzyme cut and ligation.

The insert design for plasmids pCas9-g6.7 and pCas9-g14 required an increased level of planning and attention, mostly due to the necessity of a protospacer adjacent motif (PAM) site 2-6bp downstream of the targeted site of cleavage. During CRISPR selection, once the gRNA has identified its complementary region of DNA (i.e. the protospacer in the WT genome), the Cas9 nuclease recognises the 'NGG' PAM site ('N' can be any nucleotide base) and cuts the DNA 3-4bp upstream of it. If no PAM site is present, then cleavage does not occur. Consequently, the whole CRISPR selection process is reliant on a PAM site already existing within in a small specific stretch of the WT genome upstream or downstream of the site of insertion where a target protospacer sequence can be assigned. If this is not the case, then no cleavage can occur without further modifications to the engineered genome. Figure 3.12a provides an aid for visualising the crRNA spacer - WT protospacer:PAM complex, whilst Figure 3.12b displays a Benchling excerpt of the pCas9-g6.7 sequence, displaying the different CRISPR components.

The CRISPR design for g14 was quite straightforward as, fortunately, there is a PAM site ('AGG') present just 4bp upstream from the 'T' nucleotide of the g14 stop codon (i.e. the site of NtNL insertion). Therefore, the protospacer ('ACCTACAGGTCATAGCTAAG') could comfortably span the site of insertion and the pCas9-g14 crRNA spacer was designed to compliment this. However, the g6.7 CRISPR design process was not so rudimentary as, following on from a thorough analysis of the WT genome, it was concluded that there were no PAM sites in a suitable region neighbouring the site of insertion (the closest PAM is 37bp upstream of the g6.7 stop codon, so the 20bp protospacer would not span the site of insertion). Therefore, even if a crRNA spacer was designed to compliment a protospacer spanning the NtNL site of insertion, the absence of a PAM site would inhibit Cas9 from cleaving the DNA.

Fortunately, a solution could be designed to alleviate this inconvenience. This solution was to use an initially non-viable PAM site (i.e. its corresponding protospacer does not span the site of insertion and therefore would result in cleaved WT *and* engineered genomes) and design the HR fusion cassette (and therefore the engineered genomes) to not only to include the NtNL fusion but also to remove the (previously non-viable) PAM site by changing its codon sequence to a different three base pairs that code for the same amino acid - thereby leaving the peptide translation unchanged. In this case, the PAM site 'GGG' is present immediately prior to the g6.7 stop codon (i.e. the 3' end of the 5' g6.7 upstream homology arm in the g6.7 HR fusion cassette). Without any further modifications, this PAM site would be present in both the WT and engineered genomes, but by changing the 'GGG' sequence to 'GGC' in the 5' HR homology arm, this PAM site is removed from the engineered K1F*g6.7::NtNL* genome, whilst still coding for the same amino acid (glycine). Furthermore, these changes were implemented and the g6.7 HR fusion

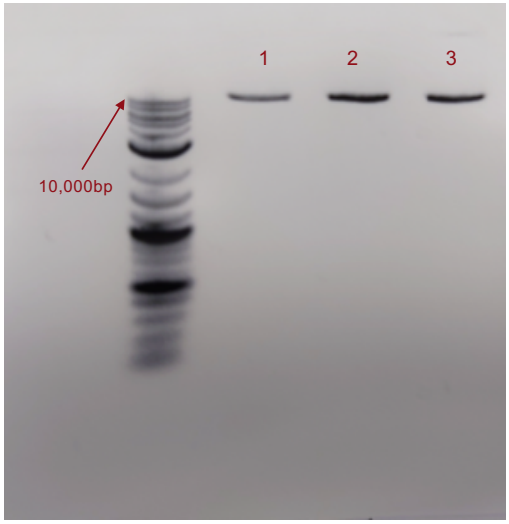


Figure 3.13: **Agarose gel electrophoresis of the pCas9 CRISPR plasmid variants.** The 1kb plus DNA Ladder (NEB) was used to estimate band sizes of **1:** pCas9 digested with XbaI, **2:** pCas9-g6.7 digested with XbaI and **3:** pCas9-g14 digested with XbaI.

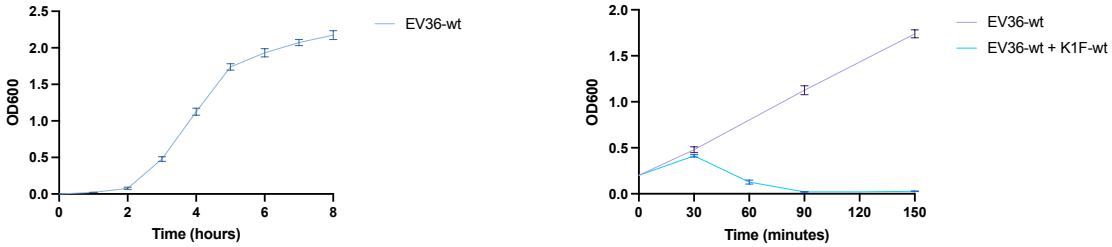
cassette was amended, allowing the 'GGG' PAM site to be used and a crRNA spacer complimentary to a protospacer prior to the g6.7 site of insertion ('GAGCCTCTGGCGTCAACATG') to be designed.

Figure 3.13 displays the empty pCas9 vector (well 1), pCas9-g6.7 (well 2) and pCas9-g14 (well 3) - all linearised with XbaI. The only difference between the three plasmids is the 20bp crRNA spacer sequence - hence the comparable fragment sizes (9316bp).

3.4.3 EV36 growth curve and K1F phage infection kinetics

Once the plasmids had been successfully designed, constructed and transformed into *E. coli* EV36, it was decided to observe what impact, if any, the burden of bearing each plasmid would have on the EV36 growth curve and its fitness as a phage K1F host.

To begin with, the EV36-wt growth curve and its K1F infection kinetics were measured to act as baseline values (Figure 3.14). Over an 8 hour period, a typical *E. coli* growth curve is observed, starting with the lag phase lasting approximately 2 hours, then the exponential phase following on (hours 2 to 5) and then the beginning of the stationary phase being observed towards the end of the curve (Figure 3.14a). The K1F infection dynamics in Figure 3.14b also represent customary findings, with K1F-wt taking approximately 30 minutes to impact the growth of EV36-wt (this is typically observed between 30-60 minutes), followed by the rapid clearance of the culture resulting in almost undetectable OD_{600} readings. Recently published K1F/EV36 data display



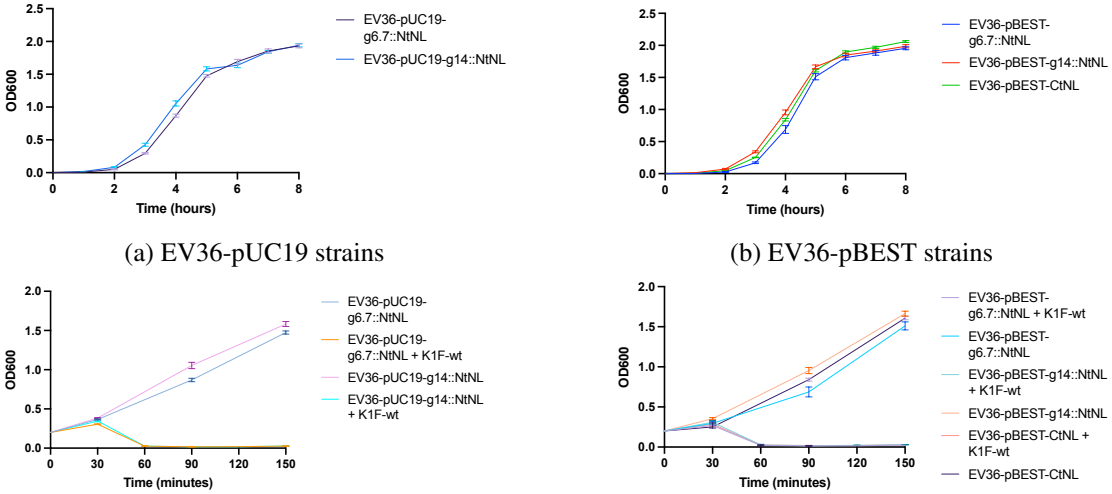
(a) EV36-wt (b) EV36-wt + K1F-wt

Figure 3.14: ***E. coli* EV36-wt growth curve and infection kinetics with K1F-wt.** The optical density at 600 nm was frequently measured to track the cell growth. If phage were added, cultures were infected at a MOI of 0.001 when an OD_{600} value of 0.2 was achieved, (+/- SD, n = 3).

similar bacterial growth rates and phage infection dynamics [135], so it was encouraging that this could be replicated here.

The data for EV36 strains bearing the various pUC19 and pBEST plasmids are displayed below in Figure 3.15. Evidently, the plasmids and the respective gene variations within the plasmid groups have little effect on the EV36 growth curve or the ability of K1F-wt to infect the cells and propagate.

The pCas9 bearing EV36 strains were the only ones that exhibited any notable degree of distinc-



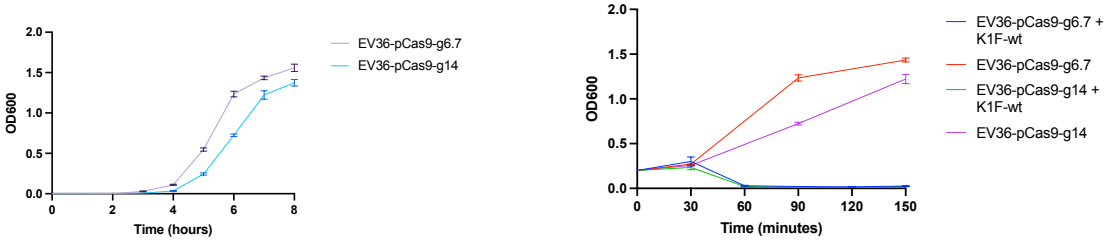
(a) EV36-pUC19 strains (b) EV36-pBEST strains

(c) EV36-pUC19 strains + K1F-wt (d) EV36-pBEST strains + K1F-wt

Figure 3.15: ***E. coli* EV36-pUC19-g6.7::NtNL, EV36-pUC19-g14::NtNL, EV36-pBEST-g6.7::NtNL and EV36-pBEST-g14::NtNL growth curves and infection kinetics with K1F-wt.** The optical density at 600 nm was frequently measured to track the cell growth. If phage were added, cultures were infected at a MOI of 0.001 when an OD_{600} value of 0.2 was achieved, (+/- SD, n = 3).

tion in growth curve kinetics, both compared to EV36-wt and internally between g6.7 and g14. As shown in Figure 3.16a, four hours pass before the cells enter the lag phase (compared to 2 hours for the other EV36 strains) and by the end of the 8-hour growth curve the cells are only just exiting the exponential phase - which begins to peak at a significantly lower OD_{600} value compared to the other strains (OD_{600} 1.5 rather than 2.0). It is also observed that the performance of EV36-pCas9-g14 is considerably inferior compared to EV36-pCas9-g6.7. One clear explanation for the much decelerated growth displayed by the EV36-pCas9 strains is the sheer size of the pCas9 protein that is being constitutively expressed by these cells. It is, in fact, a well established fact that when *E. coli* cells are forced to express an "unneeded" protein, the fitness of the cells and the speed of their growth is negatively impacted - this is known as "protein cost" [136]. Furthermore, the extent of the protein cost is dependent on the size and function of the protein. With this considered, it is not surprising that EV36 cells constitutively expressing pCas9 (160 kDa) grow at a much slower rate when compared to EV36 cells bearing either of the pUC19 plasmids (no protein cost) or pBEST plasmids (15.5 kDa and 28.2 kDa for gp6.7::NtNL and gp14::NtNL respectively).

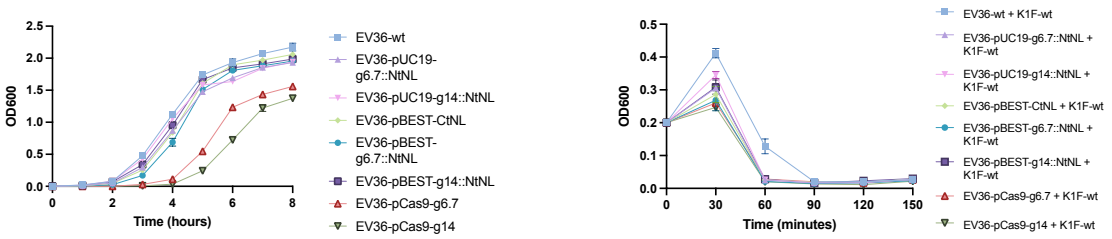
Quite surprisingly, K1F infection kinetics are not impacted by the presence of pCas9 with spacer sequences targeting K1F WT g6.7 or g14 (Figure 3.16b). This is contrary to the inference that the two pCas9 strains (EV36-pCas9-g6.7 and EV36-pCas9-g14) should target and cut their encoded spacer sequences [62] (i.e. the WT K1F genomes), which consequently would decelerate intracellular phage synthesis and therefore hinder the K1F infection rate. This observation initially caused concern, suggesting that the fitness of the pCas9 plasmids was relatively poor. However, upon conferring with the researchers who recently used the same strategy (HR + CRISPR) to engineer phage K1F*g10b::gfp* [94], it was disclosed that they also observed identical findings (i.e. their pCas9 EV36 strains did not inhibit K1F infection kinetics). Moreover, even though



(a) EV36-pCas9 strains

(b) EV36-pCas9 strains + K1F-wt

Figure 3.16: ***E. coli* EV36-pCas9-g6.7 and EV36-pCas9-g14 growth curves and infection kinetics with K1F-wt.** The optical density at 600 nm was frequently measured to track the cell growth. If phage were added, cultures were infected at a MOI of 0.001 when an OD_{600} value of 0.2 was achieved, (+/- SD, n = 3).



(a) All EV36 strains

(b) All EV36 strains + K1F-wt

Figure 3.17: **Growth curve and phage infection kinetics comparison of the *E. coli* EV36 strains.** The optical density at 600 nm was frequently measured to track the cell growth. If phage were added, cultures were infected at a MOI of 0.001 when an OD_{600} value of 0.2 was achieved, (+/- SD, n = 3).

the pCas9 fitness also appeared low during their work, it still served its function later on and successfully selected for the engineered phage. Therefore, it was decided to proceed without dwelling too much on this matter.

Finally, to provide a clear visual comparison, all data were compiled into a single graph for both the 8-hour EV36-only growth curve (Figure 3.17a) and K1F-wt phage infection growth curve (Figure 3.17b). To summarise, it can be concluded that the pUC19 and pBEST plasmids have very little effect on EV36 growth (due to the lack of or very little protein cost), whereas the pCas9 plasmids do significantly decelerate and limit EV36 growth (due to the substantial plasmid size - almost 10,000bp - and constitutive expression of the similarly substantial Cas9 protein). As expected, the unburdened EV36-wt is the fastest growing strain and it reaches the highest OD_{600} value after 8 hours of incubation. This can be attributed to the absolute lack of unneeded protein expression or plasmid replication within this WT strain and therefore optimal growth conditions are presented. It can also be observed that K1F-wt takes a longer time to clear the EV36-wt culture compared to the EV36 variants. This could be attributed to the faster growth of EV36-wt compared to the variants and therefore the phage are required to undergo a prolonged propagation before culture clearance can be achieved. In other words, it takes a longer time period for the phage to clear a more rapidly growing culture that has reached an OD_{600} value of 0.4, compared to cultures that have only reached an OD_{600} value of 0.3 (as seen in Figure 3.17b). All of the EV36 variants are cleared by K1F-wt comparably to each other.

3.4.4 Establishment of an optimised crude extract preparation method for cell-free TXTL

At the start of this project, it was envisioned that by utilising the advancements and tools from the reinvigorated field of cell-free synthetic biology, many processes of interest could be streamlined, optimised and subject to innovation. Rapid testing of protein-expressing constructs

and manipulation of phage within a cell-free system were two initial ideas that would be expanded upon over the course of completing this thesis. Furthermore, as it would be necessary to experiment with a substantial number of TXTL reactions during this project, it was decided to build an in-house cell-free TXTL system from scratch rather than extensively use one of the commercially available TXTL master-mix products from a vendor such as Arbor Biosciences (Arbor Biosciences, Michigan, USA). In doing so, a considerable amount of financial resources would be saved. For comparison, the "myTXTL" kit from Arbor Biosciences costs approximately £8 per reaction, whereas it has been estimated that an in-house system costs approximately 10-20p per reaction [72] - exhibiting an almost 100% cost reduction and therefore a worthwhile commitment.

As the core provider of the transcriptional and translational machinery, the crude cell extract can be considered as one of the more important aspects of a cell-free system. Therefore, the process of developing a crude extract was made over many iterations until optimal results were achieved. Crude extract preparation experimentation was carried out using different lysis methods informed by previously published cell-free content – bead beating, sonication, autolysis, lysozyme-assisted cell lysis and were all attempted and compared alongside a lysis amalgamation method which utilised the best performing lysis methods in unison. In each case, the crude extract was combined with the energy solution, amino acid mix and additives, then assessed for performance by measuring its GFP-expressing activity in relative fluorescence units (RFU).

The main reason for experimenting with an array of lysis methods was due to the widespread outlook from within the cell-free research community that consistent cell lysis is very hard to standardise from one laboratory to another and the subsequent TXTL expression yields are highly variable, even when following an identical protocol [137]. Whilst there have been several attempts over the past decade to provide a standardised crude extract preparation protocol [72, 73, 137–140], the underlying message appears to be that each lab should optimise their own crude extract preparation method based on the blueprint protocols available in the literature. Interestingly, whilst the crude extract method frequently changes, the vast majority of research groups use the same energy solution and amino acid mix protocols that were provided in Noireaux's 2013 publication [72]. This all but confirms the robustness of these methods and therefore they were utilised, unmodified, alongside the crude extract for the in-house cell-free system presented in this thesis.

The performance of all the in-house crude extract attempts were directly compared with myTXTL - the current industry standard solution which can typically yield 20 μg of protein per reaction [141]. The results are summarised in Figure 3.18. It can be observed that the lysis amal-

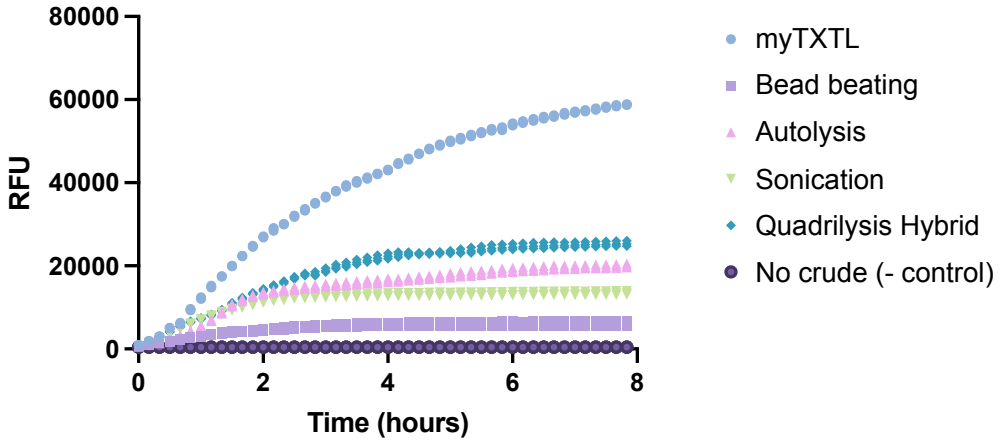


Figure 3.18: **Crude extract method comparison.** The RFU output over 8 hours was compared between TXTL systems comprising the in-house 4 crude extracts made with different methods, the myTXTL commercial kit and a negative control (no crude extract). Identical TXTL reaction conditions were used for each crude variant. Each reaction was supplemented with 18 nM pBEST-GFP plasmid, (+/- SD, n = 3).

gamation hybrid method (named "Quadrilysis") gave approximately 20% more yield compared to autolysis, which in turn gave nearly 40% more yield compared to sonication, which in turn gave nearly 100% more yield than bead beating. Finally, as Figure 3.18 shows, the commercial myTXTL system is capable of yielding approximately three times more GFP than what is obtained using the best in-house attempt (i.e. the Quadrilysis hybrid method). This is a testament to the power of myTXTL, nevertheless, it was decided that the optimisation iterations had been saturated and the in-house crude extract method had been optimised to a satisfactory level to proceed with the continuation of the project.

Informed by the optimisations above, the in-house system was constructed as per the optimised Quadrilysis crude extract preparation protocol described in this chapter's methods section and the unmodified energy solution and amino acid mix protocols described in chapter 2. The Quadrilysis hybrid method takes inspiration from multiple sources and was developed over the course of many iterations. This optimised crude extract preparation protocol follows the general principles described throughout the various pioneering cell-free renaissance papers published in the 2000s [142–144] which lead to Noireaux's well-renowned 2013 protocol [72]. It also utilises Jeff Hasty's pAD-LyseR plasmid (Addgene 99244), which enables *E. coli* autolysis via the constitutive expression of phage lambda endolysin (gene R) plus a freeze-thaw cycle [73] and the addition of lysozyme and sonication steps - resulting in a new amalgamated Quadrilysis protocol (Figure 3.19) - the term Quadrilysis referring to the four amalgamated lysis methods (freeze-thaw, autolysis, lysozyme and sonication). By extensively processing the *E. coli* cells through

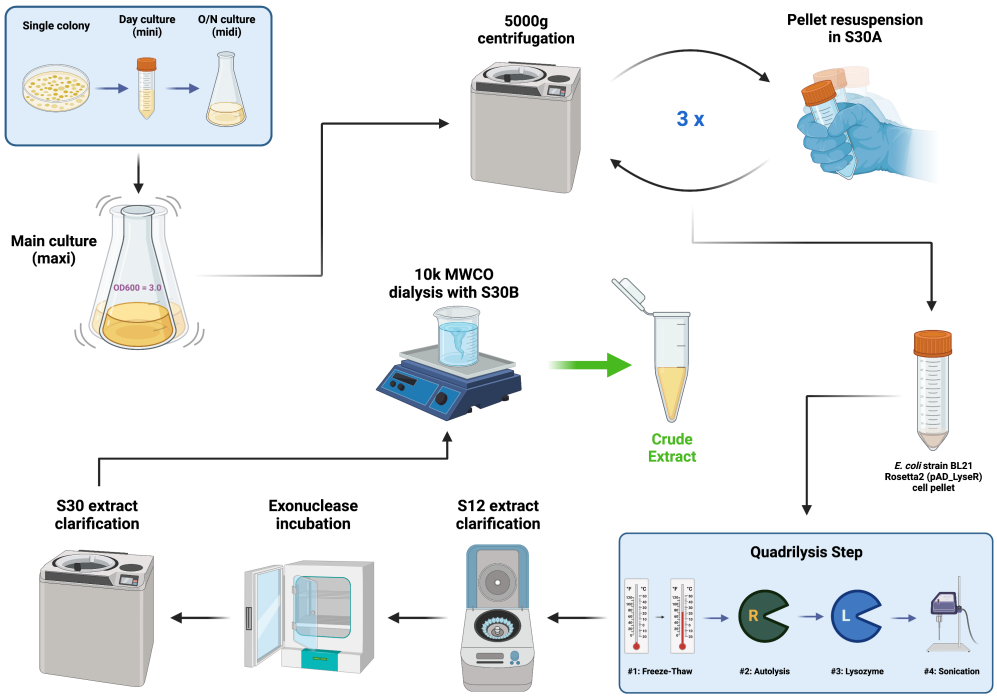


Figure 3.19: **Quadrilysis: an optimised *E. coli* crude extract preparation method**

multiple lysis steps, a thorough cell lysis is enabled and subsequently, optimal downstream results can be achieved.

The optimised Quadrilysis method, visualised above, begins with a three-fold cell growth step (single colony plate growth, then small day culture, then larger overnight culture) which prepares the cells for a rapid main culture growth that is able to reach an OD_{600} value of 3.0 in under 4 hours. Is important to culture the cells up to a density of approximately OD_{600} 3.0 as this represents the exponential phase of growth where the intracellular ribosome concentration is at its the highest and the translational machinery is most active [140]. 2xYTP media is used for all growth steps as it has previously been suggested that this growth medium optimises transcription using endogenous sigma factors [145], which will inevitably improve TXTL protein expression yields. Following on from centrifugation and resuspension in S30A buffer ("S30" is referring to the soluble extract obtained after a 30,000 g centrifuge spin [144]) is the commencement of the cell quadrilysis. As aforementioned, the freeze-thaw and autolysis steps are derived from the Hasty's 2017 publication [73]. Moreover, the constitutively expressed phage lambda endolysin gene produces low levels of intracellular endolysin throughout the main culture of the *E. coli* cells. Importantly, acting on its own, this particular endolysin has a negligible affect on *E. coli* cells until the inner membrane has been compromised and it has been shown to have no

significant impact on cell growth rates [146]. Cellular autolysis is initiated when the freeze-thaw cycle disrupts the inner membrane, thereby allowing the endogenously expressed endolysin to begin full cell lysis [147]. By combining this freeze-thaw-endolysin mediated cellular autolysis with an additional two lysis steps (lysozyme and sonication), optimal TXTL yield was achieved (Figure 3.18). These final two lysis steps are both well known *E. coli* lysis methods and have previously been demonstrated as efficient crude extract facilitators [140, 148, 149], which is why they were experimented with here, both on their own and in combination with each other in Quadrilysis. The fact that the combination of these lysis methods resulted in a better TXTL yield suggests that they elicited a more thorough lysis process and therefore provided a more optimal amount of TXTL machinery within the crude extract.

The remainder of this optimised crude extract preparation method is heavily based on Noireaux's 2013 paper [72], with multiple extract clarifications steps alongside an exonuclease incubation (where endogenous exonucleases digest any remaining nucleic acids within the extract) being carried out before a final dialysis cleansing with S30B buffer finalises the protocol.

3.4.5 Optimisation of the in-house cell-free TXTL System

Once the crude extract, energy solution and amino acid mix had been prepared, the final step was to optimise the additive concentrations for optimal TXTL yields. The additives, namely Magnesium glutamate (Mg-glu), Potassium glutamate (K-glu), Dithiothreitol (DTT) and Polyethylene glycol 8000 (PEG 8k), are used to increase the efficacy of the cell-free reaction [72]. Here, the efficacy is simply equated with the production of the desired output, and not with minimising any waste products. In this case, the desired output is measured in terms of the production of GFP, measured in RLU. So, experimentation was carried out with varying concentrations of the various additives to discover which combination gives the highest production of the desired output. The results for varying concentrations of the four different additives are summarised in Figures 3.20-3.23. Based on the TXTL optimisation results previously reported [72, 73, 148], the base level concentrations for these additives were set as follows: 4 mM for Mg-glu, 60 mM for K-glu, 1.5 mM for DTT and 2.5% for PEG 8k.

The additive concentrations were varied with respect to these baseline levels as follows. Firstly, a set of experiments was carried out wherein the concentration of only Mg-glu was changed while the other additives were kept at their baseline values. Using this experimental data, the optimum value of Magnesium glutamate was determined. In the next set of experiments, the Mg-glu concentration was fixed at this optimum value and the concentration of only K-glu was varied with the concentrations of the other additives at their baseline levels. This procedure was carried until the list of additives at disposal were exhausted. The order of such optimisation (i.e.

Mg-glu first, then K-glu etc.) is as per the logic proposed in [72], wherein it is stated that "our experiments show that the crude cell extract is most sensitive to Mg-glutamate levels, followed by K-glutamate levels" and "We have found that added DTT does not significantly affect end-expression levels". Therefore, it was decided to optimise the most essential additive first (i.e. Mg-glu), and the least essential last (i.e. DTT). Due to the fact that PEG 8k was not optimised in [72], the optimal values of the other three additives were found first, then the various PEG 8k concentrations were tested as a final step.

As Figure 3.20 shows, increasing concentration of the additive Mg-glu from 0 to 7 mM monotonically increased the output production. But beyond the critical value of 5 mM, any further addition of Mg-glu slightly reduced the output production. Hence, 5 mM can be taken to be the optimum concentration. Since magnesium ions are well known mediators of enzymatic reactions [150] and their allosteric function has been demonstrated to play a key role in cell-free TXTL enzymatic reactions [151], it is not surprising that the addition of Mg-glu has such a significant impact on TXTL fluorescence yield.

A similar trend is observed in Figure 3.21 with K-glu. Here, the optimum concentration was found to be 80 mM. In fact, the omission of K-glu (i.e. 0 mM) resulted in the near-inhibition of the TXTL reaction, clearly signifying its key role in TXTL function and suggesting it may be the most important additive, above Mg-glu. Again, the importance of K-glu was to be expected due to the significant role of potassium ions in enzymatic reactions and specifically in cell-free protein expression [152].

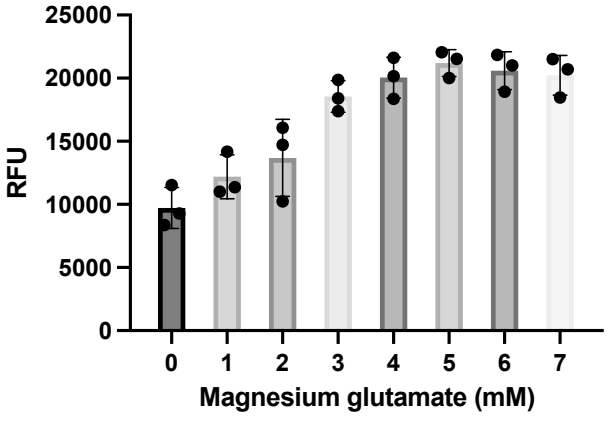


Figure 3.20: **Cell-free TXTL magnesium-glutamate optimisation.** The GFP output (measured in RFU) of in-house cell-free TXTL reactions supplemented with different Mg-glu concentrations was compared, (+/- SD, n = 3).

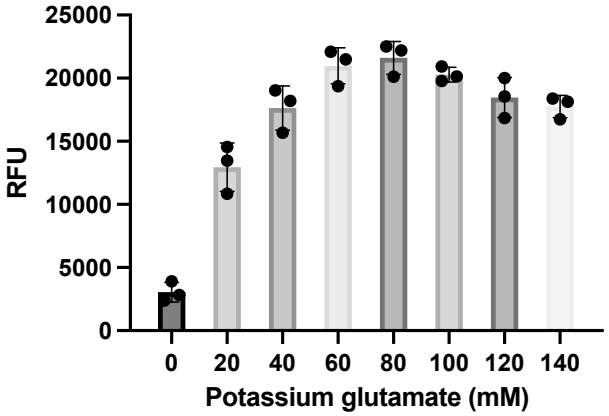


Figure 3.21: **Cell-free TXTL potassium-glutamate optimisation.** The GFP output (measured in RFU) of in-house cell-free TXTL reactions supplemented with different K-glu concentrations was compared, (+/- SD, n = 3).

As shown in Figure 3.22, the influence of supplying DTT as an additive is negligible. These results support conclusions that have previously been drawn [72, 148], where it was noted that the addition of DTT has an insignificant effect on expression. The reason for adding DTT in the first instance is because it is well characterised "enzyme stabilising" agent [153] which therefore should help maintain and optimise the enzymatic ecosystem of a cell-free system. However, due to the consistent findings that it does not have a significant impact on TXTL protein expression, it can confidently be regarded as an unnecessary additive. Nevertheless, as there was a slight observable RFU output increase (albeit insignificant) at a DTT concentration of 2.5 mM, this value was chosen for the optimised additive setup, as no harm would be done by including it.

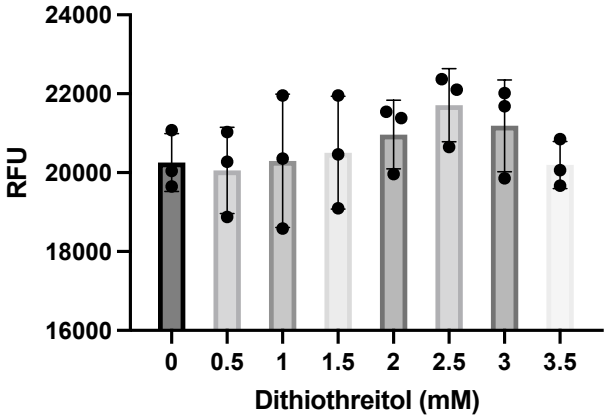


Figure 3.22: **Cell-free TXTL dithiothreitol optimisation.** The GFP output (measured in RFU) of in-house cell-free TXTL reactions supplemented with different DTT concentrations was compared, (+/- SD, n = 3).

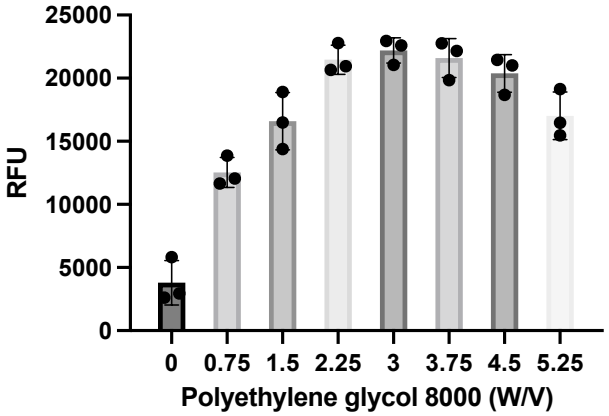


Figure 3.23: **Cell-free TXTL polyethylene glycol 8000 optimisation.** The GFP output (measured in RFU) of in-house cell-free TXTL reactions supplemented with different PEG 8k [W/V] concentrations was compared, (+/- SD, n = 3).

A similar trend to Figures 3.20 and 3.21 is again observed in Figure 3.23 with PEG 8k. Here, the optimum [W/V] concentration was found to be 3%. Similarly to K-glu, the omission of PEG 8k had quite a detrimental effect on the TXTL performance, eluding to its importance as an additive. These findings help to conclude that, as a molecular crowding agent which is added to help emulate the rate and equilibrium constants of biochemical reactions taking place within an *in vivo* cell [154], the concentration of PEG is critical to the efficacy of a TXTL reaction.

For the above optimisations, the plasmid concentration was fixed at 15 nM – this is the concentration that has previously been found to be optimal when using the pBEST plasmid in a cell-free

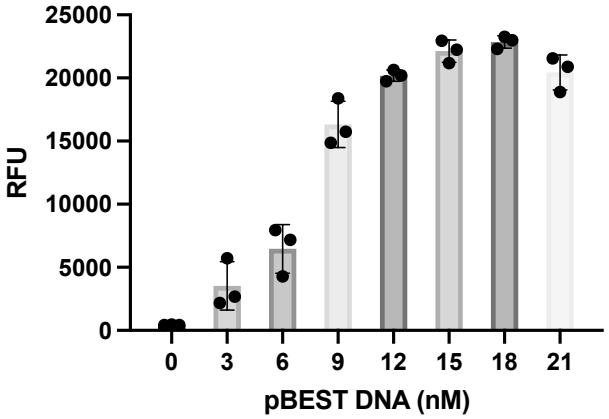


Figure 3.24: **Cell-free TXTL pBEST plasmid optimisation.** The GFP output (measured in RFU) of in-house cell-free TXTL reactions supplemented with different pBEST DNA concentrations was compared, (+/- SD, n = 3).

Table 3.1: **In-house cell-free TXTL setup after optimisations**

Concentration	**Component**
10 mg/mL	*E. coli* Rosetta 2-pAD-LyseR Crude Extract
1.5 mM	Amino Acid Mix
1x	Energy Solution 14x
3%	PEG 8000
5 mM	Magnesium-glutamate
80 mM	Potassium-glutamate
2.5 mM	DTT
18 nM	pBEST Plasmid DNA
up to 20 μL per rxn	Molecular Grade Water

system [72]. However, following the optimisations made in Figures 3.20 - 3.23, different pBEST DNA concentrations were tested in order to reveal the optimal value in this in-house system. As displayed in Figure 3.24, a slight increase in activity was achieved at a concentration of 18 nM, which was subsequently set as the optimal pBEST plasmid concentration for future experiments. It was important to optimise this as all of the ICP::NtNL expression constructs were cloned into the pBEST backbone in replace of GFP.

Following on from concluding the crude extract and additive optimisations, the final reaction setup conditions were defined (Table 3.1). This setup was used for all subsequent in-house TXTL reactions.

3.4.6 Rapid activity analysis of the fusion proteins with TXTL

Now that an in-house cell-free TXTL system had been built and optimised for pBEST plasmid expression, it could be used to rapidly test the light emitting capabilities of the ICP::NtNL fusions (which had previously been built into the pBEST backbone - Figure 3.5). Also, by doing so, this would confirm that the act of fusing NtNL to the phage ICPs does not cause suppression of spontaneous CtNL complementation to form the full NanoLuc enzyme. The TXTL workflow displayed in Figure 3.25 allows for rapid construct testing by simply expressing the various pBEST templates in TXTL reactions, followed by directly combining NtNL and CtNL reactions and finally, adding the furimazine substrate before measuring luminescence on a plate reader. Compared to a viable alternative, for example His-tag protein purification, this TXTL-powered workflow is preferable due to its speed and simplicity. Rather than go through the prolonged and laborious processes of bacterial transformation, culture growth, cell lysis and protein purification, with cell-free TXTL it is possible to obtain high yields of protein in a controlled environment within a few hours and with minimal lab work. To streamline the process further,

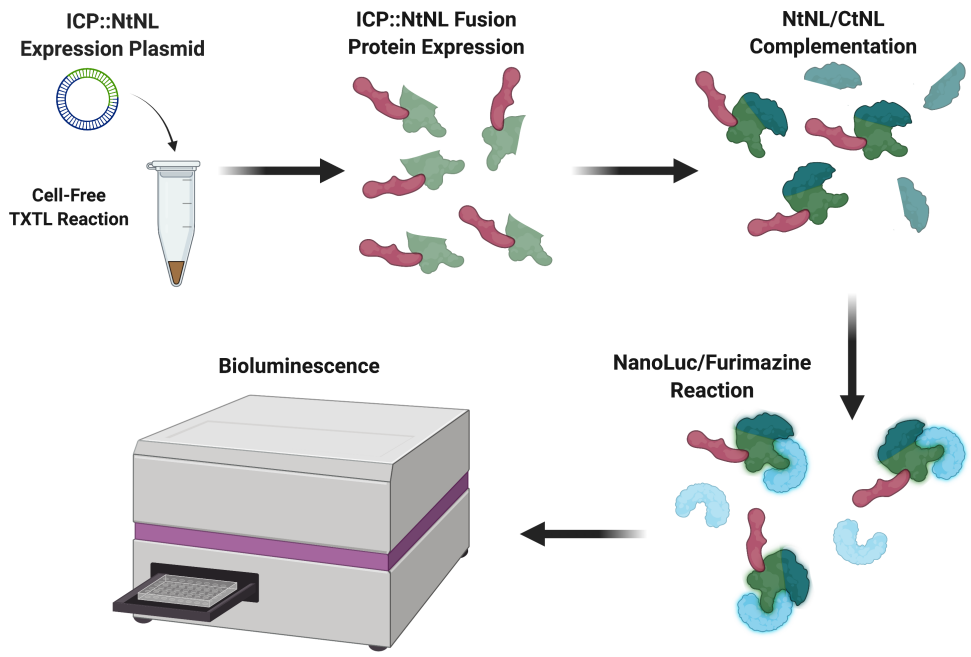


Figure 3.25: **Rapid TXTL-powered ICP::NtNL fusion activity test workflow.** The pBEST plasmid was expressed in a cell-free TXTL reaction to produce the internal capsid fusion protein. The C-terminus of NanoLuc (also expressed in TXTL) was then provided to facilitate spontaneous complementation. The NanoLuc substrate (furimazine) was added to the sample and the relative luminescence was measured by a plate reader.

it is also possible to express protein from linear DNA constructs in TXTL [155], thereby bypassing plasmid cloning, transformation, culture growth and plasmid Miniprep steps. If this is the desired route, then all that is needed is a PCR reaction of the gene construct to amplify the template DNA and then the addition of a RecBCD inhibitor, such as GamS or Chi6, to prevent the linear templates from being digested by the endogenous exonucleases that reside within the cell free reaction [156]. Furthermore, for a rapid testing application like the one displayed below, cell-free TXTL is ideal so that the minimal amount of work can be done prior to revealing whether the construct is successfully functional or not. If it is not, then the next construct can be tested rapidly too until the desired outcome is reached. This streamlined design-build-test-learn cycle is one of the key selling points of cell-free TXTL.

Following on from TXTL-testing the constructs, it was confirmed that both of the ICP::NtNL fusion proteins were capable of emitting light via spontaneous complementation with CtNL and enzymatic reaction with furimazine (Figure 3.26). As expected, all of the controls (CtNL only, ICP::NtNL only and GFP) failed to generate luminescence when combined with the Nano-Glo® assay reagents, confirming that the presence and complimentation of NtNL *and* CtNL was nec-

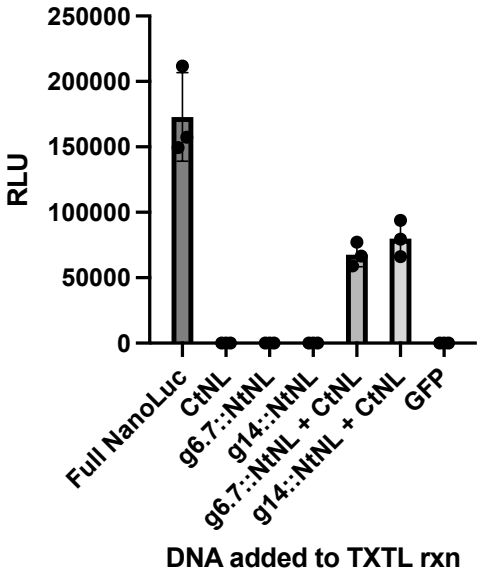


Figure 3.26: **ICP::NtNL fusion construct testing in cell-free TXTL.** Identical TXTL reactions were supplemented with 18 nM of the pBEST plasmid inserted with: full NanoLuc, CtNL, g6.7::NtNL, g14::NtNL or GFP. The reactions were incubated for 8 hours and then combined with an equal volume of Nano-Glo® assay buffer (containing furimazine). The fusion constructs were tested for luminescence both on their own and immediately after being combined with an equal volume of the CtNL TXTL reaction, (+/- SD, n = 3).

essary for furimazine catalysis. Furthermore, the gp6.7::NtNL and gp14::NtNL fusion proteins both gave comparable luminescence measurements to each other, which both represented approximately half the RLU value generated by the full NanoLuc positive control, suggesting that the act of splitting NanoLuc into two sub-units and fusing its N-terminus to a K1F ICP does have a negative impact on its ability to generate luminescence. However, this isn't detrimental to the success of this project as the light emitted by the fusions is significant compared to the controls (which do not emit light).

As it has been previously demonstrated that solely-acting CtNL or NtNL sub-units produce negligible amounts of light when presented with their substrate and necessitate full NanoLuc complimentation before a successful luminescence-emitting reaction can occur [128], it is not surprising that the ICP::NtNL fusions and CtNL tested in Figure 3.26 did not generate an RLU output when appraised on their own. Furthermore, the GFP sample was also not expected to emit chemiluminescent light when combined with the furimazine substrate, as GFP on its own generates fluorescence (detected at a longer wavelength compared to bioluminescence) and it has no known property to enzymatically react with any substrate to produce bioluminescence.

This finding (Figure 3.26) represented a significant achievement for the thesis project and confirmed that, if it was indeed possible to package these ICP::NtNL fusions inside the K1F capsids and subsequently retrieve them via host-induced ejection, viable signal generation would occur if the ejected fusion proteins were met with CtNL sub-units and furimazine substrate upon capsid expulsion. Furthermore, since the gp6.7::NtNL and gp14::NtNL fusion proteins performed comparably, it was tentatively anticipated that both engineered phage (K1F*g6.7::NtNL* and K1F*g14::NtNL*) could still viably act as the centerpiece of the proposed diagnostic phage-based *E. coli* K1 detection assay.

It was also decided to investigate if changing the incubation time of the NtNL:CtNL complementation reaction would impact the light emission capabilities of the newly formed NanoLuc complex. As the gp6.7 and gp14 NtNL fusions gave comparable results in the previous test, only gp6.7::NtNL was tested during this time course investigation as a representation of both ICP::NtNL fusions. As shown in Figure 3.27, incubating the ICP::NtNL and CtNL TXTL reactions with each other for 60 minutes prior to the Nano-Glo® assay clearly produces the highest RLU value. Moreover, as the incubation time incrementally shortens to 5 minutes, the RLU output value decreases. This indicates that, if optimal bioluminescent output is desired, it is highly important to allow the ICP::NtNL and CtNL reactions to complement with each other for a prolonged period of time prior to measuring their output.

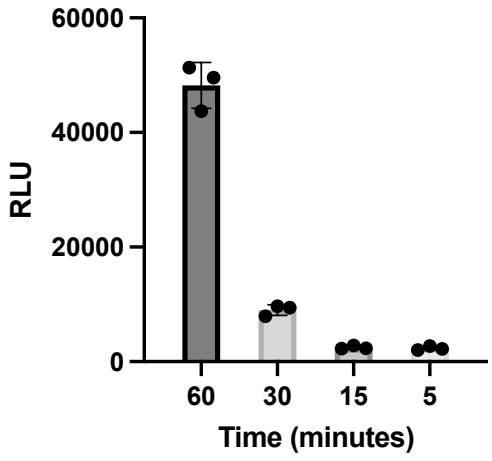


Figure 3.27: **NtNL:CtNL complementation reaction time-course data.** Identical TXTL reactions were supplemented with 18 nM of pBEST-g6.7::NtNL or pBEST-CtNL. The reactions were incubated for 8 hours and then combined with each other at equal volumes. The NtNL:CtNL reactions were left at room temperature to complement with each other and form the NanoLuc complex for either 5, 15, 30 or 60 minutes. Finally, an equal volume of Nano-Glo® assay buffer was added and the samples were measured for luminescence by a plate reader, (+/- SD, n = 3).

This finding is contradictory to what has previously been reported with these NtNL and CtNL sub-units, where the rapid reassembly of NanoLuc sub-units has been demonstrated with negligible difference in RLU output observed between 0 minutes and 60 minutes NtNL:CtNL incubation prior to bioluminescence measurement [126]. The significant difference observed here in Figure 3.27 between the varying time points can possibly be attributed to the only difference between this work and what has previously been reported - the ICP fusion - perhaps the addition of the fusion interferes, obstructs and slows down the process of NtNL and CtNL successfully complimenting with each other. Nevertheless, it clearly does not inhibit complimentation, which is the most important factor to consider. Furthermore, for the remainder of the work carried out in this thesis, the NtNL:CtNL incubation time would be conscientiously managed and standardised within experiments to ensure comparability between samples.

Following on from collecting the data displayed in Figures 3.26 and 3.27 and subsequently confirming the ICP::NtNL fusions' viability, activity and experimental reaction conditions, the next challenge - engineering the ICP::NtNL fusion phage - could begin.

3.4.7 Conclusions drawn and outlook for further investigations

The results presented in this first research chapter are, on a whole, representative of a successful start to the project. The rationale, which was the first result exhibited, offered some validation to the ICP::NtNL fusion diagnostic phage idea and gave precedent to all proceeding experiments. This rationale is important in the subsequent chapters so that estimates can be made on the necessary titer of engineered phage needed within the diagnostic assay. Subsequently, if a smaller-than-anticipated engineered phage titer is achieved but detectable signal is still generated, then conclusions can be drawn that the engineering process was more efficient than estimated at the start of this chapter. Furthermore, all DNA constructs required for the entirety of the thesis were successfully designed, ordered and cloned into their respective plasmid backbones. The correct sequences were all subsequently verified via the GATC LIGHTrun™ service (GATC Biotech, Ebersberg, Germany) and the sequencing data is available for all plasmids upon request. The impact of these plasmids on *E. coli* EV36 culture growth and K1F infection kinetics was also examined so that when the subsequent phage engineering experiments are attempted it is known what to expect with regards to each EV36 strains' culture characteristics. Finally, the establishment and optimisation of an in-house cell-free TXTL system allowed for the rapid testing and verification of the fusion constructs' luminescent activity and was subsequently then available and ready-to-go for all future construct expression and potential TXTL phage synthesis experimentation throughout the rest of the project.

One aspect of the work that could possibly be revised and improved upon is the relative propor-

tion of NtNL and CtNL sub-units comprising the split NanoLuc enzyme. The NtNL sub-unit used in the ICP fusion is a 65 amino acid fragment. Although, as previously alluded to, this is in theory small enough to fit through the K1F portal and tail without necessitating partial unfolding, it is not known whether it will negatively impact the internal composition and function of the phage when packaged into the capsid via an ICP fusion. If a negative impact or outright failure was observed in the subsequent engineering experiments, then perhaps reducing the size of the fusions' NtNL domain would increase the likelihood of phage viability due to the smaller exogenous addition having less of an obstructive or inhibitory intracapsid impediment. Furthermore, the NanoBiT complementation system has recently been developed whereby the NanoLuc enzyme is split into two disproportionately sized sub-units (1.3 kDa "Small BiT" and 18 kDa "Large BiT") which do not independently have light emitting capabilities [157]. The Small BiT, which consists of only 11 amino acids, could hypothetically be fused to the K1F ICPs and elicit reduced intracapsid obstruction compared to the 65 amino acid sub-unit currently utilised in the ICP::NtNL fusions. However, the NanoBiT complementation system was designed to study protein:protein interactions, whereby it is proposed that the NanoBiT sub-units are each fused to a different proteins and only undergo complimentation when the two proteins aggregate together. For this reason, Small BiT and Large BiT were developed so that they weakly associate with each other and predominantly undergo complimentation only when they are forced into the same proximity via the fused protein:protein interactions [158]. Moreover, in this thesis' ICP fusion phage-based diagnostic system, only the phage-borne NanoLuc sub-unit is to be fused to a protein (i.e. the ICP) and therefore, if the NanoBit system was used, only weak association would occur between the propelled ICP::Small BiT fusion and the external Large BiT domain. This is the exact reason why the currently used NtNL:CtNL system was chosen over NanoBiT in the first place, as the sub-units in this system are strongly associated and do not rely on protein:protein aggregation for complementation to occur [126]. If, however, the currently used 65 amino acid NtNL domain proves to be problematic and inhibitory of the phage engineering attempts, then perhaps a compromise will have to be made and the weakly associating NanoBiT system will be deployed as an alternative.

The success and efficacy of various phage engineering methodologies will be explored in the next chapter, with the primary goal being the successful generation of K1F*g6.7::NtNL* and K1F*g14::NtNL*.

Chapter 4

Investigating the efficacy of engineering the K1F internal capsid proteins

4.1 Abstract

Since the advent of the synthetic biology era, phage engineering has been used by researchers to modify and optimise various phage-types so that they can kill and/or detect their host with increased efficiency and higher specificity. The majority of previous attempts at creating a diagnostic phage have focused on inserting a reporter protein sequence into a region of the phage genome such that the phage structure and phenotype is not altered at all. However, in this case where the objective is to engineer the phage so that a fundamental and crucial process such as that of phage assembly or propagation is hindered, the genome editing process can become quite difficult. Novel results on this phage engineering approach are presented in this chapter. The internal capsid proteins play a key role in the most important processes during the existence of the phage and hence the process of inserting the fusion sequence into the K1F genome inevitably was difficult and time-consuming. Nevertheless, after many varying experimental approaches, a simple, yet effective non-genomic method was developed and the fusion was successfully encapsulated inside the phage capsid. In this chapter, the K1F internal capsid protein engineering peregrination is presented, concluding with the novel application of the successful non-genomic approach. This simple method offers an alternative for researchers who are attempting a high-risk phage edit, such as modifying the internal capsid proteins, thereby allowing the focus to be entirely on innovation and application rather than on extensive screening and experimental optimisation.

4.2 Introduction

This second research chapter will build upon the foundational work presented in the previous chapter, with the primary goal being the generation of either/both of the proposed engineered phage - K1F*g6.7::NtNL* and K1F*g14::NtNL*. Over the past few decades, an array of phage engineering methods have been established, which will be introduced below and experimented with throughout the chapter.

4.2.1 Traditional phage engineering

The process of homologous recombination (HR), which is described in Chapter 1, was the first technique utilised to engineer phage genomes and the first demonstrations of this process date back to the mid-20th century. In the early attempts, a process known as "phage cross" was developed [159], where researchers would infect one bacterial host cell with two genetically similar parental phage that display different phenotypes. Upon infection, cytoplasmic HR would occur between the two phage genomes and subsequently a mixed population would be generated - comprising the two WT parental phage and hybrid phage displaying chimeric phenotype characteristics derived from a combination of both parental phage [160].

Following on from the discovery of the phage cross technique, the same HR-based methodology was applied to exchange DNA between phage genomes and intracellular plasmid-borne recombination cassettes in order to edit the phage DNA and generate the desired genotype [161]. This allowed researchers more flexibility and control over the genetic edit due to the fact that they could insert any gene of interest between homology arms within the HR plasmid, rather than just creating a hybrid phage from two parental genomes. Furthermore, in an attempt to optimise the inefficient process of HR, another HR-based process known as "Recombineering" was developed, which carries out recombination with donor DNA in favourable environments such as the phage λ Red system [162]. This system comprises three genes which are usually provided on a plasmid: *gam* which inhibits the RecBCD complex from digesting the linear donor dsDNA, *exo* which converts the donor dsDNA into ssDNA via its 5' to 3' exonuclease activity and *bet* which anneals the donor ssDNA to the phage genome to generate the desired engineering result [163].

4.2.2 Phage engineering with CRISPR

The use of CRISPR for phage engineering has been vastly developed over the past few years and various combinations of point mutation, gene insertion and deletion have been demonstrated for phage targeting various bacterial species including *E. coli* [164], *K. pneumoniae* [165] and *S. thermophilus* [166]. It has also been confirmed that glycosylation and hydroxylmethyla-

tion of cytosine, as preferred by phage T4, does not prevent the action of the CRISPR/Cas9 system [164]. This approach allowed for the efficient and rapid sequential inactivation of individual genes – a very useful capability given that approximately half of the T4 genes have yet to be characterised [167, 168].

Guide RNAs can be used for precise genome engineering that, crucially, do not act on random sites [165]. This greatly facilitates the rapid selection of desired edits and coupled with the observation that CRISPR-based selection of targeted changes was successful in all tested plaques [166], demonstrates the shear potency of CRISPR. It further demonstrates that plasmid-borne spacers function with the same efficiency as their natural chromosomal counterparts. Thus, by placing spacers onto phage genomes under late promoters, the possibility of viability being restored to engineered diagnostic phage by recombination is reduced since the WT phage DNA should theoretically be digested. Furthermore, the use of plasmid-borne spacers to target genomic phage sequences was applied in an alternative approach that used CRISPR in conjunction with HR [169]. In this approach, CRISPR is directed to degrade non-edited genomes, thereby hugely expediting retrieval of successful recombinants. Further to that, in a recent paper, CRISPR was used to select genetically modified phage K1F, which had inserted a sfGFP gene on their minor protein (g10b), over wild type phage [94].

4.2.3 Research aims

The aims of this chapter are as follows:

1. To engineer phage K1F so that it incorporates the gp6.7::NtNL fusion protein within its internal capsid structure
2. To engineer phage K1F so that it incorporates the gp14::NtNL fusion protein within its internal capsid structure
3. If the phage engineering is successful, to demonstrate that the fusion can be inducibly released to generate on-demand bioluminescent signal

4.3 Methods

4.3.1 Homologous recombination and CRISPR selection

HR and CRISPR *E. coli* EV36 strains (EV36-pUC19-g6.7::NtNL, EV36-pUC19-g14::NtNL, EV36-pCas9-g6.7 and EV36-pCas9-g14) were made by following the protocols described in sections 2.3.5, 2.3.6, 2.3.7, 2.3.8 and 2.3.9. For HR, the phage were added to the host cells and incubated at 37 °C in LB medium with 100 μg/mL ampicillin until clearance was observed. The

phage lysate was then centrifuged at 3220 g for 15 minutes at 4 °C and the supernatent was retrieved and passed through a filtration unit to remove bacterial debris. If proceeding into CRISPR selection, this process was repeated three times with the pCas9 host strain. Phage lysates were diluted 100x before being added to the pCas9 host strain. Lysates were boiled, combined with screening primers, then processed and analysed via PCR and gel electrophoresis to check for successful genetic engineering.

4.3.2 Expression-assisted recombination

The two expression strains, EV36-pBEST-g6.7::NtNL and EV36-pBEST-g14::NtNL were made by following the protocols described in sections 2.3.5, 2.3.6, 2.3.7, 2.3.8 and 2.3.9. K1F-wt particles at an MOI of 0.005 were first added to the expression strains and incubated at 37 °C in LB medium with 100 μg/mL ampicillin until clearance was observed. The phage lysate was then centrifuged at 3220 g for 15 minutes at 4 °C and the supernatent was retrieved and passed through a filtration unit to remove bacterial debris. The filtered phage lysate was then diluted 100x and added to the pUC19 host for one round of HR propagation. Lysates were boiled, combined with screening primers, then processed and analysed via PCR and gel electrophoresis to check for successful genetic engineering.

4.3.3 Mass spectrometry phage protein analysis

Phage samples were passed through a 100 kDa Amicon® Ultra filtration device (Merck, New Jersey, USA) in order to carry out a buffer exchange and resuspend the phage particles in 50 mM ammonium bicarbonate. The samples were then subjected to reduction and alkylation via treatment with 10 mM tris(2-carboxyethyl)phosphine and 40 mM chloroacetamide, then incubated at 70 °C for 5 minutes. Next, the samples were treated with 0.1 μg/μL trypsin and incubated overnight at 37 °C for digestion to occur. The following day, the digested samples were treated with 1% trifluoroacetic acid in order to reduce the pH to 3-4. Finally, the samples were transferred to a Mass Spec vial and processed with a timsTOF PRO ion mobility Q-ToF mass spectrometer (Bruker, Massachusetts, USA). The generated data was analysed using the proteomics software, Scaffold.

4.3.4 Non-genomic *in vivo* ICP phage engineering

K1F-wt particles at an MOI of 0.005 were added to the EV36-pBEST-ICP::NtNL expression strains and incubated at 37 °C in LB medium with 100 μg/mL ampicillin until clearance was observed. The non-genomically engineered phage lysate was then centrifuged at 3220 g for 15 minutes at 4 °C and the supernatent was retrieved and passed through a filtration unit to remove bacterial debris. This filtered lysate was then passed through a 100 kDa Amicon® Ultra

filtration device (Merck, New Jersey, USA) three times (using a new device each time) before being resuspended in SM buffer.

4.3.5 Host-detection Nano-Glo® assay with ICP::NtNL engineered phage

1 mL of K1F phage that had been packaged with the ICP::NtNL fusion were propagated in 10 mL *E. coli* EV36-pBEST-CtNL cells at an OD_{600} value of 0.5 at 37 °C in LB medium with 100 μg/mL ampicillin until clearance was observed. The phage lysate was measured for luminescence based on the instructions of the Nano-Glo® Luciferase assay system and protocol described in section 2.3.10. An equal volume of Nano-Glo® assay reagent (containing the furimazine substrate) was mixed with the sample lysate, and this was then measured for luminescence in Relative Light Units (RLU) by a FLUOstar Omega microplate reader (BMG Labtech, Aylesbury, England).

4.4 Results and discussion

The initial K1F ICP genetic engineering strategy was heavily based on recently published work done by colleagues from the Sagona/Fehér groups [94], where a series of HR and CRISPR K1F propagations were carried out on EV36 strains bearing the relevant pUC19 and pCas9 plasmids in order to generate and select for engineered phage with a GFP-minor capsid protein fusion (K1F*g10b::GFP*). For the thesis work presented here, all engineering experiments were carried out on g6.7 and g14 in tandem and each attempt was biologically repeated three times over the course of a week. The biological repeats were analysed independently as-and-when they were carried out but also run on a single gel at the end of each phase for display purposes. For experiments that included HR followed by three rounds of CRISPR, a PCR amplification and gel was run for each propagation lysate, however - again to avoid an unnecessary overabundance of results and for display purposes - only the final propagation lysate PCR amplification/gel is shown when no difference is observed between the lysates.

4.4.1 Engineering screening

Following on from each phage engineering attempt, phage lysates were boiled at 100 °C to release their encapsulated genomic DNA. This DNA was subsequently screened via PCR for the presence of the recombinant sequences. Each engineering attempt would be screened for a positive control (WT primer pair + the K1F-wt genome to confirm the PCR was functioning correctly) and a negative control (NtNL engineering primer pair + the K1F-wt genome to confirm that no off-target, false-positive amplicons were being generated). PCR amplification screening of each phage lysate was carried out using the primer pairs highlighted in Figures 4.1 and 4.2.

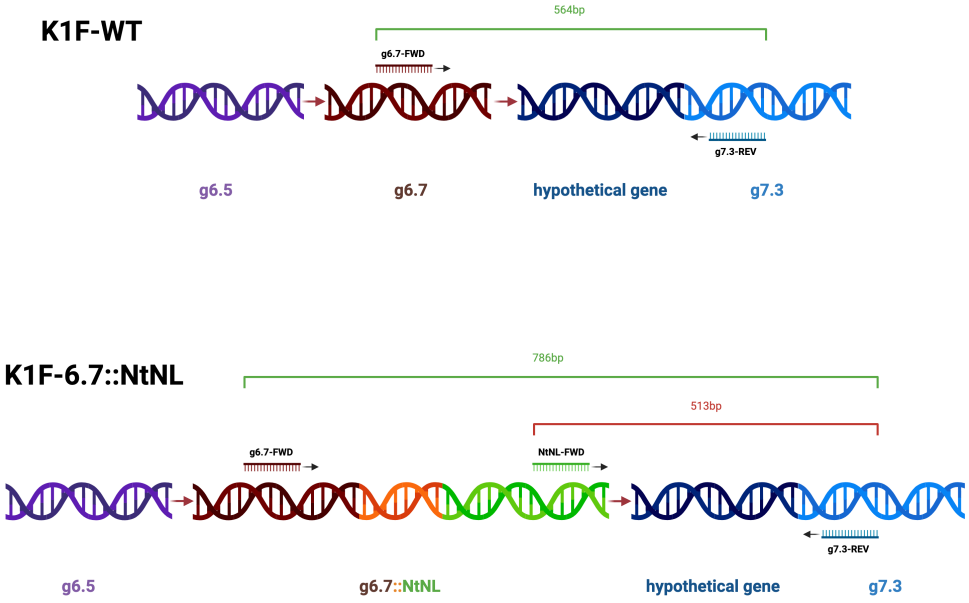


Figure 4.1: **PCR screening strategy for the identification of engineered phage K1F*g6.7::NtNL*.** PCR primers and the length of the three amplicons screened for are shown above the gene sequences of K1F-wt and K1F*g6.7::NtNL*.

The three amplicons screened for when attempting to generate and identify phage K1F*g6.7::NtNL* (Figure 4.1) were:

- **564bp fragment** generated by g6.7-FWD and g7.3-REV (i.e. the K1F g6.7 WT primer pair) - amplification of this fragment indicates the presence of K1F-wt
- **786bp fragment** also generated by g6.7-FWD and g7.3-REV - amplification of this larger fragment indicates the presence of K1F*g6.7::NtNL*
- **513bp fragment** generated by NtNL-FWD and g7.3-REV (i.e. the g6.7::NtNL engineering primer pair) - amplification of this larger fragment also indicates the presence of K1F*g6.7::NtNL*

It is important to note that the NtNL-FWD primer can bind to the inserted regions of the pBEST (Figure 3.5) and pUC19 (Figure 3.9) plasmids. However, as the g7.3-REV sequence is *only* present in the K1F-wt genome and *not* in any of the plasmids, the NtNL-FWD/g7.3-REV primer pair can only generate an amplification of the recombinant DNA - spanning the end of the g6.7::NtNL insert, the upstream hypothetical gene and the early region of g7.3. Similarly, as

NtNL-FWD does not bind anywhere in the K1F-wt genome, the risk of this primer pair generating a false-positive amplification of the K1F-wt DNA is negligible.

During the engineering attempts, the NtNL-FWD/g7.3-REV primer pair would be used to initially identify engineering success - as this pair will only produce an amplicon if HR takes place and the recombinant sequence is generated. Furthermore, if successful engineering is achieved, the g6.7-FWD/g7.3-REV primer pair would be deployed for sequencing purposes and also to verify that the larger 786bp amplicon can be generated - confirming the successful generation of K1F*g6.7::NtNL*. The reason why the g6.7-FWD/g7.3-REV primer pair is not used to begin with is because, if only a small amount of recombinant phage are initially produced (which is to be expected), then the amplification of the K1F-wt genome with this pair (the 564bp fragment) is likely to produce incredible amounts of background noise, enthralling the majority of the PCR reaction bandwidth and potentially *not* identifying the small amount of engineered genome present. Therefore, only when engineering success is confirmed and potentially enriched via CRISPR will this primer pair be used for engineering identification purposes.

Deploying the same strategy as g6.7, just with different primers, the three amplicons screened

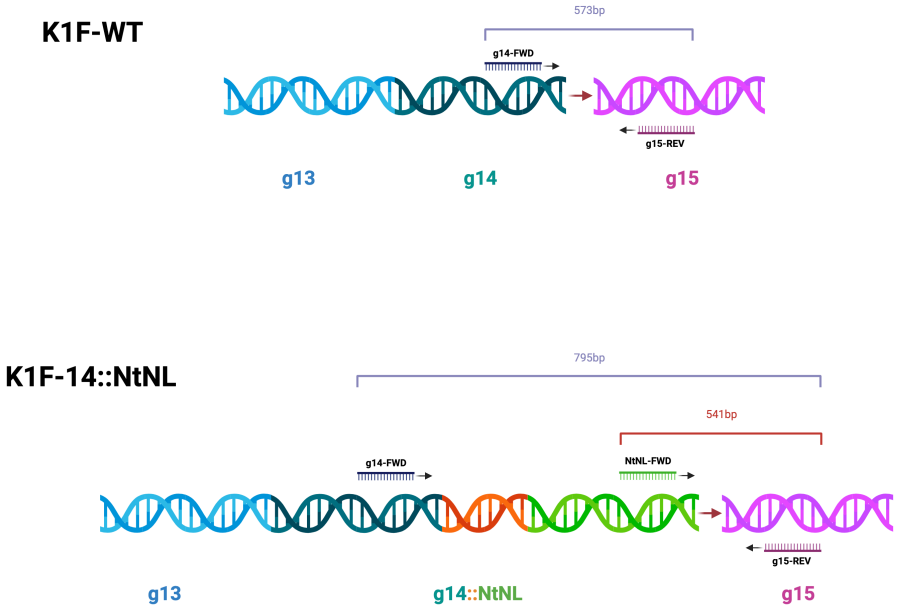


Figure 4.2: **PCR screening strategy for the identification of engineered phage K1F*g14::NtNL*.** PCR primers and the length of the three amplicons screened for are shown above the gene sequences of K1F-wt and K1F*g14::NtNL*.

for when attempting to generate and identify phage K1F*g14::NtNL* (Figure 4.2) were:

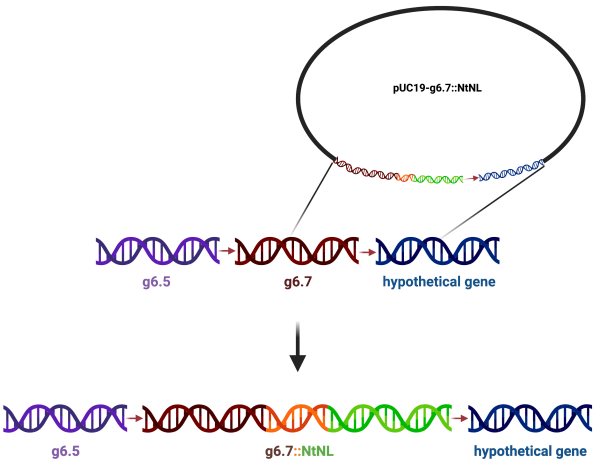
- **573bp fragment** generated by g14-FWD and g15-REV (i.e. the K1F g14 WT primer pair) - amplification of this fragment indicates the presence of K1F-wt
- **795bp fragment** also generated by g14-FWD and g15-REV - amplification of this larger fragment indicates the presence of K1F*g14::NtNL*
- **541bp fragment** generated by NtNL-FWD and g15-REV (i.e. the g14::NtNL engineering primer pair) - amplification of this larger fragment also indicates the presence of K1F*g14::NtNL*

Apart from the upstream genomic location, the plan for utilising these primer pairs and the reasons for doing so are indistinguishable from the aforementioned g6.7 engineering screening rationale, therefore to avoid replicatory prose, this won't be discussed further.

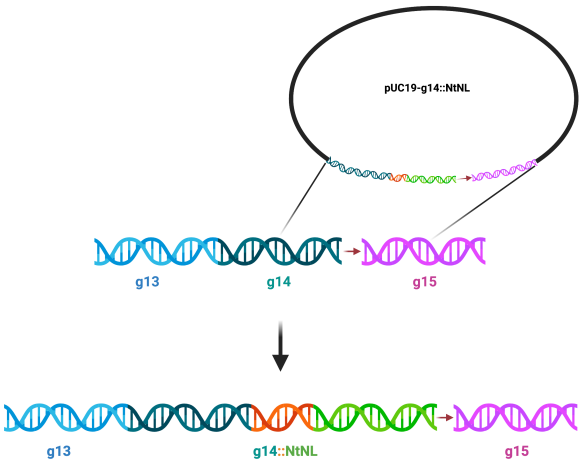
4.4.2 Homologous Recombination

To begin with, a single round of HR-mediated phage genome engineering was carried out by infecting *E. coli* EV36 cells harboring the pUC19-g6.7/g14::NtNL donor plasmid with K1F-wt and incubating the propagation until phage-induced lysis occurred. In theory, upon infection the K1F-wt genome is injected into the bacterial cytoplasm and HR is initiated at the sites specified by the homology arms on the donor plasmids. This can result in the insertion of the donor cassette into the phage genome. Due to there being no evolutionary advantage to the phage that uptake the genome modifications attempted here, they are naturally repelled by such activity and thus the success rate of HR is hindered. This typically results in, at best, a mixed phage progeny of predominantly WT and a small proportion of recombinant phage, or a failed engineering attempt. A visual representation of this for g6.7 and g14 is displayed in Figure 4.3.

In the desired scenario where HR does successfully occur, it is only likely that a very small percentage of the phage lysate population would be recombinant - such is the inefficient nature of the phage engineering HR process [166]. However, following on from the relevant PCR amplification (Figures 4.1 and 4.2), this small amount of engineered DNA should be amplified and subsequently appear on a gel. Gel electrophoresis images visualising the first PCR-amplified HR genome engineering attempts are displayed in Figure 4.4. Furthermore, as is shown in Figure 4.4, the initial HR attempt was not successfully identified during any of the biological repeats for g6.7 or g14. The only amplicon generated for both gene engineering attempts was the positive control (well 1 - i.e. the WT primer pair + the K1F-wt lysate), confirming that the PCR was functional. This, perhaps, is not surprising due to the well reported low-efficiency of solely-acting HR as a phage engineering strategy [159]. It was therefore decided to repeat this



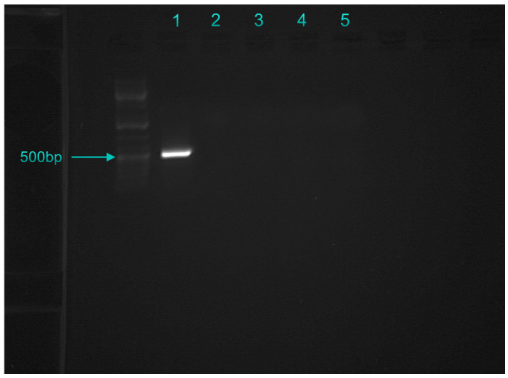
(a) g6.7::NtNL



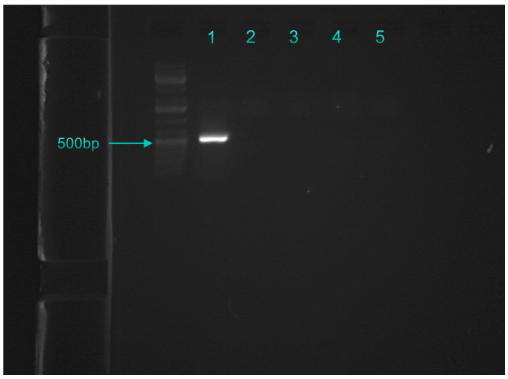
(b) g14::NtNL

Figure 4.3: **Homologous recombination between the K1F-wt genome and pUC19 ICP engineering cassettes.** Homologous recombination phage engineering for inserting the **a:** g6.7::NtNL and **b:** g14::NtNL fusion sequences into the K1F genome.

experiment with three rounds of CRISPR following the initial HR propagation - perhaps only a very small amount of HR occurred and was subsequently evading PCR amplification. Moreover, if HR engineering was successfully occurring, CRISPR selection would enrich the recombinant population, subsequently increasing the likelihood of engineered sequence PCR amplification and gel visualisation.



(a) g6.7::NtNL homologous recombination engineering attempts



(b) g14::NtNL homologous recombination engineering attempts

Figure 4.4: **Agarose gel electrophoresis PCR screening of the homologous recombination K1F engineering attempts.** PCR results obtained using the sequence identification strategy highlighted in 4.1 and 4.2 on **1a:** K1F-wt + EV36-wt w/ g6.7-FWD + g7.3-REV or **1b:** K1F-wt + EV36-wt w/ g14-FWD + g15-REV **(+ control)**, **2a:** K1F-wt + EV36-wt w/ NtNL-FWD + g7.3-REV or **2b:** K1F-wt + EV36-wt w/ NtNL-FWD g15-REV **(- control)**, **3a-5a:** K1F-wt + EV36_pUC19-g6.7::NtNL w/ NtNL-FWD + g7.3-REV or **3b-5b:** K1F-wt + EV36_pUC19-g14::NtNL w/ NtNL-FWD + g15-REV **(experimental results - 3 x biological repeats)**. The homologous recombination strategy is highlighted in 4.3a and 4.3b.

4.4.3 Homologous Recombination + CRISPR/Cas9

Following on from the generation of a mixed WT and recombinant phage population through HR, previous publications demonstrate that the recombinant population can be enriched using CRISPR/Cas selection [94,166,169,170]. To achieve this here, the possible WT/recombinant phage mix is propagated on *E. coli* EV36 cells transformed with either pCas9-g6.7 or pCas9-g14, which were presented in the previous chapter. For the selection of engineered phage (K1F*g6.7::NtNL* or K1F*g14::NtNL*), the pCas9 spacer sequences were designed to identify complementary protospacer sequences in the K1F-wt g6.7 or g14 regions flanked by a PAM site. The successful complementation of the spacer and protospacer activates the Cas9-mediated cleavage of the WT genome. By tactically designing the spacer sequence to span the site of insertion for the NtNL fusion, any engineered phage that are produced will not be detected by the pCas9 machinery - therefore providing them with an evolutionary advantage and allowing them to be enriched within the population. A schematic of this is displayed above in Figure 4.5.

Following on from one round of HR and three rounds of CRISPR selection, phage lysates were boiled at 100 °C to release their genomic DNA, which was subsequently screened via PCR for the presence of recombinant phage. However, gel electrophoresis analysis revealed that this second engineering attempt had also failed during each biological repeat for each gene (Figure 4.6). Moreover, after a consecutive series of failed HR + CRISPR selection engineering attempts,

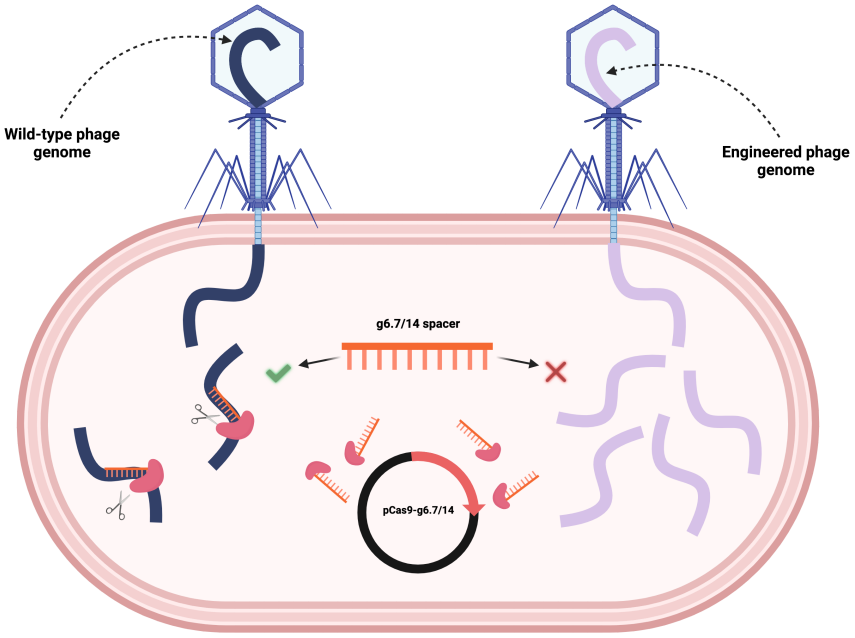
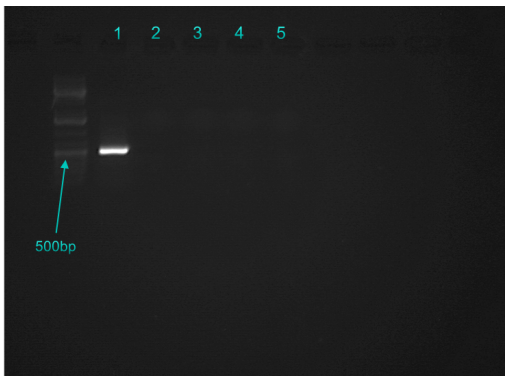


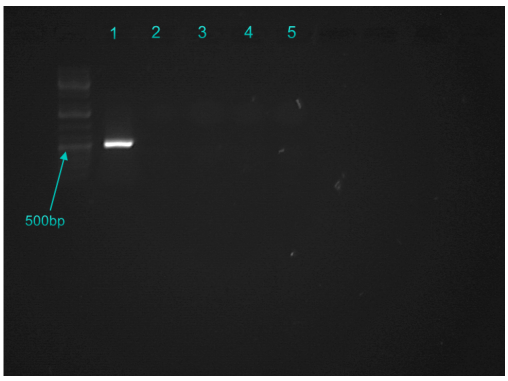
Figure 4.5: **K1F engineering CRISPR schema.** The pCas9 spacer sequence targets the K1F-wt genome and leads to targeted cleavage via the Cas9 enzyme. The g6.7/14::NtNL gene insert allow the engineered phage to evade cleavage and enriches the engineered population.

it was hypothesised that the cause of repeated failure could be an innate biological repulsion preventing K1F g6.7 and g14 from being fused to NtNL - i.e. the HR process isn't able to produce viable phage particles and therefore is not occurring frequently enough to be PCR amplified, even after CRISPR selection.

Regarding the work carried out in generating the K1F*g10b::GFP* phage [94], members of the Fehér group (who performed the genetic engineering experiments for that paper) have disclosed that the verification of successful HR via PCR amplification was not the difficult or unpredictable aspect of the work - they were able to generate engineered amplicons as soon as the experiments commenced. Moreover, the process of isolating a pure recombinant phage population via CRISPR selection was indeed the most challenging and time consuming aspect of the work. This suggests that, if HR was going to work, it would have already presented itself in the form of a successful engineering amplicon during the work carried out for Figures 4.4 and 4.6. Furthermore, the fact that the ICPs are a crucially integral proponent of phage propagation [106] - certainly, they are a more technical aspect of the phage structure compared to the minor capsid protein (g10b) which was the subject of the K1F*g10b::GFP* phage engineering - could confer that any modifications to their structure may result in a non-viable phage. Finally, even though the ICPs do not play an important role in non-propagating phage, the fact that they inhabit such a



(a) g6.7::NtNL homologous recombination + CRISPR selection engineering attempts



(b) g14::NtNL homologous recombination + CRISPR selection engineering attempts

Figure 4.6: **Agarose gel electrophoresis PCR screening of the homologous recombination + CRISPR selection K1F engineering attempts.** PCR results obtained using the sequence identification strategy highlighted in 4.1 and 4.2 on **1a:** K1F-wt + EV36-wt w/ g6.7-FWD + g7.3-REV or **1b:** K1F-wt + EV36-wt w/ g14-FWD + g15-REV **(+ control)**, **2a:** K1F-wt + EV36-wt w/ NtNL-FWD + g7.3-REV or **2b:** K1F-wt + EV36-wt w/ NtNL-FWD g15-REV **(- control)**, **3a-5a:** K1F-wt + 1 x propagation in EV36_pUC19-g6.7::NtNL then 3 x propagation in EV36_pCas9-cr6.7 w/ NtNL-FWD + g7.3-REV or **3b-5b:** K1F-wt + 1 x propagation in EV36_pUC19-g14::NtNL then 3 x propagation in EV36_pCas9-cr14 w/ NtNL-FWD + g15-REV **(experimental results - 3 x biological repeats)**. The CRISPR strategy is highlighted in Figure 4.5.

tightly packed and methodically organised interior space within the phage capsid, alongside the highly pressurised genomic DNA [113], might concur that even the slightest addition of exogenous protein (i.e. the NtNL fusion domain) causes severe internal disruption to the extent that the engineered phage is not viable.

Nevertheless, it was decided that the efforts to engineer the ICPs of K1F should not be halted due to this early impediment. Subsequently, an array of optimisation and selection strategies were devised and investigated, which are reported on below.

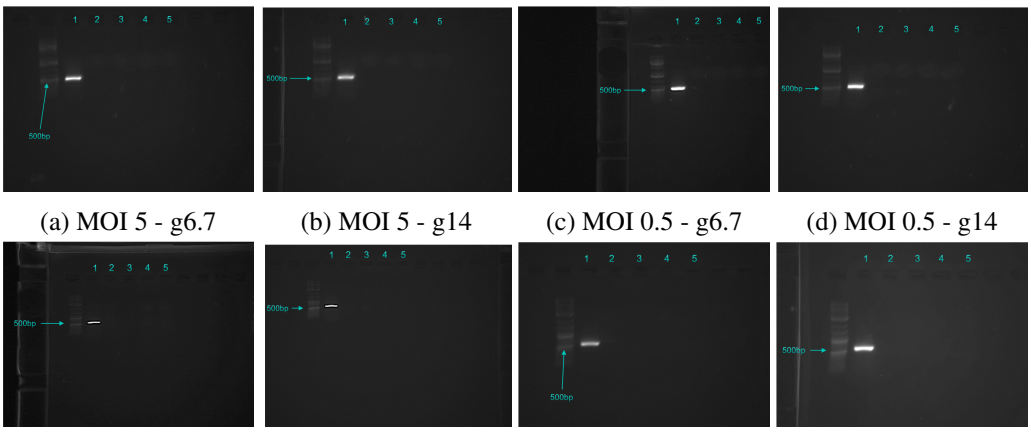
4.4.4 Consideration of the multiplicity of infection

The multiplicity of infection (MOI) is described as the ratio of agents to susceptible targets [171]. In the context of this work, the MOI is the number of plaque forming K1F phage units per EV36 bacterial cell (Figure 4.7). One reason why this may be impacting HR efficiency is that at higher MOI ratios (i.e. more phage particles than bacterial cells), it is conceivable that only a proportion and not all of the original phage population would be presented the opportunity to infect a bacterial cell and therefore confront the endogenous homologous material and potentially undergo HR. This is due to the possibility that at high MOI ratios the bacterial population is destroyed and diminished at an accelerated rate (high-speed culture clearances are observed

Figure 4.7: **Multiplicity of infection schema.** The MOI number relates to the number of phage particles per bacterial cell.

in the lab at high MOI ratios) and therefore, not every phage particle in the solution comes into contact with the endogenous material of a bacterial cell. This is problematic because, if HR engineering was successfully occurring, the 'unprocessed' phage population would dilute the engineered phage population, even before considering the efficiency of HR for phage particles that do infect a bacterial cell. In summary, at a lower MOI ratio, a confident presumption can be made that the entire phage population is given the opportunity to interact and exchange genetic material with the endogenous bacterial HR machinery. Furthermore, any K1F particles present in the post-infection lysate is very likely to be progeny that has burst out of a HR-plasmid bearing EV36 cell. Subsequently, the probability of HR occurring and detecting any successfully engineered phage via PCR amplification and gel analysis is increased.

For the previous two engineering attempts (HR only and HR + CRISPR), perhaps naively, little attention was paid to the MOI and subsequently a working MOI of 5 was being used. This ratio is unnecessarily high, potentially to an inhibitory degree. Therefore, for the subsequent experiments a serial dilution of MOI ratios were investigated: 5, 0.5, 0.05 and 0.005. As expected, the infection occurred at an increasingly slower rate as the MOI dilution increased, but phage from all of the MOI ratios were capable of lysing the bacterial culture eventually. For this MOI investigation, one round of HR and three rounds of CRISPR were carried out, as in the previ-



(a) MOI 5 - g6.7 (b) MOI 5 - g14 (c) MOI 0.5 - g6.7 (d) MOI 0.5 - g14

(e) MOI 0.05 - g6.7 (f) MOI 0.05 - g14 (g) MOI 0.005 - g6.7 (h) MOI 0.005 - g14

Figure 4.8: **Agarose gel electrophoresis PCR screening of the MOI optimisations for homologous recombination + CRISPR selection K1F engineering attempts.** PCR results obtained using the sequence identification strategy highlighted in 4.1 and 4.2 on **1a/c/e/g:** K1F-wt + EV36-wt w/ g6.7-FWD + g7.3-REV or **1b/d/f/h:** K1F-wt + EV36-wt w/ g14-FWD + g15-REV **(+ control)**, **2a/c/e/g:** K1F-wt + EV36-wt w/ NtNL-FWD + g7.3-REV or **2b/d/f/h:** K1F-wt + EV36-wt w/ NtNL-FWD g15-REV **(- control)**, **3-5a/c/e/g:** K1F-wt + 1 x propagation in EV36_pUC19-g6.7::NtNL then 3 x propagation in EV36_pCas9-cr6.7 w/ NtNL-FWD + g7.3-REV or **3-5b/d/f/h:** K1F-wt + 1 x propagation in EV36_pUC19-g14::NtNL then 3 x propagation in EV36_pCas9-cr14 w/ NtNL-FWD + g15-REV **(experimental results - 3 x biological repeats)**. The MOI was set to 5, 0.5, 0.05 and 0.005 phage particles per bacterial cell for each respective optimisation.

ous engineering attempt. The starting MOI was set for the HR propagation round and then each lysate was diluted 100-fold before infecting the next CRISPR selection culture. Displayed above in Figure 4.8 are the gel electrophoresis images resulting from PCR screening for all of the MOI optimisation engineering attempts.

As shown in the gel electrophoresis images above, no successful recombination was detected for any of the MOI ratios. Again, the only amplicon generated throughout was the positive control (i.e. the 564bp fragment for g6.7 - Figure 4.1 - and the 573bp fragment for g14 - Figure 4.2). Although this investigation did not yield satisfactory engineering results, the exercise of considering MOI ratios in the context of phage engineering was useful and it was decided that an MOI of 0.005 would be set for all subsequent engineering experiments due to the increased and more widespread exposure of phage material to the hosts' intracellular engineering machinery.

4.4.5 Simultaneous Homologous Recombination + CRISPR Selection

Following on from the continuation of genetic engineering failures, a conclusion was drawn that it was the HR process that was failing. As aforementioned, previous work indicates that a PCR amplification signal should be detectable on a gel after just one round of HR, then

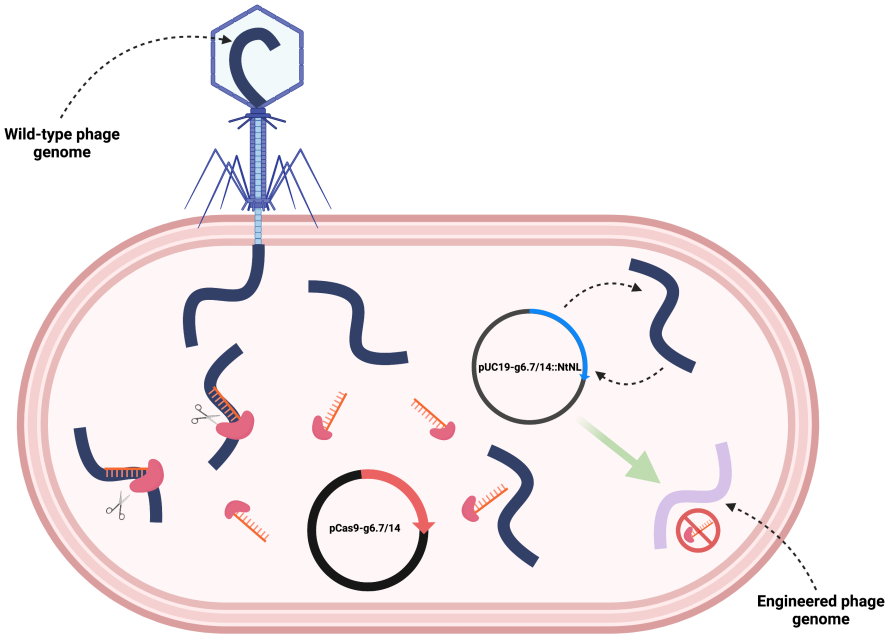


Figure 4.9: **Simultaneous homologous recombination + CRISPR selection schema.** It is envisioned that the selective presence of the pCas9 plasmid increases the engineering efficacy of the pUC19 plasmid in real time.

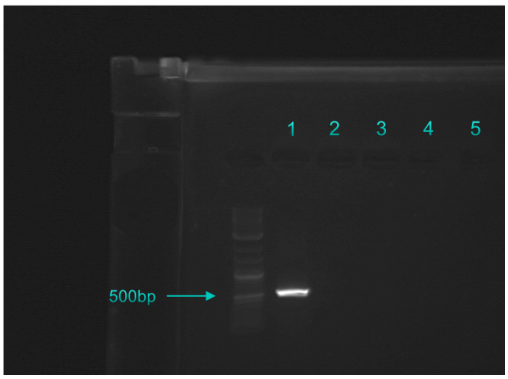
the signal should progressively increase as the CRISPR rounds are completed. Therefore, due to the fact that no bands are appearing at all (other than the positive controls), the culpability of this can be attributed directly to the HR process. Even if some degree of HR was occurring, it must be too inefficient to render a detectable signal or pass on enough recombinant phage to the subsequent CRISPR selection cultures.

It was therefore decided that perhaps an additional, simultaneous selective pressure was necessary to manufacture a more compelling engineering process for the phage. Furthermore, by having the pCas9 machinery present during the HR propagation, this might result in real-time selection for phage that have undergone HR with the pUC19 plasmid and does not rely on these recombinants being passed over to subsequent propagations. In fact, a literature review revealed that this simultaneous strategy has in fact been successfully deployed before for efficiently engineering phage targeting *V. cholerae* [172]. The simultaneous HR + CRISPR strategy is visualised in Figure 4.9.

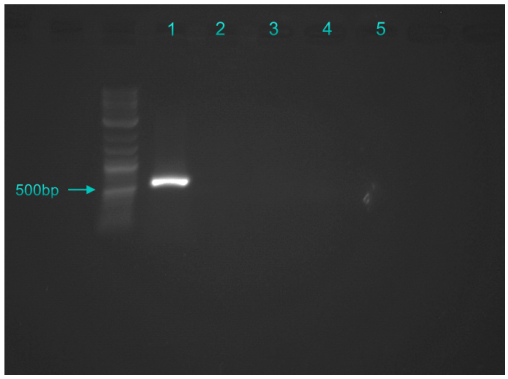
An alternative, albeit similar mechanism has also previously been presented in a paper whereby CRISPR and HR plasmids are used in tandem. However, instead of the CRISPR plasmid adding a real-time selective pressure as is hypothesised here, this publication suggests that the directed

nucleic damage caused by the Cas-mediated cleavage stimulates homology-directed repair with the simultaneously present HR donor plasmid, which subsequently increases engineering efficiency [173]. Either mechanism, whether it be via CRISPR-mediated real-time HR selection or CRISPR-stimulating homology-directed repair, would produce the desired outcome (i.e. engineered phage) *and* they both can be achieved with the same protocol and setup (i.e. both mechanisms rely on the CRISPR and HR plasmids being simultaneously present in the host cells but do not specifically need anything else). Subsequently, due to the promising prior art and dual-mechanism engineering potential, these experiments were approached with some optimism.

However, prior to committing to the simultaneous HR + CRISPR experiments, it was first necessary to verify the plasmid compatibility of pCas9 and pUC19. The two key items to consider are antibiotic resistance and origin of replication (*ori*). Firstly, the two plasmids have different antibiotic resistance genes - chloramphenicol (which binds to the *E. coli* ribosomes and therefore inhibits protein synthesis) resistance for pCas9 and ampicillin (which fatally prevents *E. coli* cells from constructing their cell walls) resistance for pUC19. Furthermore, the pUC19 *ori* is a pMB1 derivative, and the pCas9 *ori* is p15A. Fortunately, pMB1 and p15A do not compete for the same replication machinery and therefore are compatible within the same cell, so co-



(a) g6.7::NtNL simultaneous homologous recombination + CRISPR selection engineering attempts



(b) g14::NtNL simultaneous homologous recombination + CRISPR selection engineering attempts

Figure 4.10: **Agarose gel electrophoresis PCR screening of the simultaneous homologous recombination + CRISPR selection K1F engineering attempts.** PCR results obtained using the sequence identification strategy highlighted in 4.1 and 4.2 on **1a:** K1F-wt + EV36-wt w/ g6.7-FWD + g7.3-REV or **1b:** K1F-wt + EV36-wt w/ g14-FWD + g15-REV **(+ control)**, **2a:** K1F-wt + EV36-wt w/ NtNL-FWD + g7.3-REV or **2b:** K1F-wt + EV36-wt w/ NtNL-FWD g15-REV **(- control)**, **3a-5a:** K1F-wt + 3 x propagation in EV36_pUC19-g6.7::NtNL_pCas9-cr6.7 w/ NtNL-FWD + g7.3-REV or **3b-5b:** K1F-wt + 3 x propagation in EV36_pUC19-g14::NtNL_pCas9-cr14 w/ NtNL-FWD + g15-REV **(experimental results - 3 x biological repeats)**. The simultaneous homologous recombination + CRISPR selection strategy is highlighted in Figure 4.9.

transformation of pUC19-g6.7/14::NtNL and pCas9-g6.7/14 into *E. coli* EV36 cells was carried out and subsequently, the simultaneous HR + CRISPR selection experiments commenced.

As shown in Figure 4.10, PCR amplification screening for all biological repeats failed to detect any recombinant phage after three rounds of simultaneous HR + CRISPR propagation. Due to the previously published demonstrations of utilising HR, CRISPR or a simultaneous combination of the two in order to successfully engineer phage [94, 159, 169, 172, 173], alongside the fact that, hitherto, the experimental design and setup used for the ICP engineering attempts was heavily based on the K1F*g10b::GFP* publication [94] and was advised on by researchers that carried out the K1F*g10b::GFP* engineering work, it is not expected that the design, setup or precedent of the experiments are causations of the ICP engineering failure. Furthermore and as previously eluded to, one of the more likely possible reasons for this repeated ICP engineering failure is an innate biological intolerance for the modification of the K1F ICPs. However, this suggestion should not be considered a fact and further investigation is necessary before arriving at any conclusions.

4.4.6 Expression-assisted recombination

At this stage it was quite clear that, by using previously established methodology, the HR process was not occurring between the K1F-wt genome and the pUC19 donor plasmid for g6.7 or g14 regardless of the optimisation or selection strategy that was deployed. As a result of this, the next set of experiments presented were based on novel, hypothetical ideas that postulated a new mechanism for compelled phage engineering. If successful, it is hoped that this new method could be used to force the engineering process through in scenarios where the phage is at a severe disadvantage after the engineered DNA sequence alteration is incorporated into the genome (so much so that traditional engineering methods are rendered obsolete).

When hypothesising about what might exhibit a more complying phage towards genetic engineering when the genetic adaption itself is significantly unfavourable, one idea that transpired was that if the phage already possessed the adaption in its physical form (i.e. the protein that the engineered DNA codes for), then it may be in a more favourable condition for genetic engineering than a completely WT phage because the adaption is already present in the phenotype. In the context of this work, if the ICP::NtNL fusion protein can be packaged inside the phage capsid in a non-genetic manner prior to the HR propagation, then perhaps that will create a more favourable environment for a successful HR engineering. A schema describing this hypothesis, named "Expression-assisted Recombination" (ExRec), is displayed in Figure 4.11. A more detailed account of the ExRec protocol is given in this chapter's Methods section.

It is noted that this hypothesis is heavily experimental and does not lean on any prior art for

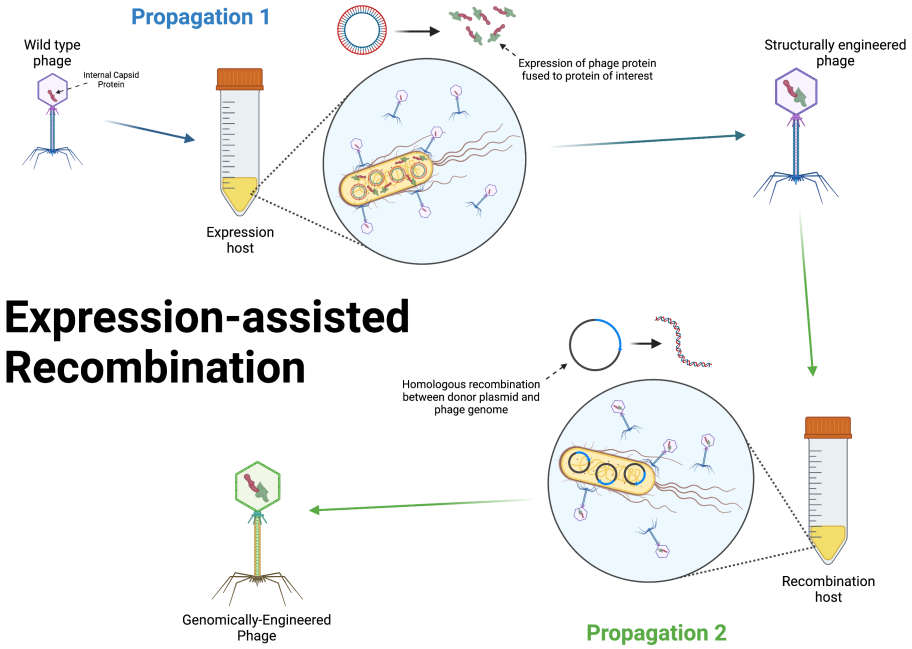
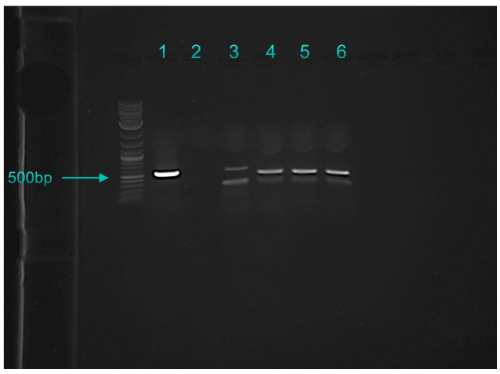


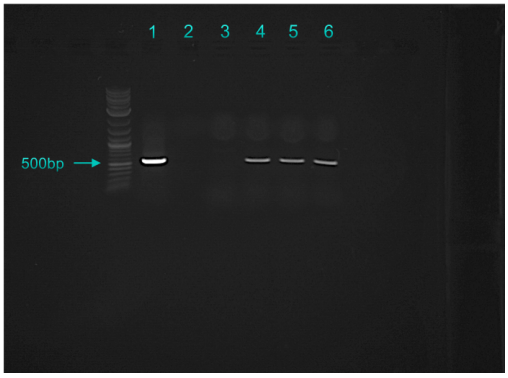
Figure 4.11: **Expression-assisted recombination hypothetical strategy.** One round of propagation is carried out with K1F-wt + EV36-pBEST (expression strain) then the lysate is propagated in EV36-pUC19 (recombination strain) in an attempt to generate the engineered phage.

credibility or premise, however, at this stage there wasn't a significant amount of time or money at stake due to the expression and HR plasmids already being constructed - only a simple transformation of the expression plasmid into EV36 was necessary prior to the ExRec experiments being attempted. The PCR screening gel electrophoresis results for ExRec are shown in Figure 4.12. Quite surprisingly, initial gel analysis suggested that successful HR had occurred, with the g6.7::NtNL (513bp) and g14::NtNL (541bp) fusion sequences being amplified by the NtNL primer pairs for the first time thus far. Moreover, this successful engineered amplicon generation occurred during each biological repeat of ExRec (wells 4, 5 and 6 on both gels). If further investigations (e.g. verification of the engineered genome via sequencing) proved this initial finding to be true, then the ExRec protocol could represent a significant step forward within the phage engineering community - especially when applied to precarious engineering attempts where all established methods fail to produce the desired result.

Whilst there were no off-target amplicons generated for g14 - which displayed exactly the results that were desired (Figure 4.12b), the g6.7 screening did yield two unanticipated amplicons (Figure 4.12a). Well 3 on the g6.7 gel implies that two off-target amplifications occurred during the expression propagation. The sample used in well 3 was taken from the expression lysate (i.e. be-



(a) g6.7::NtNL expression-assisted recombination engineering attempts



(b) g14::NtNL expression-assisted recombination engineering attempts

Figure 4.12: **Agarose gel electrophoresis PCR screening of the expression-assisted recombination (ExRec) K1F engineering attempts.** PCR results obtained using the sequence identification strategy highlighted in 4.1 and 4.2 on **1a:** K1F-wt + EV36-wt w/ g6.7-FWD + g7.3-REV or **1b:** K1F-wt + EV36-wt w/ g14-FWD + g15-REV **(+ control)**, **2a:** K1F-wt + EV36-wt w/ NtNL-FWD + g7.3-REV or **2b:** K1F-wt + EV36-wt w/ NtNL-FWD g15-REV **(- control)**, **3a:** K1F-wt + EV36_pBEST-g6.7::NtNL w/ NtNL-FWD + g7.3-REV or **3b:** K1F-wt + EV36_pBEST-g14::NtNL w/ NtNL-FWD + g15-REV **(experimental control)**, **4a-6a:** K1F-wt + 1 x propagation in EV36_pBEST-g6.7::NtNL then 1 x propagation in EV36_pUC19-g6.7::NtNL w/ NtNL-FWD + g7.3-REV or **4b-6b:** K1F-wt + 1 x propagation in EV36_pBEST-g14::NtNL then 1 x propagation in EV36_pUC19-g14::NtNL w/ NtNL-FWD + g15-REV **(experimental results - 3 x biological repeats)**. The expression-assisted recombination strategy is highlighted in 4.11.

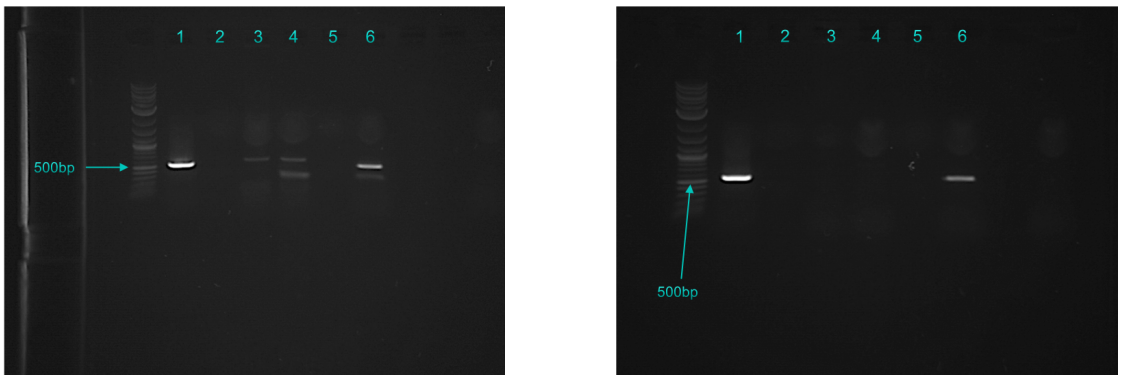
fore the HR propagation), therefore, the causation of the off-target amplifications could be either due to a pBEST plasmid amplification or HR occurring between the phage genome and homologous regions of the pBEST plasmid. The pBEST-g6.7::NtNL plasmid and K1F-wt genome do share homology over the entire g6.7 sequence, however, if recombination was to occur it would be entirely dependent on homology upstream of the NtNL sequence, as there is no homology at all after the end of g6.7. This is possible, but unlikely due to the usual requirement of upstream *and* downstream homology, as is present on the HR plasmid. It is also conceivable that the off-target amplicons were generated from the pBEST plasmid alone, as the NtNL-FWD primer does bind to the insert region of this plasmid. Hypothetically, it would only take the g7.3-REV primer to incorrectly bind at two locations upstream to produce the two off-target amplicons. However, after carefully studying the pBEST sequence, it is not at all obvious where the REV primer might incorrectly bind to produce the two off-target amplicons that are observed (which are approximately 300-400bp and 600-700bp in size - Figure 4.12a). Furthermore, the shorter off-target amplification for g6.7 appears to persist after the HR propagation as displayed in wells 4, 5 and 6 - which are biological repeats taken from the final lysate of the ExRec protocol. However, the longer off-target amplification is not observed after the subsequent HR propagation, suggesting that the template material that generated the smaller amplicon was transferred to the subsequent

HR propagation, whereas the template material that generated the larger off-target amplicon is specific to the EV36-pBEST-g6.7::NtNL expression strain and ceased to exist following on from the expression propagation. Reassuringly, no amplification of the correct size is observed after the expression propagation, removing the possibility of a false-positive result culminating from the expression propagation. Also, the fact that no off-target amplifications occurred for g14 indicated a successful HR process.

It can confidently be deduced that the inclusion of an expression propagation before the round of HR propagation (i.e. the ExRec protocol) is what has instigated the generation of these amplicons - both on- and off-target. This is because the NtNL primer pairs for g6.7 and g14 have been repeatedly utilised to screen the HR propagation lysates (Figures 4.4, 4.6, 4.8 and 4.10) and therefore it is known that the HR propagation alone (or with CRISPR) has not been capable of generating amplicons with the NtNL primer pairs - conspicuously attributing the sudden amplicon generation capabilities to the combination of expression and HR propagations. Furthermore, whilst the presence of DNA bands of the correct size resulting from the NtNL primer pairs for the first time was a promising observation, it was decided that a further round of controls was necessary to help confirm if the genetic engineering had been successful or not.

In Figure 4.13, the gel electrophoresis results from PCR amplifications with the screening primers are shown. Alongside the standardised positive and negative controls (wells 1 and 2), PCR results are displayed for: the pBEST expression plasmid only (well 3), the ExRec expression lysate (well 4), the pUC19 HR plasmid only (well 5) and the ExRec recombination lysate (well 6). As displayed in Figure 4.13b, these extra controls confirmed that the g14::NtNL engineering primer pair was not generating amplicons from either of the g14 plasmids (pBEST or pUC19), additionally, no recombination was occurring during the expression propagation for g14 - suggesting that the amplicon visible in well 6 is in fact a result of successful HR and the NtNL fusion sequence being inserted into the phage genome.

For g6.7, the faint band visible in well 3 indicates that, despite no viable location being found in the plasmid sequence, the g7.3-REV primer has bound 700bp upstream of NtNL in pBEST-6.7::NtNL and together with NtNL-FWD, generated an amplicon. This is a reasonable conclusion, as this amplicon is again generated from the expression lysate (well 4), where pBEST is present, but is not generated from the subsequent recombination lysate (well 6), where pBEST is not present. Well 4 also suggests that the smaller off-target amplicon (approximately 400bp) is an artifact of some level of recombination occurring between the pBEST plasmid and K1F-wt genome, which would produce an incorrectly engineered phage - presumably with a slightly smaller genetic region around the site of insertion (i.e. around 400bp instead of 513bp). This



(a) g6.7::NtNL controls

(b) g14::NtNL controls

Figure 4.13: **Agarose gel electrophoresis PCR screening of the controls for verifying K1F*g6.7::NtNL* and K1F*g14::NtNL* engineering success.** PCR results obtained using the sequence identification strategy highlighted in 4.1 and 4.2 on **1a:** K1F-wt + EV36-wt w/ g6.7-FWD + g7.3-REV or **1b:** K1F-wt + EV36-wt w/ g14-FWD + g15-REV **(+ control)**, **2a:** K1F-wt + EV36-wt w/ NtNL-FWD + g7.3-REV or **2b:** K1F-wt + EV36-wt w/ NtNL-FWD g15-REV **(- control)**, **3a:** pBEST-g6.7::NtNL w/ NtNL-FWD + g7.3-REV or **3b:** pBEST-g14::NtNL w/ NtNL-FWD + g15-REV, **4a:** K1F-wt + EV36_pBEST-g6.7::NtNL w/ NtNL-FWD + g7.3-REV or **4b:** K1F-wt + EV36_pBEST-g14::NtNL w/ NtNL-FWD + g15-REV, **5a:** pUC19-g6.7::NtNL w/ NtNL-FWD + g7.3-REV or **5b:** pUC19-g14::NtNL w/ NtNL-FWD + g15-REV, **6a:** K1F-wt + 1 x propagation in EV36_pBEST-g6.7::NtNL then 1 x propagation in EV36_pUC19-g6.7::NtNL w/ NtNL-FWD + g7.3-REV or **6b:** K1F-wt + 1 x propagation in EV36_pBEST-g14::NtNL then 1 x propagation in EV36_pUC19-g14::NtNL w/ NtNL-FWD + g15-REV.

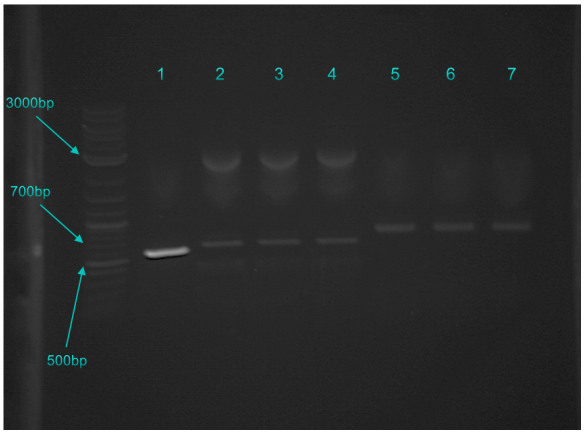
smaller amplification is also detected after the recombination propagation (well 6), suggesting the incorrectly engineered phage had survived and propagated. Whilst these off-target amplifications for g6.7 were unexpected and undesirable, there was still no sign of a false-positive result and therefore it was still encouraging that, after completing the ExRec protocol, the correctly sized amplicon was generated for both g6.7 and g14. This suggests that the hypothesis of creating a more 'favourable' phage for genetic engineering by first propagating the phage in a host containing the engineering subject (i.e. the fusion protein) - thereby potentially non-genomically packaging the fusion into the phage capsids - was worth investigating further and had possibly produced the engineered phage - K1F*g6.7::NtNL* and K1F*g14::NtNL*.

4.4.7 Plaque PCR analysis

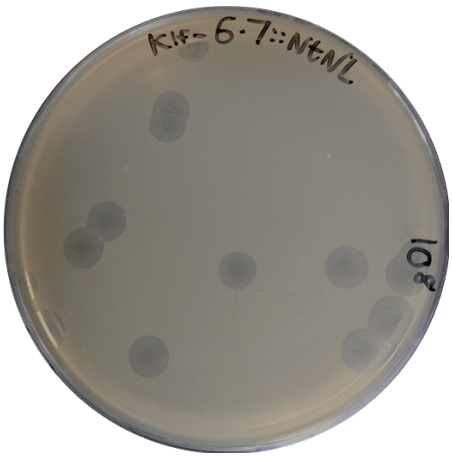
After gaining preliminary confirmation that the engineered phage had been produced via the ExRec protocol, the next step was to isolate the engineered population through a plaque assay so that a purified stock of the engineered phage could be made and stored. Within a plaque assay, each individual plaque represents a uniform phage population generated from a single plaque forming unit (PFU) infecting a single host cell [174], therefore, by isolating a single plaque which is capable of generating the correct amplicon with the NtNL engineering primer

pair, the engineered phage can be isolated and then subsequently purified and stored for future use. The gel electrophoresis results displayed in Figure 4.14a show the plaque PCR amplicons generated from three g6.7 plaques (wells 2-4) and three g14 plaques (wells 5-7). The plates that the plaques were taken from are also shown (Figures 4.14b and 4.14c). From these plates, the PFU calculation was applied to estimate the phage titers (in PFU/mL) after the ExRec protocol: 3.1×10^8 for K1F*g6.7::NtNL* and 1.0×10^8 for K1F*g14::NtNL*.

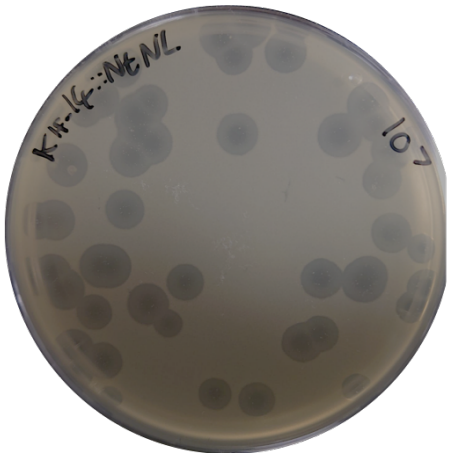
The amplicons generated from the plaques derived from ExRec-engineered K1F*g6.7::NtNL* (wells 2-4) in Figure 4.14a appear to be the same incorrectly sized PCR products that were



(a) Plaque PCR results for ExRec-generated engineered phage on WT-EV36.



(b) Plaque assay of ExRec-generated engineered K1Fg6.7::NtNL on EV36-wt.



(c) Plaque assay of ExRec-generated engineered K1Fg14::NtNL on EV36-wt.

Figure 4.14: **Plaque PCR and plaque assay results for ExRec-generated engineered phage on EV36-wt.** PCR results obtained using the sequence identification strategy highlighted in 4.1 and 4.2 on **1:** K1F-wt + EV36-wt w/ g6.7-FWD + g7.3-REV **(+ control)**, **2-4:** K1Fg6.7::NtNL + EV36-wt w/ NtNL-FWD + g7.3-REV, **5-7:** K1Fg14::NtNL + EV36-wt w/ NtNL-FWD + g15-REV.

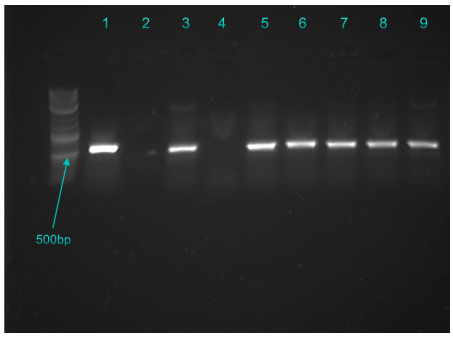
previously observed for the expression lysate in Figure 4.12a, along with an additional off-target amplicon approximately 3000bp in size. For K1F*g14::NtNL* plaques, the amplicon produced had shifted to approximately 700bp (wells 5-7). A correct amplification of the engineered phage for neither g6.7::NtNL nor g14::NtNL was observed, suggesting that the desired engineered phage had not been isolated via the plaque assay. It should also be noted that several more plaques were tested for each phage and identical results were found.

When considering why this may have happened, one key difference to note between the propagation processes that produced the ExRec PCR results and plaque PCR results is that the latter was propagated with EV36-wt, whereas the former was with EV36-pUC19-g6.7/14::NtNL (i.e. EV36 cells bearing the HR plasmid). Conversations with the Fehér group (who engineered the K1F*g10b::GFP* phage [94]), disclosed that engineered phage populations can become unstable when they are not propagated with the HR plasmid present. This is a potential explanation for K1F*g6.7/14::NtNL* when propagated on EV36-wt prior to plaque PCR screening. Perhaps, because the engineered inclusion of the ICP fusion construct provides an evolutionary disadvantage to the phage (as aforementioned), as soon as the engineering pressure (i.e. the HR plasmid) is removed, the phage rapidly revert back to either WT or a less obstructive engineered iteration - possibly explaining the altered amplicon sizes observed in Figure 4.14a. Furthermore, the template material that previously produced the off-target amplicons for g6.7 in Figure 4.12a may have become dominant once the pressure of the HR plasmid was removed. If this did occur, it would become enriched in the sample and subsequently account for the PCR amplifications observed in Figure 4.14a. However, there is no such explanation for the new g14 off-target amplicon (or the 3000bp g6.7 amplicon) observed above.

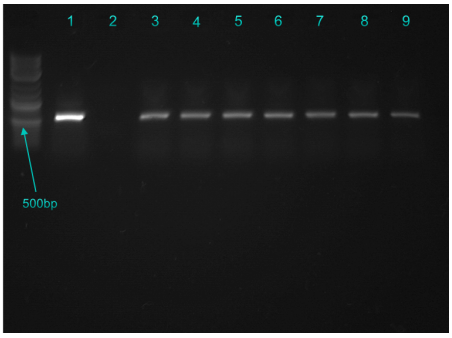
In an alternative attempt to explain the new off-target amplicons, it could be argued that during the EV36-wt propagation the ICP::NtNL sequence may be randomly altered via HR with similar regions of DNA, thereby producing a new sequence with a reduced evolutionary disadvantage compared to the original engineering construct. This new phage would be selected for and would propagate, thus, diminishing the original ICP engineered phage population. The premise for microbial genome HR-mediated "microevolution" has previously been presented [175], however, the likelihood of this or another type of genetic evolution (e.g. mutation based) occurring identically within independent plaques is low. Therefore, it cannot be considered a plausible conclusion at this stage.

Regardless of the cause or origin of the new off-target amplicons, it was decided to observe what would happened when repeating the plaque PCR experiments with ExRec-engineered phage propagated on EV36-pUC19-6.7/14::NtNL (i.e. the HR strains). The results for this further

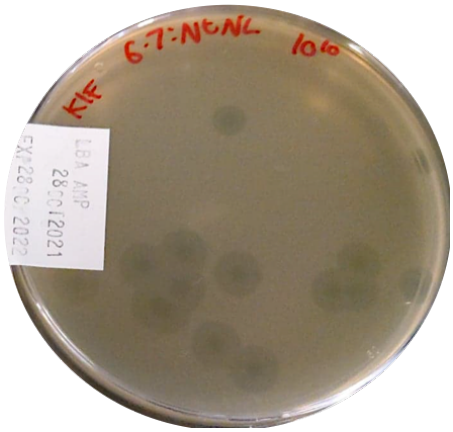
investigation are displayed in Figure 4.15. The plaque assay results revealed that a significantly higher titer was achieved for K1F*g6.7::NtNL* when propagated on the EV36 strain bearing the HR plasmid compared to EV36-wt - titers of 3.4×10^{10} and 1.2×10^{8} were achieved for K1F*g6.7::NtNL* and K1F*g14::NtNL* respectively (Figures 4.15c and 4.15d). One hypothesis for this is that, whilst the stock titer of engineered phage was identical for both plaque assays, the HR-bearing host provided viable conditions for the engineered phage population to propagate on, therefore allowing the engineered PFUs to remain viable and contribute to the titer.



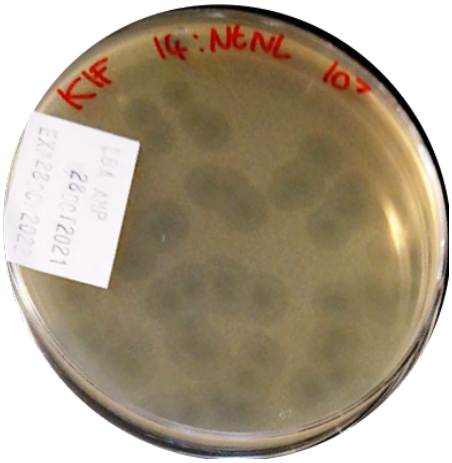
(a) ExRec-generated K1Fg6.7::NtNL - Plaque PCR with EV36-pUC19-g6.7::NtNL



(b) ExRec-generated K1Fg14::NtNL - Plaque PCR with EV36-pUC19-g14::NtNL



(c) Plaque assay of ExRec-generated engineered K1Fg6.7::NtNL on EV36-pUC19-g6.7::NtNL.



(d) Plaque assay of ExRec-generated engineered K1Fg14::NtNL on EV36-pUC19-g14::NtNL.

Figure 4.15: **Plaque assay and plaque PCR results for ExRec-generated engineered phage generated on the EV36-pUC19 strains.** PCR results obtained using the sequence identification strategy highlighted in 4.1 and 4.2 on **1a:** K1F-wt + EV36-wt w/ g6.7-FWD + g7.3-REV or **1b:** K1F-wt + EV36-wt w/ g14-FWD + g15-REV **(+ control)**, **2a:** K1F-wt + EV36-wt w/ NtNL-FWD + g7.3-REV or **2b:** K1F-wt + EV36-wt w/ NtNL-FWD g15-REV **(- control)**, **3a-9a:** K1Fg6.7::NtNL + EV36_pUC19-g6.7::NtNL w/ NtNL-FWD + g7.3-REV or **3b-9b:** K1Fg14::NtNL + EV36_pUC19-g14::NtNL w/ NtNL-FWD + g15-REV **(experimental results - 7 x individual plaques)**.

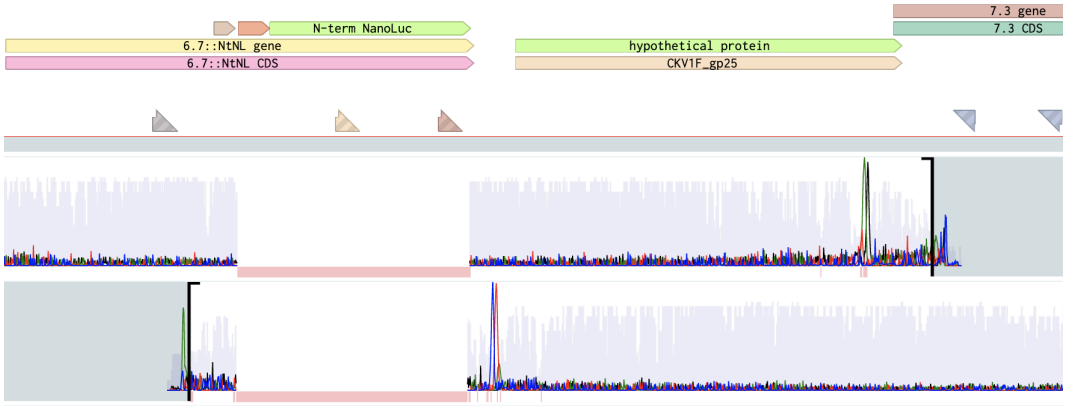
Whereas, previously with EV36-wt, the selection activity (e.g. microevolution or WT-reversion) that may have been occurring (as aforementioned), might have caused the PFU to drop. This increase was only observed for K1F*g6.7::NtNL*, potentially suggesting that the engineered population within this phage stock solution was higher in comparison to K1F*g14::NtNL*. Moreover, if the engineered population is higher in the stock solution, then there is a higher percentage of phage that may lose their viability when propagated on EV36-wt, which would explain the drop in PFU for K1F*g6.7::NtNL*.

Interestingly, almost 100% of the plaques that were screened for g6.7 and g14 generated correctly sized amplicons - indicating successfully isolated engineered phage populations (Figures 4.15a and 4.15b). As with the previous plaque PCR results, many more plaques were screened (data not shown) producing comparable results. Nonetheless, to confirm successful engineering it was now necessary to sequence-verify the inserted region and PCR-verify the increased amplicon size that should be generated from g6.7-FWD + g7.3-REV (564bp for WT and 786bp for K1F*g6.7::NtNL*) and g14-FWD + g15-REV (573bp for WT and 795bp for K1F*g14::NtNL*) as displayed in Figures 4.1 and 4.2.

4.4.8 Sequencing and suspected chimera

Genomic DNA was extracted from several plaque-isolated engineered phage samples and sent for sequencing with the WT primer pairs (g6.7-FWD + g7.3-REV for K1F*g6.7::NtNL* or g14-FWD + g15-REV for K1F*g14::NtNL*). All sequencing results gave comparable results, one of which is represented in Figure 4.16a for the g6.7 engineering attempt and Figure 4.16b for the g14 engineering attempt. All sequencing data successfully identified regions from the K1F-wt genome but *not* the engineered genomes for either K1F*g6.7::NtNL* or K1F*g14::NtNL*. Moreover, base pairs upstream and downstream of the NtNL insert were sequenced, however the NtNL fusion insert itself was not identified - represented by the red unsequenced regions for each missed base pair. This was also repeated with phage isolated from EV36-wt plaques with comparable results.

This finding was problematic due to the fact that each plaque represents a single, uniform phage population. Therefore, the sequencing data strongly suggests that each of the plaques comprised phage K1F-wt only, even though prior PCR screening indicated engineering success (i.e. in Figures 4.12, 4.13 and 4.15). This signifies the first suggestion of a potential artificial PCR product or some degree of contamination leading to false positives. Additional PCR analysis was carried out to investigate this further (Figure 4.17). However, this time, the purified phage genome was used as the template instead of the boiled phage lysate to avoid any background noise from the lysate content (e.g. plasmids, *E. coli* DNA, etc.). The Phenol:Chloroform method was used to



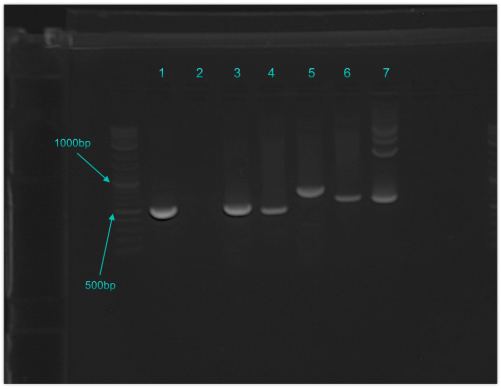
(a) K1F*g6.7::NtNL* failed sequencing attempt

(b) K1F*g14::NtNL* failed sequencing attempt

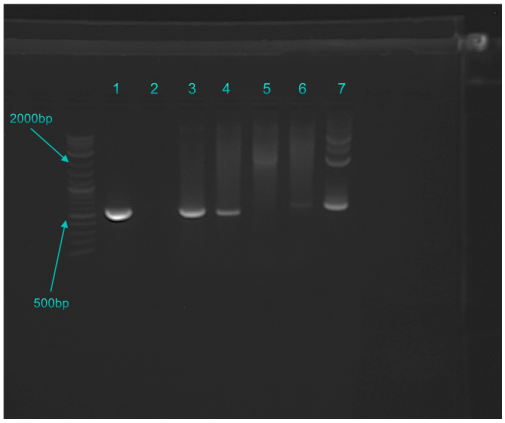
Figure 4.16: **K1F engineering failed sequencing attempts.** Screenshot from the Benchling sequence alignment tool for **a:** spanning g6.7, the NtNL fusion insert and the upstream hypothetical gene. The expected K1Fg6.7::NtNL sequence is aligned with the actual sequencing results generated from g6.7-FWD + g7.3-REV and **b:** spanning g14, the NtNL fusion insert and the upstream g15. The expected K1Fg14::NtNL sequence is aligned with the actual sequencing results generated from g14-FWD + g15-REV. Successfully sequenced base pairs are shown as multicoloured frequency lines whereas unsuccessfully sequenced base pairs are highlighted in red.

extract pure, high concentration genomic DNA from phage samples that were propagated from ICP::NtNL amplicon-generating phage (i.e. phage that generated the engineering amplicon and were thought to be successfully engineered).

Prior to running this gel, is was suggested that potentially an artificial PCR product was being detected, thereby producing a false positive result. It is conceivable that PCR-amplified fragments from the intracellular expression or HR plasmids (which NtNL-FWD can bind to) and K1F-wt genome fragments (which the REV primers can bind to) were contaminating the PCR reaction and creating chimeras. This is why it was decided to use the pure phage genome as the PCR template rather than the boiled lysate. Following on from isolating the engineered phage genomes using the Phenol:Chloroform method, initial PCR analysis using NtNL-FWD and g7.3/15-REV was quite promising for both phage - as displayed in well 4 in Figures 4.17a



(a) ExRec-generated K1Fg6.7::NtNL - final screening



(b) ExRec-generated K1Fg14::NtNL - final screening

Figure 4.17: **Final PCR screening analysis to conclusively determine engineering success.** PCR results on **1a:** K1F-wt + EV36-wt w/ g6.7-FWD + g7.3-REV or **1b:** K1F-wt + EV36-wt w/ g14-FWD + g15-REV, **2a:** K1F-wt + EV36-wt w/ NtNL-FWD + g7.3-REV or **2b:** K1F-wt + EV36-wt w/ NtNL-FWD g15-REV, **3a:** K1Fg6.7::NtNL DNA w/ g6.7-FWD + g7.3-REV or **3b:** K1Fg14::NtNL DNA w/ g14-FWD + g15-REV, **4a:** K1Fg6.7::NtNL DNA w/ NtNL-FWD + g7.3-REV or **4b:** K1Fg14::NtNL DNA w/ NtNL-FWD + g15-REV, **5a:** K1Fg6.7::NtNL DNA w/ NtNL-FWD + g8-REV or **5b:** K1Fg14::NtNL DNA w/ NtNL-FWD + g16-REV, **6a:** K1Fg6.7::NtNL DNA w/ amp-FWD + amp-REV or **6b:** K1Fg14::NtNL DNA w/ amp-FWD + amp-REV, **7a:** pUC19-g6.7::NtNL w/ amp-FWD + amp-REV or **7b:** pUC19-g14::NtNL w/ amp-FWD + amp-REV.

and 4.17b - suggesting the engineered fragment was present in the absence of any external phage lysate contaminates. The sample used to generate the fragment shown in well 5 was identical to well 4, the only difference was that an alternative REV primer was used to extend the amplicon further. This was done in an attempt to more persuasively verify correct NtNL insertion into the genome by exhibiting an uninterrupted large DNA fragment consisting of the engineered insert and upstream WT material. For K1F*g6.7::NtNL*, g8-REV was used and for K1F*g14::NtNL*, g16-REV was used - producing an 800bp and 2000bp amplicon respectively. The new primer pairs both produced correctly sized amplicons from the purified DNA - again indicating engineering success.

Given the promising results generated from the pure genomes (wells 4 and 5), it was concerning that, still, it was not possible to successfully sequence verify the insert using a WT upstream FWD primer (comparable results to Figure 4.16 were repeatedly observed) or generate a larger PCR fragment using a WT upstream FWD primer (the well 3 fragments in Figures 4.17a and 4.17b show comparable fragments to wells 1 and 4, whereas if successful engineering had occurred it would generate a fragment that is approximately 200bp larger, as explained in Figures 4.1 and 4.2).

Due to these conflicting observations, it was decided to test if the HR plasmid was in fact present in the pure phage genome stocks and therefore potentially explaining the persistence of PCR chimera formation (i.e. false positive results in wells 4 and 5). Well 7 is a control showing the amp-FWD and amp-REV primers correctly binding to the ampicillin gene region of the pUC19 plasmid, generating approximately a 600bp amplicon that is not present in any of the phage genomes. Unexpectedly, this ampicillin amplicon was also generated from the "engineered phage genome only" sample in well 6 for both g6.7 and g14. Hypothetically, this may suggest that the pUC19 plasmid had persisted to remain in the stocks regardless of the stringent purification steps carried out during Phenol:Chloroform phage genome extraction, thereby allowing for PCR chimeras to be generated from the K1F-wt genome and pUC19 DNA.

In PCR, the most common mechanism of chimera formation is incomplete primer extension during the extension phase. Moreover, during subsequent PCR cycles, partially extended (aborted) products can act as primers and bind to heterologous templates, which can then be extended to form artificial PCR products (i.e. chimeras) [176]. This is potentially an explanation for what has occurred here. Furthermore, repeated genomic extractions, PCR amplifications and gel electrophoresis analysis confirmed the presence of the ampicillin resistance gene in the so-called engineered phage DNA stocks - adding increased likelihood that it was in fact false positives occurring rather than successful engineering. Figure 4.18 exhibits a potential mechanism for the formation of chimeras during this work, using K1F g6.7 as an example.

It is hypothesised that during PCR amplification of the phage lysate or contaminated genome stocks, a long, incomplete PCR fragment is generated and aborted via the NtNL-FWD primer binding to the pUC19 plasmid. This aborted fragment could possibly consist of part of the NtNL gene and up to 150 base pairs of the upstream homology arm within the HR plasmid. This fragment then has the potential to bind to the K1F-wt genome (acting as a new FWD primer) with its regions of homology (i.e. the 3' end of the fragment, leaving the NtNL 5' end of the fragment unattached but still present). This new FWD primer can then pair up with the existing REV primer (g7.3-REV in this example) to produce an artificial, fully formed fragment of DNA mimicking the desired engineered phage amplicon which can subsequently continue to be amplified by the original primer pair.

Furthermore, in an attempt to find out what exactly had been amplified during the pseudo-successful ExRec engineering attempts, the correctly sized engineering amplicons seen in Figure 4.15 were gel extracted, purified and sent for sequencing with the NtNL engineering primer pairs. These sequencing results are displayed in Figure 4.19.

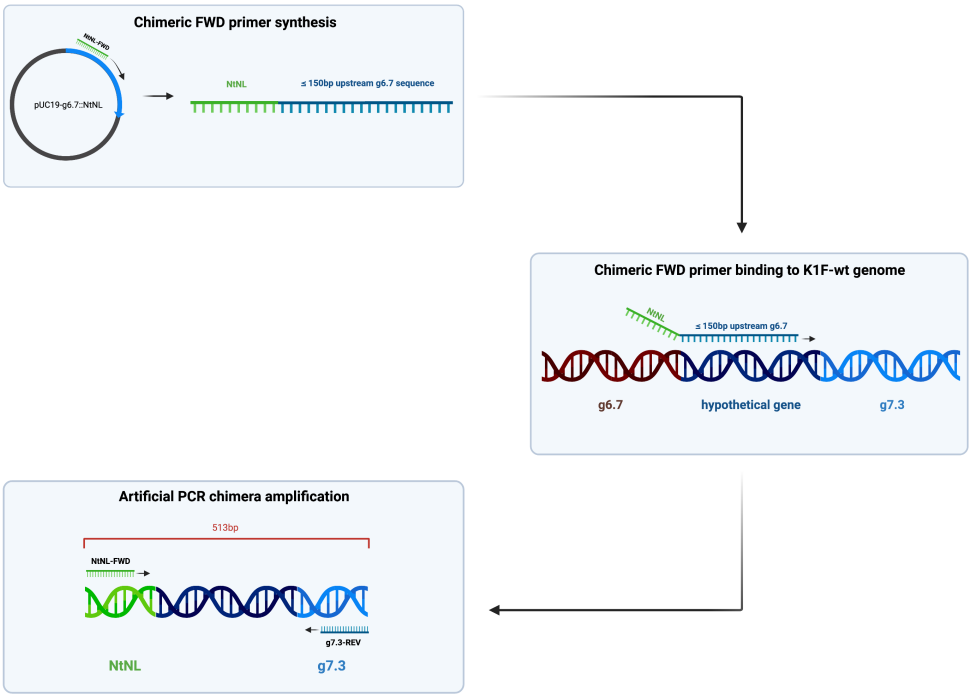
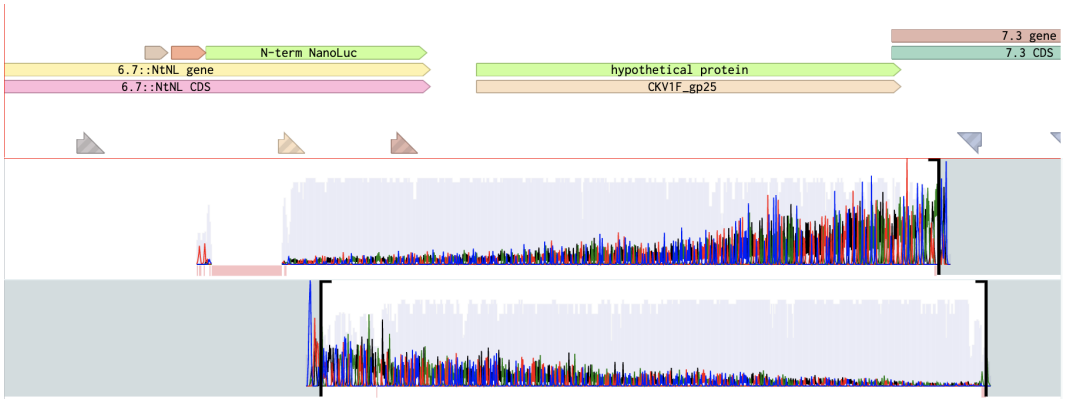


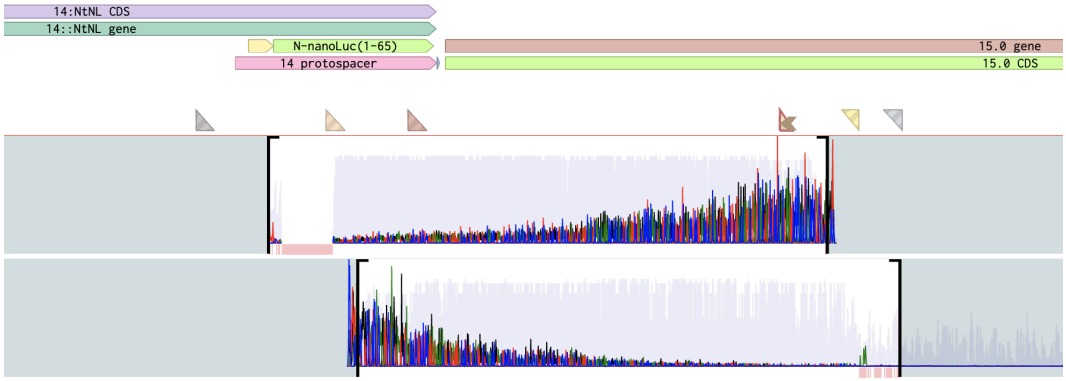
Figure 4.18: **PCR chimera schema.** False positives may be generated by artificial chimera fragments. The potential mechanism of this occurring is outlined using the K1F g6.7 engineering process as an example.

As shown in Figure 4.19, DNA fragments for both g6.7 and g14 spanning the NtNL fusion and 500bp upstream were correctly sequenced. Moreover, the fully-sequenced amplicons were of the correct configuration that would be expected from the engineered genomes. Whilst the NtNL-FWD primer can bind to the pUC19 HR plasmids, these plasmids only have a 150bp upstream homology arm, therefore the 500bp upstream REV primers cannot bind or instigate amplicon generation. Furthermore, whilst the REV primers do bind to the K1F-wt genome, the NtNL sequence is not present. Therefore, the first obvious conclusion to make would be that the engineered genomes were present and had been amplified. However, considering the data presented in Figure 4.17 and the hypothesis displayed in Figure 4.18, it is, unfortunately, more likely that chimeric formation during PCR is the source of these amplifications.

At this stage, when taking into account all of the data collected, it was highly indicative that, although the ExRec protocol did cause a shift in results, the genetic engineering experiments had not produced the desired engineered phage. Nevertheless, before moving on, one last investigation was carried out - ExRec followed by three rounds of CRISPR propagation. Perhaps, if a small amount of recombination was occurring during the ExRec protocol then CRISPR selec-



(a) K1F*g6.7::NtNL* amplicon sequencing with NtNL-FWD and g7.3-REV



(b) K1F*g14::NtNL* amplicon sequencing with NtNL-FWD and g15-REV

Figure 4.19: **Sequencing the suspected PCR chimera amplicons.** Screenshots from the Benchling sequence alignment tool for **a:** fragment spanning g6.7, the NtNL fusion insert, the upstream hypothetical gene and g7.3. The expected K1Fg6.7::NtNL sequence is aligned with the actual sequences results generated from NtNL-FWD + g7.3-REV, or **b:** fragment spanning g14, the NtNL fusion insert and the upstream g15. The expected K1Fg14::NtNL sequence is aligned with the actual sequences results generated from NtNL-FWD + g15-REV. Successfully sequenced base pairs are shown as multicoloured frequency lines whereas unsuccessfully sequenced base pairs are highlighted in red.

tion would enrich this population. The agarose gel electrophoresis PCR screening results of this final genetic engineering attempt, which was only attempted for g6.7 due to time constraints, are displayed in Figure 4.20.

As shown on the following page, the amplicon signal generated from the ExRec sample (well 4) gradually decreases over the course of the first two CRISPR propagations (wells 5 and 6) and isn't detectable after the final CRISPR propagation (well 7). This is the opposite of what should occur during CRISPR selection, where the amplicon band size should increase with enrichment, suggesting that, indeed, chimeric formation is the explanation for false positive results seen in

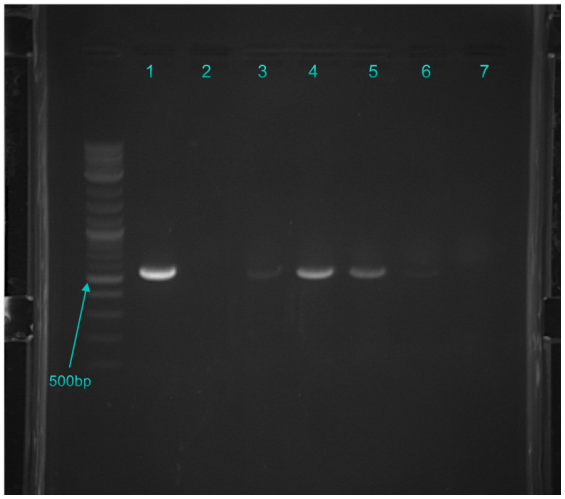


Figure 4.20: **Agarose gel electrophoresis PCR screening of ExRec followed by three rounds of CRISPR.** PCR results obtained using the sequence identification strategy highlighted in 4.1 and 4.2 on **1:** K1F-wt + EV36-wt w/ g6.7-FWD + g7.3-REV, **2:** K1F-wt + EV36-wt w/ NtNL-FWD + g7.3-REV, **3:** K1F-wt + 1 x round of ExRec w/ g6.7-FWD + g7.3-REV, **4:** K1F-wt + 1 x round of ExRec w/ NtNL-FWD + g7.3-REV, **5** previous sample + CRISPR round 1 w/ NtNL-FWD + g7.3-REV, **6** previous sample + CRISPR round 2 w/ NtNL-FWD + g7.3-REV, **7** previous sample + CRISPR round 3 w/ NtNL-FWD + g7.3-REV.

wells 4, 5 and 6. If this is the case and in line with the hypothesis presented in Figure 4.18, then the pUC19 HR plasmid present in the ExRec lysate must be passed on from lysate-to-lysate during the initial CRISPR propagations until its presence is diminished and undetectable by the time the third CRISPR propagation is completed. Well 3 (ExRec-processed phage + the WT FWD primer instead of NtNL-FWD) contributes to confirming the lack of successful engineering by displaying a comparably sized amplicon to well 1 (an amplicon generated from K1F-wt with the same primers used for the well 3 sample), whereas (as previously stated and explained in Figures 4.1 and 4.2), if successful phage engineering had occurred, this amplicon would be approximately 200bp larger due to the NtNL fusion increasing the size of the gene.

4.4.9 Mass spectrometry analysis

Although it was now assumed that it was highly likely that false-positive amplicons were being generated, as a final means of confirming the outcome of the genomic engineering experiments, it was decided to analyse the engineering amplicon-generating samples via mass spectrometry (MS). In doing so, if the unlikely scenario that the engineering had successfully worked and therefore the ICP::NtNL fusion packaged phage were present, it would be possible to detect the fusion proteins with the mass spectrometer.

In a streamlined protocol optimised for phage protein analysis via MS (which is provided in

this chapter's Methods section), K1F-wt alongside quasi-K1F*g6.7::NtNL* and -K1F*g14::NtNL* samples were resuspended in an MS-friendly solution (ammonium bicarbonate), then further prepared and digested with trypsin prior to analysis. Trypsin is a protease widely used to process proteins into more readable sets of peptides for MS analysis due to its reliable and consistent cleavage of lysine and arginine residues [177]. Displayed in Figure 4.21 are the proteins that were screened via MS, along with their amino acid sequences and percentage of identified residues. The successfully identified residues are highlighted in green.

The MS experts from the University of Warwick Proteomics department have advised that approximately 60% and above amino acid coverage is very good and 2 or more unique peptides identified is sufficient for confirming the presence of a protein. Agreeing statements and data can also be found in the literature [178, 179]. As displayed above, the K1F-wt proteins (gp6.7 and

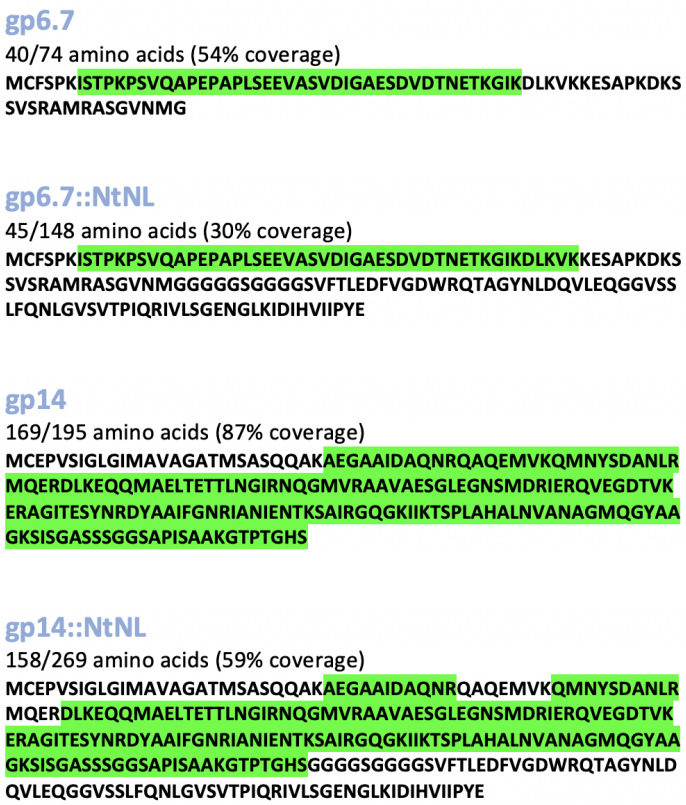


Figure 4.21: **Mass spectrometry analysis of the K1F genome engineering samples.** The full amino acid for each protein is provided, with the successfully identified residues highlighted in green. For the two WT proteins (gp6.7 and gp14) a K1F-wt sample was screened, for the gp6.7::NtNL protein an engineering amplicon-generating quasi-K1F*g6.7::NtNL* sample was screened and for the gp14::NtNL protein an engineering amplicon-generating quasi-K1F*g14::NtNL* sample was screened.

gp14) were identified to a satisfactory degree (albeit with gp6.7 slightly below 60% coverage). Furthermore, 2 and 33 unique peptides were identified for gp6.7 and gp14 respectively, confirming their identification. Whilst, on paper, the fusion proteins did yield a considerable percentage of coverage each, this coverage was entirely on the N-terminal end of both proteins (i.e. the WT domains). As seen in Figure 4.21, no amino acid residues were identified from the (GGGGS)2 linker onward and neither were any unique peptides found for the linker or NtNL domain. This confirms that the exogenous C-terminal NtNL domain was not present in any of the samples.

In conclusion, the MS data clearly verifies that no fusion proteins had been packaged inside the phage capsids and therefore, no successful genetic engineering had occurred. It does, however, provide a blueprint method for rapid protein screening if the genetic engineering was successfully completed in the future. Therefore, it was not a completely futile exercise.

4.4.10 Non-genomic *in vivo* ICP phage engineering

After almost certainly establishing that the genetic engineering experiments had failed to produce neither K1F*g6.7::NtNL* nor K1F*g14::NtNL*, it was decided that a fresh approach was necessary. Furthermore, leaning on some of the ideas and (albeit counterfeit) results generated from the ExRec experiments (Figure 4.11), a non-genomic engineering strategy was devised. This rapid and simple non-genomic phage engineering strategy, displayed in Figure 4.22 (a more detailed protocol is provided in this chapter's Methods section), was created in an attempt to generate structurally engineered K1F*gp6.7::NtNL* and K1F*gp14::NtNL* at the protein level rather than the genomic level - note that 'g6.7/14' has been changed to 'gp6.7/14' to represent the non-genomic, transient approach. Briefly, K1F-wt is propagated on host cells that constitutively express the relevant ICP::NtNL fusion protein. It is hypothesised that, upon the inception of intracellular phage assembly, some phage particles will include the fusion protein within their structure and therefore become an ICP::NtNL engineered phage. A 100 kDa Amicon® Ultra filtration unit (Merck, New Jersey, USA) is then used in an attempt to separate the <100 kDa fusion protein present in the lysate from the >100 kDa mixed WT/engineered phage population. If successful, this method could be used to manufacture single-use, host-dependent diagnostic phage (or any other engineered phage generated using the method) at industrial scale. The only limiting factors are, firstly, the size of culture that the user is capable of processing and secondly, the size of the filtration unit they can source. Moreover, if a bespoke solution was built then this method could be used to produce a vast amount of single-use, engineered phage.

During attempting this new phage engineering method, aliquots from several stages of the process were taken and measured for luminescence via the Nano-Glo® assay. This data is shown in Figure 4.23. The data shown in Figures 4.23a and 4.23b are largely comparable to one an-

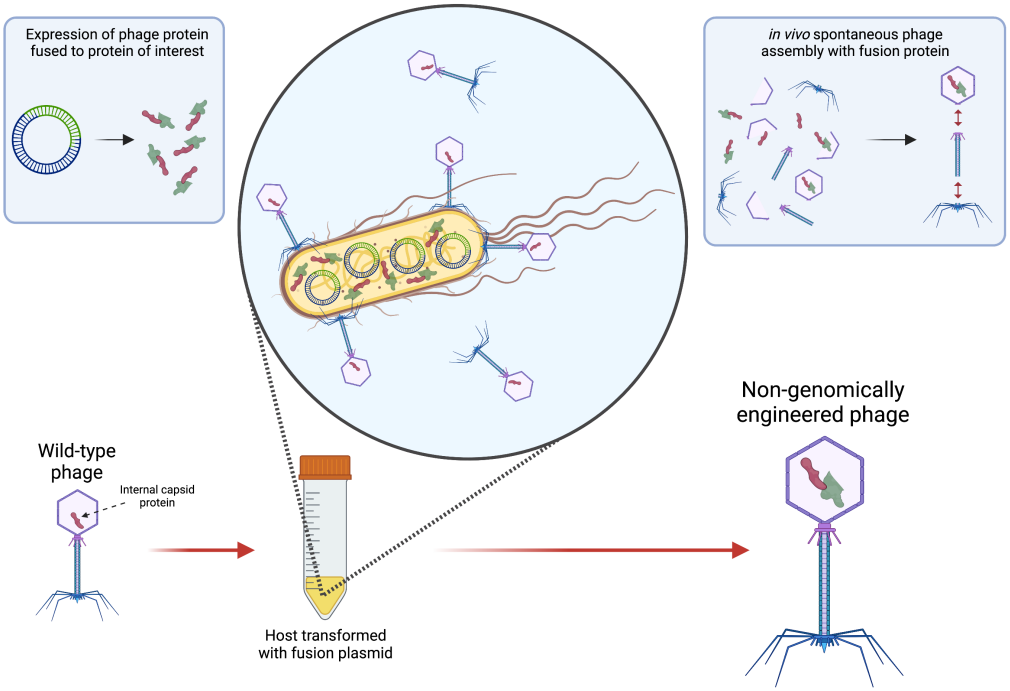
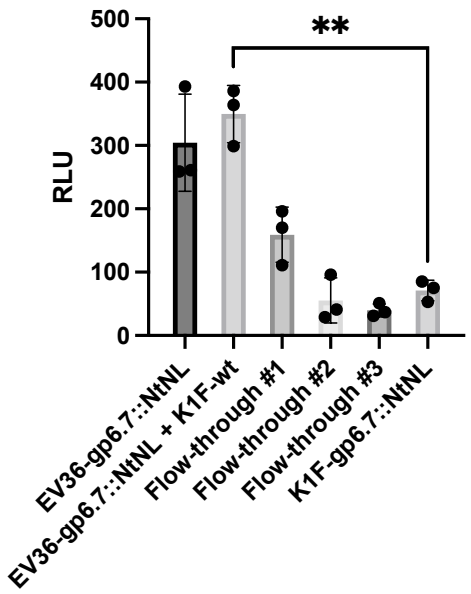
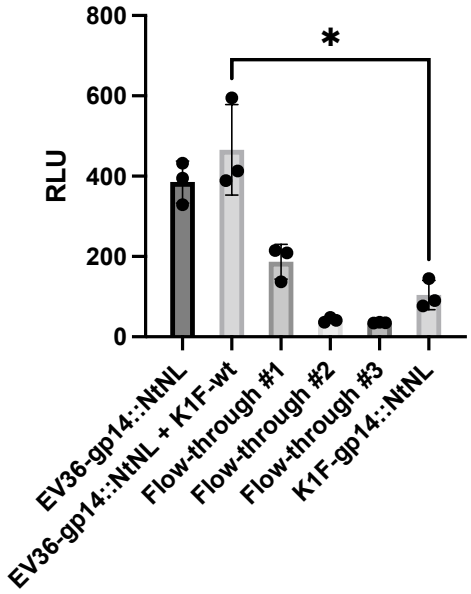


Figure 4.22: **Non-genomic *in vivo* ICP phage engineering strategy.** It is hypothesised that this method can be used to package the ICP::NtNL fusion into the phage capsid by propagating K1F-wt on an EV36 strain constitutively expressing the fusion protein.

other. For both, whilst the difference between the first and second data-points are not significant, the observable rise in luminescence after K1F-wt has been added and subsequently cleared the EV36-ICP::NtNL culture could be due to the enhanced lysis that the phage provides alongside the lysis buffer, therefore releasing more endogenous ICP::NtNL protein, compared to the lysis buffer only for the first data-point. Data-points 3, 4 and 5 are the RLU readings after each filtration step. The first filtration flow-through maintained approximately half of the signal from data-point 2 and the next two flow-through samples emitted only small amounts luminescence near the baseline value. However, contrary to the desire for the 15.5 kDa (gp6.7::NtNL) and 28.2 kDa (gp14::NtNL) fusion proteins to entirely pass through the 100 kDa filtration unit, it would appear that a portion of the protein was in fact retained alongside the phage. Furthermore, the sixth data-point (i.e. the filtered engineered phage) exhibits an increase in luminescence above the baseline level observed in the second and third flow-through samples. It was therefore essential that this background signal was taken into account during the subsequent testing of the non-genomically engineered phage.



(a) K1F*gp6.7::NtNL* non-genomic in vivo phage engineering workflow data-points

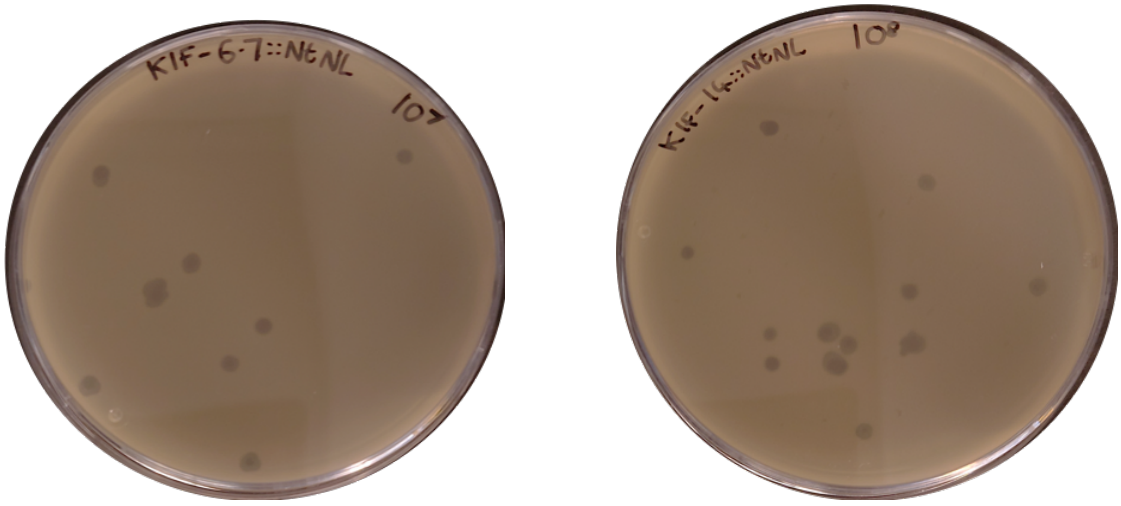


(b) K1F*gp14::NtNL* non-genomic in vivo phage engineering workflow data-points

Figure 4.23: **Non-genomic *in vivo* engineering workflow data-points.** Luminescence was measured on aliquots taken at several stages during the **a:** K1Fgp6.7::NtNL and **b:** K1Fgp14::NtNL non-genomic in vivo engineering process. Flow-through samples were taken after passing the phage through a 100 kDa Amicon® Ultra filtration device three times, using a new device each time. All samples were treated with lysis buffer and combined with an equal volume of EV36-CtNL lysate, then left for 1 hour to allow for spontaneous CtNL::NtNL complementation prior to the Nano-Glo® assay, (+/- SD, n = 3).

One possible reason for only half of the signal from data-point 2 being passed through the filtration unit and being present in data-point 3 is the blocking or fouling of the filtration membrane. Furthermore, the interactions between proteins and filtration membranes and how this impacts the filtration performance has been well documented [180, 181], with the general consensus being the more used a filtration membrane is, the more increased the flux decline is. This was the core reason why three different, brand new filtration units were used for each filtration step during the phage engineering workflow. However, this did not appear to resolve the issue, as the second and third flow-through samples did not produce a substantial luminescent output, suggesting that only a negligible amount of unpackaged fusion was available to pass through the new filtration membranes. Moreover, the rise in RLU for data-point 6 suggests that any unfiltered signal may have aggregated with the phage particles, thereby preventing their filtration and contributing towards the filtered phage RLU output as background noise.

Plaque assays were also carried out on the engineered phage after they had been filtered in order to confirm the presence of phage and quantify their concentration (Figure 4.24). The PFU



(a) Non-genomic in vivo engineered K1F*gp6.7::NtNL* phage on EV36-wt

(b) Non-genomic in vivo engineered K1F*gp14::NtNL* phage on EV36-wt

Figure 4.24: **Plaque assays of non-genomic *in vivo* engineered phage on EV36-wt.**

calculation was applied to estimate the phage titers and this revealed an engineered phage stock of 2.5×10^7 for K1F*gp6.7::NtNL* and 3.4×10^8 for K1F*gp14::NtNL*. This confirmed that a large proportion of the phage particles present in the non-genomic *in vivo* engineering lysate were able to be retained throughout the filtration process.

4.4.11 Heat-induced signal release by *in vivo* non-genomically engineered phage

After carefully sifting through the literature and finding relevant data concerning the expulsion of T7/K1F genomic DNA and other internal capsid contents, a quick acid test to verify successful signal packaging was devised. This was based on the findings that, upon heating to temperatures between 65°C - 80°C, it has recently been shown that phage T7 prematurely ejects its genomic DNA [182, 183]. As the ejection of the ICPs precede DNA ejection [121], it is also hypothesised here that the phage ICPs are prematurely expelled at high temperatures too. Therefore, if successful ICP::NtNL fusion packaging had taken place, an increased luminescent signal should be detected at high temperatures when the internal capsid contents (containing ICP::NtNL) are heat-ejected out of the phage. However, before testing this, it was first important to observe what impact heating to different temperatures had on the ICP::NtNL fusions' ability to successfully complement with CtNL and emit light when provided with substrate.

The data in Figure 4.25 reveal that, at higher temperatures (45 °C - 90 °C), the activity of ICP::NtNL is significantly diminished. Furthermore, the luminescent readout gradually decreases as the temperature is increased from 45 °C to 90 °C. This could be attributed to the high

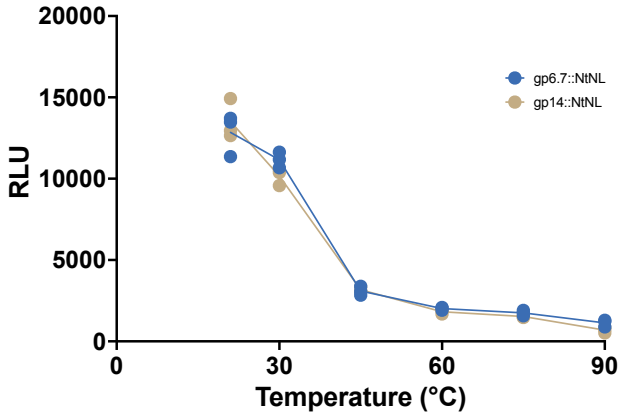
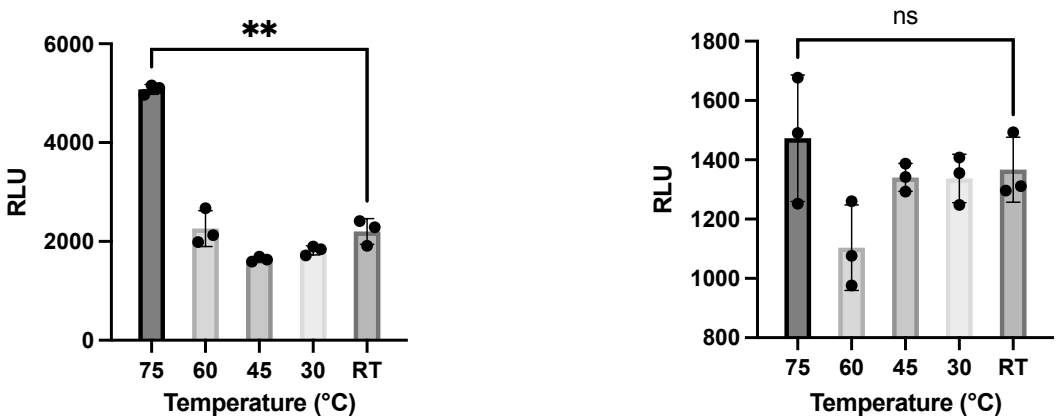


Figure 4.25: **ICP::NtNL fusion thermal stability.** TXTL-expressed fusion proteins (gp6.7::NtNL - blue or gp14::NtNL - gold) were exposed to different temperatures (RT, 30 °C, 45 °C, 60 °C, 75 °C and 90 °C) on a heat block, then let to return to RT. The samples were then combined with an equal volume of TXTL-expressed CtNL and let to incubate for 1 hour before carrying out the Nano-Glo® assay to measure their luminescent activity.

temperature-induced denaturation causing a decrease in the fusions' ability to either 1). complement with CtNL or 2). catalyse the furimazine to furimamide reaction once it has complimented with CtNL to form the full NanoLuc enzyme, or 3). a combination of the two. The RLU values are similar at RT and 30 °C, suggesting negligible protein denaturation had occurred, which is to be expected. Data for gp6.7::NtNL and gp14::NtNL are comparable. Interestingly, previously published data suggest that the melting temperature of the full NanoLuc enzyme is 58 °C [125]. Moreover, the data presented in Figure 4.25 would suggest that, when the enzyme is split into two sub-units, its melting temperature is reduced to approximately 45 °C. This is suggested by the drastic reduction in RLU output from 30 °C to 45 °C, eluding to the enzyme activity and thus, its composition. It, perhaps, is not surprising though that the thermostability of NanoLuc is reduced when split into sub-units, as the optimal stability is to be expected when the fully formed and folded protein composition is displayed.

The ICP::NtNL thermal stability data from Figure 4.25 must be considered when analysing the heat-induced ejection results displayed in Figure 4.26. Moreover, the ICP::NtNL heat ejection hypothesis was tested for the non-genomically engineered phage and the data collected are shown in Figure 4.26. If the non-genomic engineering experiments had failed and the only ICP::NtNL fusion content present in each sample was unpackaged background (i.e. external to the phage capsids) then it would be expected that at higher temperatures (45 °C - 90 °C), the signal generated would be lower than at 30 °C and RT, due to the high temperatures denaturing the fusion protein (as displayed in Figure 4.25). However, the results in both Figures 4.26a and 4.26b show that when the engineered phage were heated to 75 °C, an increased signal com-



(a) K1F*gp6.7::NtNL* heat-induced signal release

(b) K1F*gp14::NtNL* heat-induced signal release

Figure 4.26: **Luminescence measurements after heating non-genomically *in vivo* engineered phage to different temperatures.** All samples were exposed to the desired temperature, then allowed to return to room temperature, then combined with an equal volume of CtNL TXTL reaction and left to incubate for 1 hour at room temperature to allow for spontaneous CtNL::NtNL complementation prior to the Nano-Glo® assay, (+/- SD, n = 3).

pared to the RT sample is observed - this increase is highly significant for K1F*gp6.7::NtNL* but although the increase is observable for K1F*gp14::NtNL*, it is not significant.

The RT samples for both phage represent the background (unpackaged) ICP::NtNL and therefore should be viewed as baseline values. For K1F*gp6.7::NtNL*, the signal observed in the 30 °C and 45 °C samples slightly drop below the baseline RT value, suggesting some denaturing of the gp6.7::NtNL fusion. At 60 °C, despite the significant denaturing that has been shown to occur (Figure 4.25, the RLU value rises back to the baseline value, suggesting a possible combination of denaturation and heat-ejection other samples are comparable. The 75 °C sample should represent the lowest RLU value, however the significant increase in signal suggests a large amount of encapsulated fusion has been released.

For K1F*gp14::NtNL*, the difference in values between samples are not significant, however some conclusions can be drawn. The 60 °C sample clearly generated the least signal, suggesting fusion denaturation, however at 75 °C, the signal is vastly increased - again, contrary to what *should* happen if successful phage engineering was not achieved (i.e. only background signal was present). When considering that high temperatures negatively impact ICP::NtNL activity, the data shown in Figure 4.26 are highly indicative that the non-genomic engineering process *is* capable of successfully packaging ICP::NtNL fusion into the K1F phage capsid.

4.4.12 Conclusions drawn and outlook for further investigations

Ultimately, the genetic engineering experiments attempted over the course of this chapter are likely to have failed, however, they did encourage innovative thought processes which eventually lead to the non-genomic *in vivo* phage engineering protocol and subsequently, the key research aims were somewhat inadvertently achieved. Moreover, the research aims for this chapter were to, firstly, engineer K1F so that it incorporates the ICP::NtNL fusion proteins within its internal capsid structure and secondly, to demonstrate that the fusion can be inducibly released to generate on-demand bioluminescent signal. The data shown in Figure 4.26 strongly suggest that both of those research aims have been achieved.

If the genetic engineering experiments were to be reattempted, it is suggested that a strong emphasis should be placed on optimising the crRNA designs so that WT phage propagation is inhibited. Also, it is worth experimenting with the λ red system in order to enhance the intracellular environment for HR. In a recent advancement within the CRISPR phage engineering field, multiple crRNA designs were screened for optimal activity, then their performance was measured in Efficiency of Plating (EoP), whereby the crRNA causing the lowest EOP was selected as the optimally performing design [170]. EoP is essentially a measurement for how well the crRNA inhibits phage propagation. Even though, as previously mentioned, the CRISPR system that successfully selected for K1F*g10b::GFP* did not inhibit K1F-wt propagation [94], it would be preferable if a crRNA design could be optimised to efficiently cut K1F-wt and prevent propagation so that it can be confirmed that the system is working prior to attempting the genetic engineering experiments. Furthermore, if the CRISPR system more efficiently digests WT DNA, then it is likely that a more stringent selection process would be commissioned, as seen in the literature whereby optimising the crRNA design leads to engineering success rates as high as >99% [170]. Moreover, if the HR and CRISPR processes, armed with an optimised crRNA design, were combined within the λ red system, as has previously been demonstrated for editing *E. coli* chromosomal DNA [184], then truly optimal results could be achieved.

Whilst the non-genomic *in vivo* phage engineering method did arise from the quasi-promising ExRec strategy, the rationale behind its core mechanism - complimenting intracellular phage assembly with plasmid provided phage proteins - is well grounded. Furthermore, it has previously been demonstrated that when a gene of interest is knocked out of the T7 genome, the physical translation of that gene can be provided via intracellular plasmid expression, which can subsequently compliment assembly and "rescue" the mutant phage [97]. In this referenced publication, gp6.7 and gp14 are both provided intracellularly when attempting to propagate T7-Δgp6.7 and T7-Δgp14 knock-out mutants. It has also been shown that by providing the T7 gp17 (tail fibers) to tail-less phage, "activation" of the previously unviable particles occurs whereby they assemble

with the provided tail protein [185]. Whilst there is no mention in either of these publications of the plasmid-mediated phage assembly rescue/activation method being utilised as an engineering method, essentially the same methodology to this is displayed in the non-genomic *in vivo* phage engineering method (Figure 4.22), with the only difference being that an exogenous fusion protein is supplied rather than a WT protein. The key contribution from this chapter to the aforementioned prior art is that these seemingly unassuming phage rescue demonstrations have been transformed into a rapid fusion-based engineering method to that can, not only be used to alter the phage structure, but also package proteins inside the phage capsid using the provided ICP fusion technology. Furthermore, in the future, this could be utilised for an array of innovative applications (e.g. for therapeutic or diagnostic purposes) that could benefit from the production of single-use engineered phage.

Chapter 5

Exploring cell-free TXTL as a K1F capsid packaging method and establishing the *E. coli* K1 phage-based diagnostic assay

5.1 Abstract

Given the serious threat posed by antimicrobial resistance, the development of a truly rapid and reliable bacterial detection system has been one of the great ambitions of next-generation phage biology. Moreover, whereas the results derived by other researchers have produced diagnostic phage that can detect their host pathogen in a matter of hours, the work presented here focuses on attempting to construct a system which can achieve successful detection within minutes. This chapter's results include the development of a diagnostic test utilising the engineered phage developed during the previous two chapters, the establishment and use of cell-free K1F synthesis to optimise the non-genomic phage engineering process and experimenting with an amalgamation of cell-free and electron microscopy to visually analyse phage assembly. Furthermore, it is envisaged that the engineering strategy developed here could be replicated with a plethora of different phage-types, thereby paving the way for a multi-use detection tool that could potentially screen dozens of pathogenic bacteria at once within a few minutes. The implications of such a device are huge and would play a key role in tackling antimicrobial resistance by allowing specific therapies to be rapidly administered on a case-by-case basis, rather than resorting to the widespread use of antibiotics.

5.2 Introduction

The increasing incidence of bacterial infections worldwide, alongside the often uninformed overuse of antibiotics, have harboured an urgent need for rapid pathogen detection. A fast and specific diagnosis would allow for early, targeted therapy to be administered and for antibiotics to be spared for necessary cases only, subsequently leading to a decrease in medical burden and AMR [22].

5.2.1 Existing *E. coli* bacterial detection systems

The traditional culture-based method has been the gold standard of bacterial identification for decades. However, due to the time-consuming (>24 hours) and laborious nature of bacterial culturing, many alternative detection methods have been established and continue to be developed. Existing alternative methodologies/technologies for bacterial detection include:

- **phage-based** - high specificity phage (typically bearing a reporter gene within a non-structural region of their genomic DNA) are utilised to infect and propagate within the target bacteria. Subsequently, a signal is generated and/or released [186]
- **enzyme linked immunosorbent assay (ELISA)** - utilises high-affinity antibody-enzyme complexes which, when attached to specific bacterial antigens and provided with substrate, react and release a signal [187]
- **gold nanoparticle (AuNP) aggregation** - utilises the principle of surface plasmon resonance (i.e. the light-stimulated collective oscillation of conduction electrons at a polarised interface) to detect changes in nanoparticle aggregation states [188]. AuNPs are coupled with highly specific detectors (e.g. phage, antibodies or nucleic acids) in order to achieve specificity and initiate bacterial-mediated aggregation [189]
- **polymerase chain reaction (PCR)** - specific primer pairs bind are deployed to target regions of bacterial DNA. Amplification/detection is initiated via thermal cycling and the activity of DNA polymerase [190].
- **nucleic acid sequence-based amplification (NASBA)** - high sensitivity method utilising RNA polymerase for the amplification of bacterial mRNA coding for a gene of interest. The starting material and end product of this method is always RNA, therefore there is no interference of background DNA [191]
- **loop-mediated isothermal amplification (LAMP)** - utilises DNA polymerase and multiple primers which bind to different distinct regions of the bacterial target DNA, initiating the formation of a hairpin loop structure. Subsequently, amplification is carried out via

strand displacement DNA synthesis cycles which do not necessitate thermal cycling. The final products are a mixed collection of hairpin loop DNA molecules with various lengths and structures [192]

- **recombinase polymerase amplification (RPA)** - the rapid isothermal amplification process is mediated by: 1). recombinases which bind primer pairs to bacterial target dsDNA, 2). single-stranded DNA-binding proteins which protect the primers prior to extension, 3). strand-displacing DNA polymerase which carries out DNA synthesis, facilitating exponential amplification which results in a similar product to PCR [193]

For all nucleic acid-based detection methods, amplification can be visualised via conjugating the product with a probe (e.g. biotin or FITC) and combining this with another method such as an ELISA. Furthermore, a comparative display of recent *E. coli* detection systems built utilising these methods and their capabilities is shown in Table 5.1. Whilst these detection methods each have their own advantages and disadvantages [194, 195], they are all uniform in the fact that they require additional technical expertise or equipment in order to complete their respective diagnostic tests. In this instance, technical expertise or equipment is defined as necessitating at least basic laboratory skills and/or equipment and therefore rendering the test unusable in a standard General Practice (GP) or test-at-home scenario. There are also numerous commercial lateral flow test (LFT) *E. coli* antigen detection systems available [196, 197]. However, before a diagnosis can be given they require the user to carry out an enrichment step on their sample to cultivate the bacteria, meaning that only next-day results (16+ hours) are achievable.

Table 5.1: ***E. coli* detection method case-study comparison**

Detection Method	Limit of Detection	Speed of Detection	Technical Expertise Required?	Reference
Phage-based	10^5 or 10^2 CFU/mL	3 or 7 hours	Yes	Wang, Chen and Nugen (2017) [198]
ELISA	10^4 CFU/mL	2.5 hours	Yes	Zhao et al. (2020) [199]
AuNP aggregation	50 CFU/mL	1 hour	Yes	Zheng et al. (2019) [200]
PCR	4 CFU/mL	2 hours	Yes	Kim and Oh (2020) [201]
NASBA	<10 CFU/mL	3.5 hours	Yes	Heijnen and Medema (2009) [202]
LAMP	10 CFU/mL	5 hours	Yes	Xia et al. (2021) [203]
RPA	4.4 CFU/mL	10 minutes	Yes	Hu et al. (2020) [204]

5.2.2 Cell-free TXTL synthesis of phage

In what was a monumental achievement at the time, the first demonstration of cell-free infectious phage synthesis was a decade ago [29], where phage T7 and ϕX174 were used as model organisms. Phage T7 was also synthesised a few years later [31]. Quite surprisingly, though, there haven't been many examples of innovation or application using this technology since and its high potential remains to be exploited, as discussed in a recent publication [205].

There are, however, two rare cases of cell-free phage synthesis being innovatively applied available in the literature. The first uses the technology to assemble whole phage genomes that had been constructed *de novo* via gibson assembly, presenting a new phage DNA engineering method which does not at all rely on the host [206], thereby removing the inefficiencies often found in host-mediated phage engineering. The second example, derived from a paper which is still under review [207], quite remarkably presents cell-free as a non-genomic engineering platform for producing single-use phage in a manner that is very similar to the work presented in Chapters 4 and 5 of this thesis. In the unpublished work, phage T7 is manipulated in cell-free so that it assembles with simultaneously provided gp10 fused to NanoLuc, thereby producing a phage with NanoLuc displayed on its outer capsid shell. These recent innovations, along with the work presented in this thesis and predictions from the esteemed Noireaux group [205], confirm the high potential status of cell-free phage synthesis technology, which is set to be exploited in the coming years for an array of fundamental, therapeutic and diagnostic applications.

5.2.3 Research aims

The aims of this chapter are as follows:

1. To demonstrate the host-induced release of packaged fusion protein for K1F*gp6.7::NtNL* and K1F*gp14::NtNL*
2. To explore the potential of cell-free TXTL as a phage engineering and analysis tool
3. To present an optimised diagnostic model for rapidly detecting *E. coli* K1

5.3 Methods

5.3.1 Electron microscopy analysis of cell-free TXTL phage synthesis

10 μL drops of TXTL K1F assembly reaction from different time intervals were applied to the centre of the mesh and were incubated for 1 minute. The samples were removed and the mesh washed twice with 10 μL drops of water and finally negatively stained with 10 μL 2% uranyl acetate for 1 minute. Images were acquired using the Jeol 2100 transmission electron microscope (Jeol, Tokyo, Japan).

5.4 Results and discussion

The amalgamation of the data generated in Chapters 3 and 4 represent the foundational work for creating an engineered phage K1F with diagnostic capabilities. This collection of data, thus far, demonstrated successful ICP::NtNL fusion design, implementation and activity

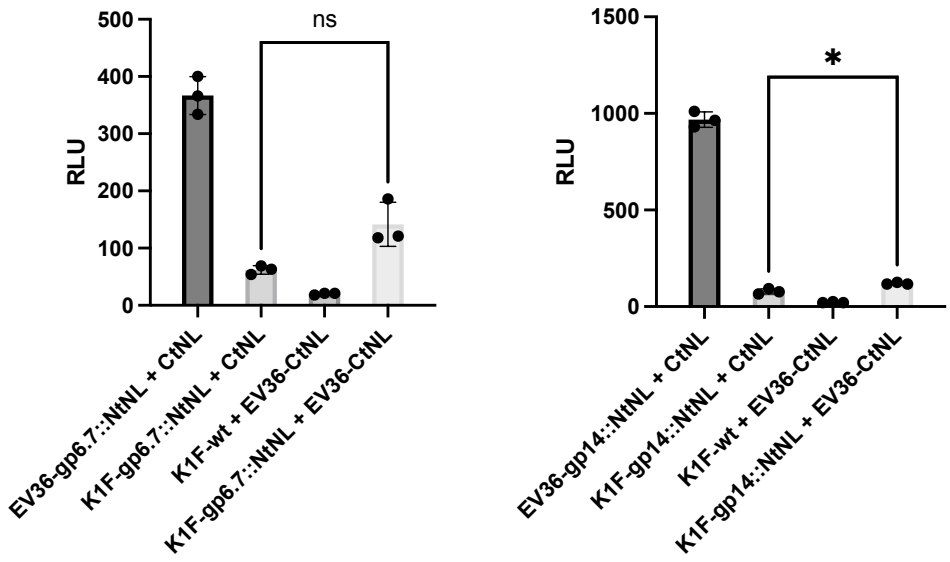
and also, a propitious *in vivo*, non-genomic engineering method for packaging an ICP fusion inside the phage capsid. This final chapter will apply the previously collected data to diagnostic scenarios and explore the efficacy of building a rapid, point-of-care *E. coli* K1 detection device.

5.4.1 Host-induced signal release by *in vivo* non-genomically engineered phage with endogenously supplied CtNL

Following on from the non-genomic *in vivo* ICP phage engineering experiments (Figure 4.22) and subsequently demonstrating the successful packaging of the fusion via heat-induced release of the internal capsid contents (Figure 4.26), the proposition of host-induced signal ejection could now be investigated. It was decided that the starting point for this would be to use a host that is constitutively expressing the CtNL protein (i.e. *E. coli* EV36 transformed with pBEST-CtNL), rather than supplying CtNL externally. It was hypothesised that if the cytoplasm of targeted host cells were affluent in CtNL protein, then this would increase the probability of the newly injected gp6.7::NtNL fusion protein successfully complimenting with its constitutively and cytoplasmically expressed CtNL counterpart, a process which may spare it from the endogenous proteases which patrol the intracellular environment and that would usually degrade gp6.7 during WT phage infection [97]. Theoretically, if following the same fate as WT gp14, the gp14::NtNL fusion would form the outer pore of the ejectosome to aid with genomic translocation into the host [121]. Subsequently, upon phage-induced host lysis, these membrane-lodged gp14::NtNL fusion proteins would be freed and exposed to the divulging CtNL-rich host cytoplasm.

Displayed in Figure 5.1 are the data generated from the first host-induced ICP::NtNL signal ejection attempt, using a CtNL-expressing EV36 host. The data here, whilst not entirely convincing, do potentially portray the characteristics of a host-dependent diagnostic phage. For both engineered phage (modified g6.7 and g14), the 'K1F-gp6.7/14::NtNL + CtNL' (phage only) data represent the background signal that is external to the phage particles in the stock engineered phage solution. For K1F*gp6.7::NtNL* (Figure 5.1a), the difference between 'phage only' and 'phage + host' is observable (when the host is present, the signal does increase), however this difference is not statistically significant and therefore at this stage it would be optimistic to state that the difference is definitely due to the presence of the host. However, the data collected for K1F*gp14::NtNL* (Figure 5.1b) do exhibit a statistically significant difference between the 'phage only' and 'phage + host' RLU outputs, with the 'phage + host' sample generating a higher signal. Furthermore, when K1F-wt is combined with EV36-pBEST-CtNL, a negligible amount of signal is produced - confirming the necessity of both NtNL and CtNL.

It was also subsequently noted that the 'phage + host' samples were being combined at a ratio



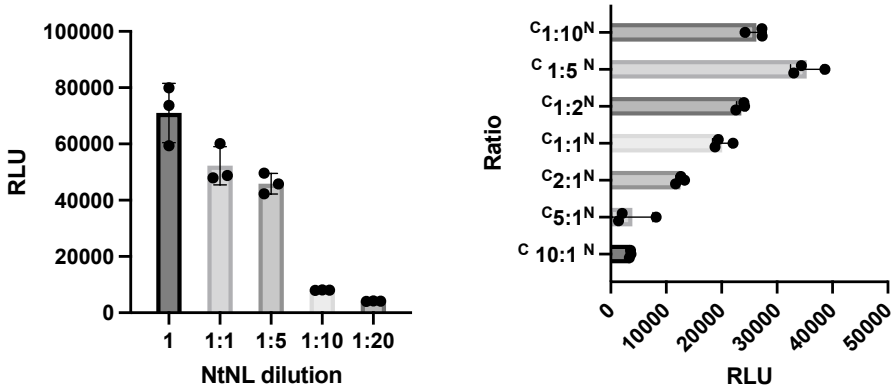
(a) Host-dependent signal generation with K1F*gp6.7::NtNL*

(b) Host-dependent signal generation with K1F*gp14::NtNL*

Figure 5.1: **Host-dependent diagnostic phage with endogenous CtNL.** Luminescence was measured on the engineered phage only, K1F-wt propagated on the CtNL-expressing host and the engineered phage propagated on the CtNL-expressing host. All samples were treated with lysis buffer. For the samples that included the CtNL-expressing host, additional CtNL lysate supplementation was not necessary due to CtNL already being present in the sample. To present the opportunity for spontaneous CtNL::NtNL complementation prior to the Nano-Glo® assay, the K1F-ICP::NtNL only samples were supplemented with EV36-CtNL lysate at a ratio of 1:10 to imitate the dilution made in the K1F-ICP::NtNL + EV36-CtNL experimental samples, (+/- SD, n = 3).

of 1:10 (i.e. 1 mL phage in 10 mL host culture), whereas the 'phage only' samples were being combined with CtNL at a ratio of 1:1. Therefore, the NtNL fusion protein present in the 'phage + host' samples was being diluted to a considerably higher degree compared to the 'phage only' samples and subsequently this may have impacted the results. Furthermore, the 'phage only' samples might generate less signal at a higher dilution and therefore this may improve the statistical significance of the results, whilst simultaneously representing a more equitable comparison between the two samples.

The impact of combining NtNL with water at various dilutions before combining with CtNL is displayed in Figure 5.2a - as NtNL is increasingly diluted, the RLU output monotonically decreases. Moreover, there is approximately a 10-fold difference between the 1:1 and 1:10 dilution, therefore it is crucial to ensure the NtNL dilution in all samples is equivalent (e.g. all samples set at 1:10). It was also decided to measure the RLU output at varying NtNL:CtNL ratios to find the optimal complementation conditions. As shown in Figure 5.2b, and rather surprisingly, the CtNL 1:5 NtNL sample generated the highest signal. Furthermore, the RLU



(a) NtNL dilution series

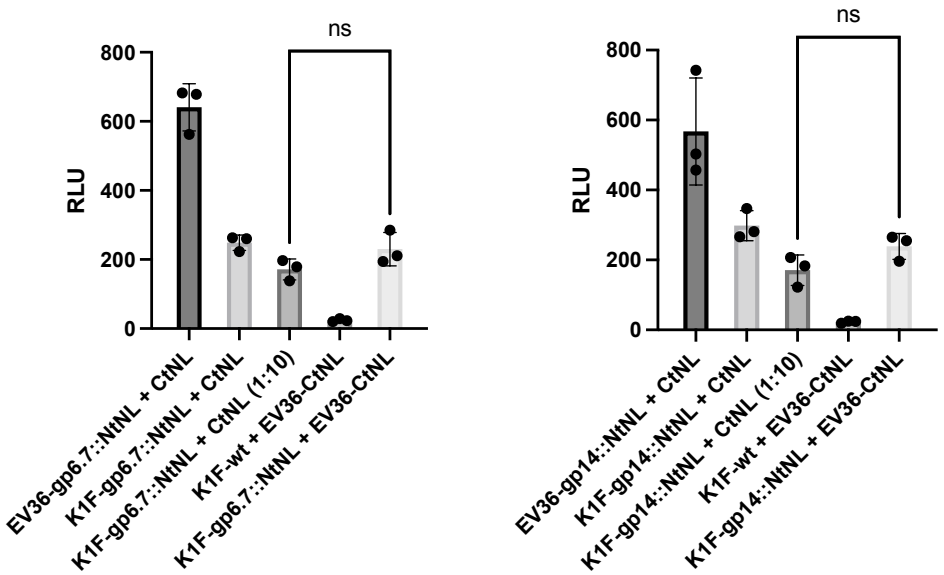
(b) NtNL:CtNL ratio

Figure 5.2: **NanoLuc sub-unit optimal conditions.** NtNL optimisation data for **a:** gp6.7::NtNL was mixed with molecular grade water at various dilutions before combining with an equal volume of CtNL and measuring the RLU output. **b:** gp6.7::NtNL and CtNL were combined at different ratios and the RLU output was measured. The NtNL and CtNL proteins were expressed via cell-free TXTL for **a** and **b**, (+/- SD, n = 3).

output was consistently higher when the ratio included more NtNL than CtNL and the generated signal decreased as the proportion of CtNL in the ratio was increased.

The two graphs in Figure 5.2 suggest that the experimental conditions previously used for the CtNL-expressing host-induced detection assay were not at all optimal for NtNL:CtNL complimentation. Not only were the 'phage + host' samples diluting NtNL ten-fold, but this dilution was being carried out in the CtNL lysate, therefore representing the CtNL 10:1 NtNL sample in Figure 5.2b (i.e. the least optimal complimentation ratio). If CtNL could be supplied externally, then measures could be taken to improve the reaction conditions, but whilst the assay was being carried out in the CtNL-expressing host (i.e. non-external CtNL), these non-optimal reaction conditions would recommence. Nevertheless, it was decided to repeat the endogenous CtNL detection experiment with the inclusion of a 1:10 diluted 'phage only' sample to allow for comparable results between 'phage only' and 'phage + host' (Figure 5.3). For now, the issue of combining NtNL and CtNL at a ratio of 1:10 was not addressed, but as all samples were being subjected to the same conditions, the results (whilst being sub-optimal) would be comparable.

The second attempt at demonstrating host-dependent signal release with the CtNL-expressing host did not yield significant results. As shown in Figure 5.3, whilst diluting the 'phage only' samples 1:10 with CtNL for both K1F*gp6.7::NtNL* and K1F*gp14::NtNL* did incur a slight reduction in RLU output, it was not at all comparable to the 10-fold reduction seen in Figure 5.2a. This is potentially due to the fact that the luminescence generated in the detection assay is at



(a) Second attempt: K1F*gp6.7::NtNL*

(b) Second attempt: K1F*gp14::NtNL*

Figure 5.3: **Second attempt: Host-dependent diagnostic phage with endogenous CtNL.** Luminescence was measured on the engineered phage only, K1F-wt propagated on the CtNL-expressing host and the engineered phage propagated on the CtNL-expressing host. All samples were treated with lysis buffer upon phage-induced lysis. For the samples that included the CtNL-expressing host, additional CtNL lysate supplementation was not necessary due to CtNL already being present in the sample. To present the opportunity for spontaneous CtNL::NtNL complementation prior to the Nano-Glo® assay, the K1F-ICP::NtNL only samples were supplemented with EV36-CtNL lysate at a ratio of 1:10 to imitate the dilution made in the K1F-ICP::NtNL + EV36-CtNL experimental samples, (+/- SD, n = 3).

the very lower end of detectable signal (RLU = 100s), whereas in the dilution assay the signal generated was much higher (RLU = 10,000s). This could be attributed to the fact that the proteins expressed for the dilution assay were produced via cell-free TXTL, whereas the proteins expressed for the detection assay were produced *in vivo*.

In Figures 5.3a and 5.3b the 'phage + host' samples did generate observably higher signal compared to their 1:10 'phage only' counterpart, however, as the results were not statistically significant, further investigation was necessary before any conclusions could be made. Furthermore, it should also be considered that in a true *in vivo* setting (i.e. in a test-at-home detection device), the bacterial host would not be carrying a plasmid instigating the expression of CtNL. Therefore, it is imperative that further experiments are conducted to display a diagnostic model where CtNL can be provided extracellularly along with the other test components (i.e. lysis buffer and NanoLuc substrate - furimazine). One concern with this proposition is that the ejected fusion proteins may be degraded by endogenous host proteases, as has previously been demonstrated for gp6.7 [97, 208], before it is possible to expose them to the CtNL and furimazine. A solution

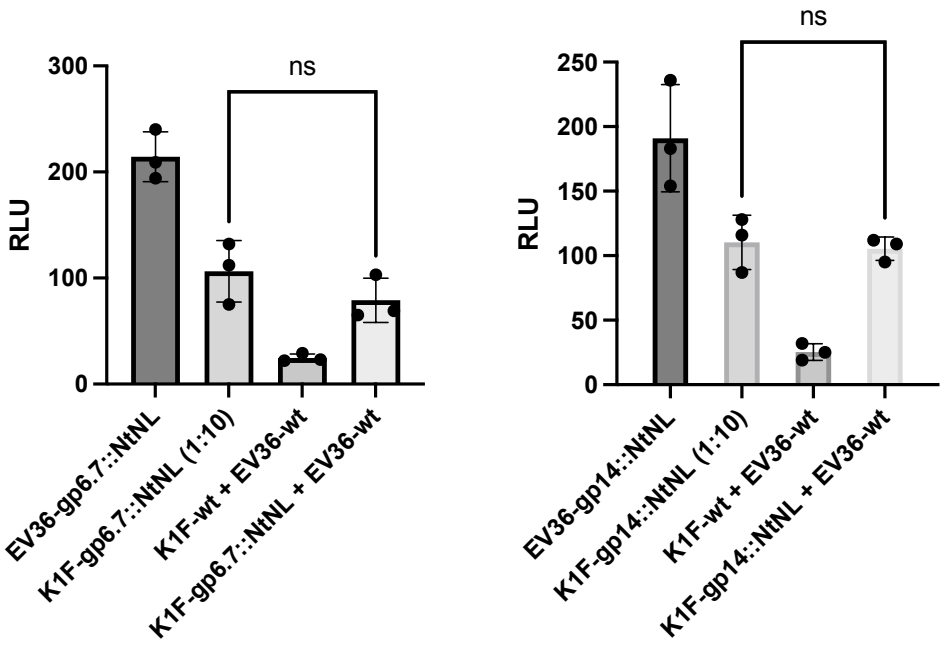
to this could be to supply a protein inhibitor cocktail upon lysis to suppress the protease activity. This will be explored in the subsequent detection attempts.

5.4.2 Host-induced signal release by *in vivo* non-genomically engineered phage with externally supplied CtNL

Although there was a suggestion that the diagnostic assay had worked in the previous section, as no reproducibly significant results could be obtained it was next decided to dissociate from the endogenous CtNL host so that the unfavourable CtNL 10:1 NtNL ratio could be ameliorated. It was suspected that this ratio could be debilitating to the amount of signal generated and therefore could be increasing the difficulty of demonstrating the host-induced signal generation. Another reason for moving on from EV36-pBEST-CtNL, as aforementioned, was to ensure that any results generated could be compared to a real life diagnostic scenario. Moreover, when attempting to detect a pathogen *in vivo*, the pathogen would not be endogenously expressing CtNL, therefore, externally provided CtNL would be a necessary component of the test assay.

For the next detection attempt, the engineered phage were incubated with EV36-wt at a ratio of 1:10 until clearance was observed, then the phage lysates were combined with an equal volume of a TXTL CtNL reaction, left to incubate at room temperature for an hour and finally combined with the Nano-Glo® assay reagent and measured for luminescence. As shown in Figure 5.4, there is no significant difference in the RLU output between the 'phage only' and 'phage + host' samples. In fact, for K1F*gp6.7::NtNL*, the 'phage only' sample generated a slightly observable increased signal compared to the 'phage + host' sample (Figure 5.4a). These experiments were identically repeated multiple times and comparable results were yielded.

Two possible explanations for why the 'phage only' samples generated equal/more amounts of luminescent signal compared to the 'phage + host' samples are, firstly, that a higher efficiency phage packaging method was necessary to increase the amount of encapsulated NtNL and secondly, the ejected fusion proteins were being degraded by endogenous proteases before the opportunity for CtNL was being presented. To elaborate on the first point, perhaps the *in vivo* method actually is a relatively low-performing capsid packaging mechanism and as there is nothing to compare it to (neither in this thesis nor in the literature), it is difficult to gauge this. In fact, *E. coli* EV36 cells are not characterised as being a recombinant protein expression strain in the same way that, for example, BL21 is [209]. Therefore, it is not expected that vast amounts of fusion protein will be expressed within these K1F host cells. Subsequently, if a smaller amount fusion protein is being expressed then it can be reasonably hypothesised that a smaller amount of capsid packaging events will occur and therefore, fewer engineered phage will be generated. If the phage could be synthesised and packaged in cell-free TXTL - a well renowned high expres-



(a) K1F*gp6.7::NtNL* with externally provided CtNL

(b) K1F*gp14::NtNL* with externally provided CtNL

Figure 5.4: **Host-dependent diagnostic phage test with externally provided CtNL.** Luminescence was measured on the engineered phage only, K1F-wt propagated on EV36-wt and the engineered phage propagated on the EV36-wt. All samples were treated with lysis buffer upon phage-induced lysis. To present the opportunity for spontaneous CtNL::NtNL complementation prior to the Nano-Glo® assay, all samples were treated with an equal volume of a CtNL TXTL reaction and left to incubate for 1 hour at room temperature, (+/- SD, n = 3).

sion and highly controllable environment - then this low-efficiency engineering problem could be mitigated.

The second possible explanation for the failed attempt displayed in Figure 5.4 is that the majority of the fusion proteins were being degraded by host proteases prior to CtNL complementation, despite the presence of a protease inhibitor in the lysis buffer that was added upon phage-induced lysis (n.b. lysis buffer was added to all samples, including phage-lysed samples, to allow for a congruous comparison). Perhaps, by the time phage-induced lysis had occurred, the endogenous proteases had already sufficiently degraded a large proportion of the fusion proteins - indeed, it does take approximately 60 minutes for phage-induced lysis to occur, therefore the proteases do have a prolonged period of time to act before the protease inhibitor cocktail can suppress their activity. However, as the publication regarding ICP protease degradation displays, gp14 is not degraded to the same extent as gp6.7 upon ejection into the host [97], therefore, the hypothesis that protease action is suppressing the signal produced by 'phage + host' samples is somewhat

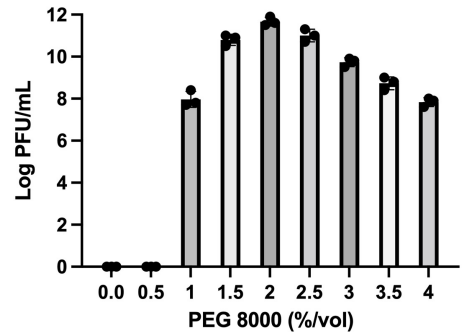
problematic. However, the referenced publication does display that a small amount of gp14 is degraded and as Figure 5.4 shows, the gp14 'phage + host' sample is equal to the 'phage only' sample, whereas the gp6.7 'phage + host' sample is reduced compared to the 'phage only' sample - potentially suggesting that gp14::NtNL is being degraded to a lesser extent compared to gp6.7::NtNL, but still enough to prevent the ejected gp14::NtNL fusion proteins from generating an increased signal compared to the 'phage only' sample.

Nevertheless, regardless of what impact the proteases are having individually to each sample, one pragmatic approach to counteract their effect on a whole is to lyse the cells at the earliest possible time-point with a lysis buffer incorporating a protease inhibitor cocktail, in an attempt to free the ejected fusion into a protected environment before the entirety of it is degraded. It has previously been shown that the phage T7 eclipse time (i.e. the minimum time required for the host to produce the first phage progeny post-infection) can be as quick as 10 minutes [210], therefore, it can be assumed that a proportion of the engineered phage K1F population are capable of ejecting their ICPs 10 minutes after supplying the host.

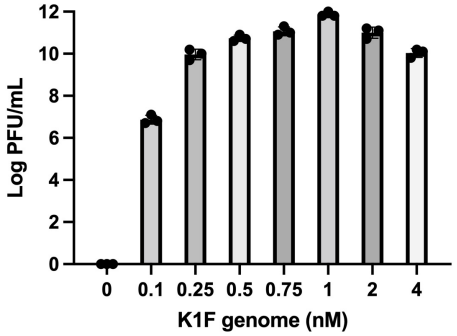
These potential solutions and a new engineering approach are explored in the coming sections. Furthermore, this new proposed approach would utilise the cell-free TXTL phage synthesis in order to facilitate high-efficiency *in vitro* capsid packaging with the ICP::NtNL fusion. Moreover, prior to generating the diagnostic phage via TXTL engineering, preliminary data needed to be collected in order to establish cell-free phage synthesis in the laboratory. These results are presented next, followed by the continuation of the diagnostic phage work.

5.4.3 Cell-free TXTL synthesis of K1F phage

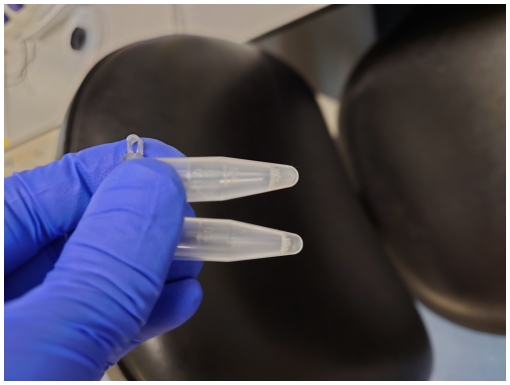
Prior to experimenting with the TXTL engineering method, it was first essential to successfully synthesize K1F phage in a TXTL reaction. The TXTL synthesis of phage was first demonstrated a decade ago, but has been published only a few times since by the same group [29–31] and very recently by a different research group [148, 211]. Fortunately, one of the TXTL phage synthesis publications was regarding T7 phage - for which the genome is comparable to K1F. Therefore, the reaction conditions used in that publication could be used as a starting point for the optimisation experiments for K1F TXTL synthesis. As the demands of phage synthesis on a TXTL system are much higher than single protein expression, an optimal system is preferable. Therefore, it was decided to use the highly powerful myTXTL kit (Arbor Biosciences, Michigan, USA), as this had already been purchased to compare with the in-house TXTL system in Chapter 3. Moreover, high-throughput use of the system was not anticipated during these TXTL packaging experiments, so it was expected that enough reagents for this application were already available in the lab (approximately 90 reactions).



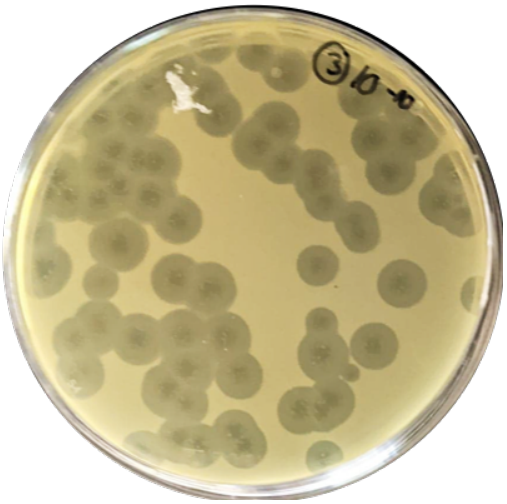
(a) K1F TXTL synthesis as a function of PEG 8000 concentration



(b) K1F TXTL synthesis as a function of genome concentration



(c) Comparison between a TXTL reaction expressing phage and a TXTL reaction expressing a single recombinant protein



(d) Plaque assay of TXTL-synthesised K1F-wt on EV36-wt

Figure 5.5: **myTXTL optimisations for cell-free K1F-wt synthesis.** The optimal PEG 8k %/volume (**a**) and K1F genome concentration (**b**) were found by testing a range of different values in the myTXTL system, (+/- SD, n = 3). The 12 μL reactions were carried out in triplicate and incubated for 16 hours at 29 °C. Successful TXTL phage synthesis reactions typically appear more opaque (**c**). Plaque assays were carried out on a series of TXTL reaction dilutions with EV36-wt to calculate the phage titer (**d**).

It has subsequently been demonstrated that K1F can be synthesised to a high titer using the in-house cell-free system described in Chapter 3 and publication of this work is impending. However, for the entirety of the TXTL phage engineering work presented in this chapter, the myTXTL kit was used. Figure 5.5 displays the data and images generated from optimising K1F synthesis in the myTXTL system. One early sign of a successful TXTL phage synthesis reaction (prior to validating via a spot test of plaque assay) is observing whether the reaction has become opaque or not [211]. As shown in the upper tube in Figure 5.5c, the K1F synthesis reaction has caused the liquid to turn opaque, whereas the reaction in the bottom tube (a recombinant protein expression) has remained relatively clear. Further analysis via a plaque assay enables an accurate

titer to be calculated. Optimal reaction conditions in myTXTL produced a very high K1F titer of approximately 10^{12} PFU/mL.

The optimal amounts of PEG and K1F genomic DNA are also displayed above. As Figure 5.5a shows, a minimum of 1%/vol of PEG 8k is necessary for the reaction to be viable, and when increasing PEG 8k from 1% - 2% (the optimal amount), a 10,000-fold increase in PFU/mL is exhibited - validating the importance of PEG 8k in cell-free reactions. Previous research has found the optimal genome concentration for T7 in a TXTL reaction to be 0.25 nM [31], however as shown in Figure 5.5b, the optimal genome concentration for K1F found here is 1 nM.

5.4.4 An investigative platform comprising cell-free TXTL and electron microscopy for studying phage synthesis

Following on from successfully optimising the myTXTL platform for K1F synthesis, a side project became of interest whereby it was hypothesised that the open and controllable cell-free environment would be well-suited for time-course imaging of phage. Traditionally in phage-related studies, electron microscopy (EM) would be used solely for capturing stand-alone images

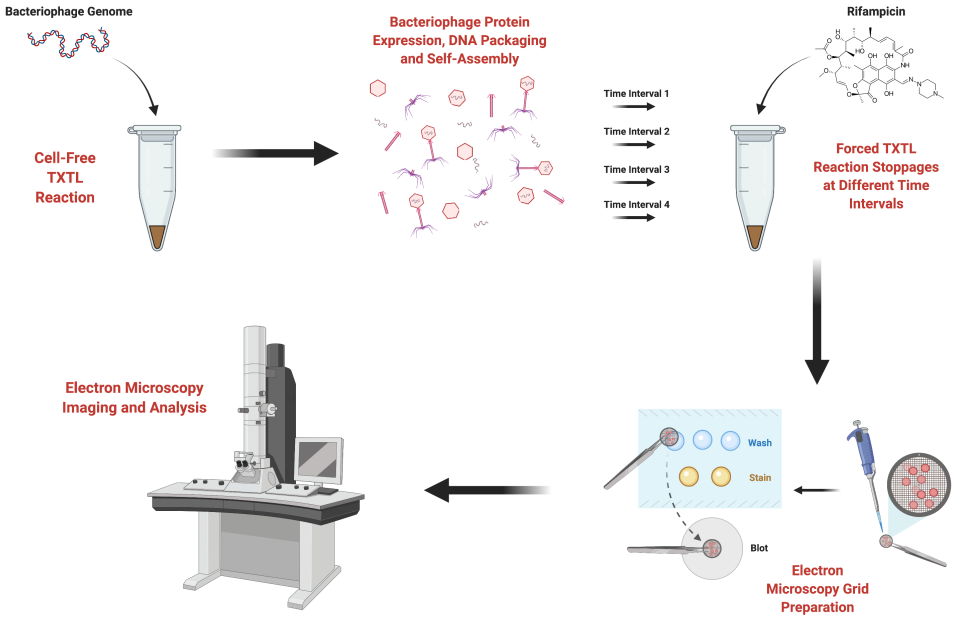


Figure 5.6: **Workflow for TXTL synthesis-mediated electron microscopy phage assembly analysis.** Phage are assembled in a cell-free reaction and the reactions are stopped at different time intervals. Each time interval reaction is then visualised via electron microscopy.

of phage in certain scenarios. However, this novel proposition enables the user to discover and track the characteristics of phage synthesis over the course of a controlled TXTL reaction whilst aligning the images with the corresponding phage titer for further insight. Furthermore, due to the highly controllable TXTL environment, it is possible to precisely quantify the start and end points of DNA expression (i.e. when the genome is added to the reaction and when the transcriptional inhibitor, rifampicin, is added to the reaction). This defined level of control is not easily attained *in vivo* due to the turbulent nature of phage propagation and opaque composition of membrane-encapsulated bacterial cells.

Figure 5.6 displays the proposed workflow, where cell-free TXTL phage synthesis reactions are stopped at various time intervals with the addition of 100 μg/mL rifampicin - an RNA-polymerase deactivator which directly inhibits transcription and therefore any subsequent translation and phage synthesis processes within the reaction are also ceased. Then, 10 μL of the inhibited TXTL reactions are directly applied to electron microscopy grids and imaged with EM - subsequently producing a 'real-time' image of each TXTL phage synthesis time interval reaction. By carrying out a plaque assay on each time interval reaction it is also possible to map the images to their corresponding phage titer.

To display a representation of the different stages of TXTL phage synthesis (whilst considering

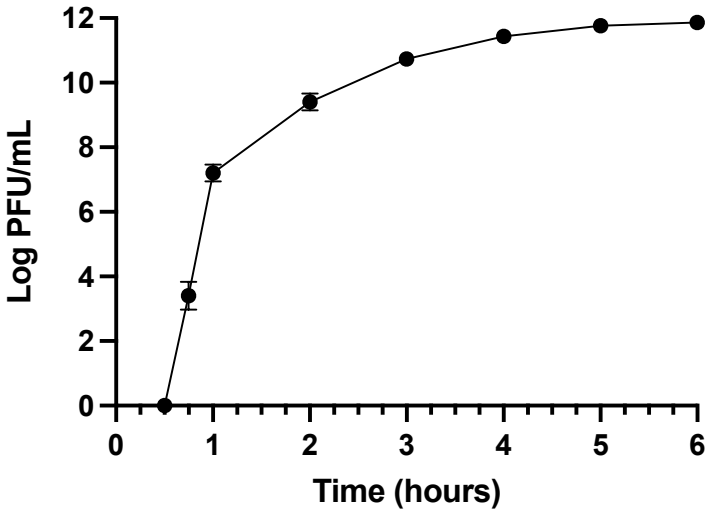


Figure 5.7: **Kinetics of K1F TXTL synthesis.** A concentration of 1 nM K1F genome and 2% PEG were added to the myTXTL master mix for this experiment. Plaque assays were carried out to calculate titers for each time interval. Each sample was done in triplicate. No phage is produced during the first 30 minutes of incubation. The first recorded PFU is at 45 minutes and K1F synthesis typically reaches plateau within 4-5 hours at approximately 10^{12} PFU/mL, (+/- SD, n = 3).

the high cost of EM imaging), five time interval reactions were selected for imaging: 15 mins, 30 mins, 45 mins, 1 hour and 3 hours. As shown in Figure 5.7, the region of rapid K1F phage synthesis in TXTL spans the aforementioned time intervals and after 3 hours the phage titer begins to plateau between 10^{11} - 10^{12} PFU/mL. Between 0 - 1 hours a rapid increase in K1F synthesis is observed, with the PFU/mL rising from 0 - 10^{7}. Then, from 1 - 3 hours a further 1000-fold increase in phage titer is observed. It was therefore of high interest to use this novel TXTL/EM approach to visually observe any changing characteristics in phage assembly or propensity over the course of these first 3 hours of TXTL synthesis.

The first two K1F TXTL synthesis time interval reactions that were imaged were 15 and 30 minutes. No observations could be made 15 minutes after the reaction had started (Figure 5.8). The images captured mainly consisted of the grey/black 'background noise' and sporadic lipid/protein clusters that are typical of a standard *E. coli* lysate EM image. In fact, a small aliquot of an empty myTXTL reaction (i.e. not expressing any DNA) was also analysed via EM and this produced comparable images to those seen at the 15 minute time interval. This is not surprising though, given that no K1F TXTL synthesis PFU titer is observed until the 45 minute mark (Figure 5.7).

A number of sporadic phage capsids could be seen at the half-hour mark (Figure 5.9), suggesting that protein expression and phage assembly were underway within 30 minutes of beginning the K1F TXTL synthesis reaction. The fact that still no PFU titer is observed at the 30 minute time interval (Figure 5.7) suggests that little or none of the phage capsids seen in these images represent viable phage, and will therefore be referred to as procapsids. As a typical component of dsDNA phage assembly and maturation, a procapsid is defined as a DNA-free phage capsid

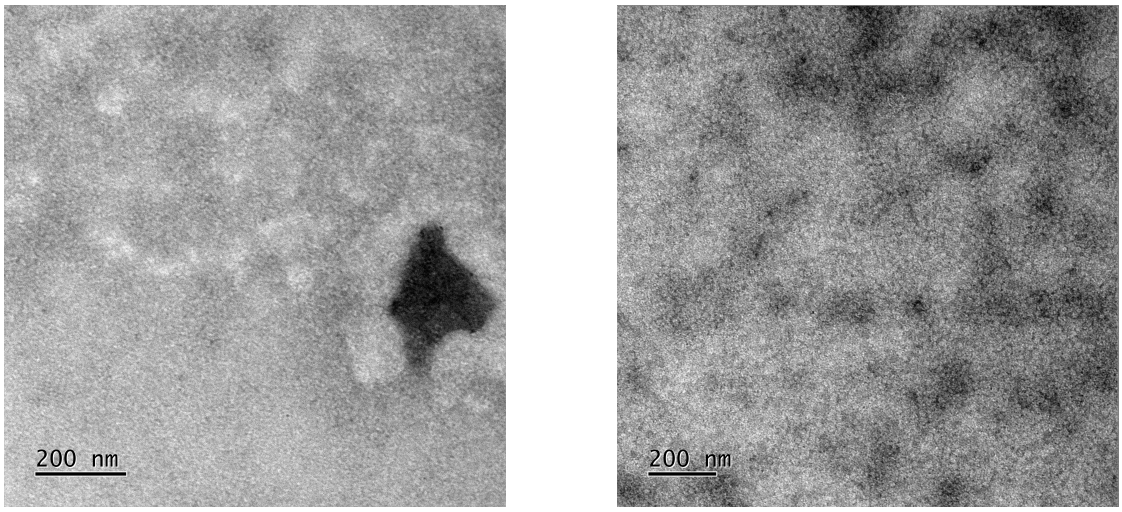


Figure 5.8: **15 minute time interval: K1F TXTL synthesis visualised with EM.**

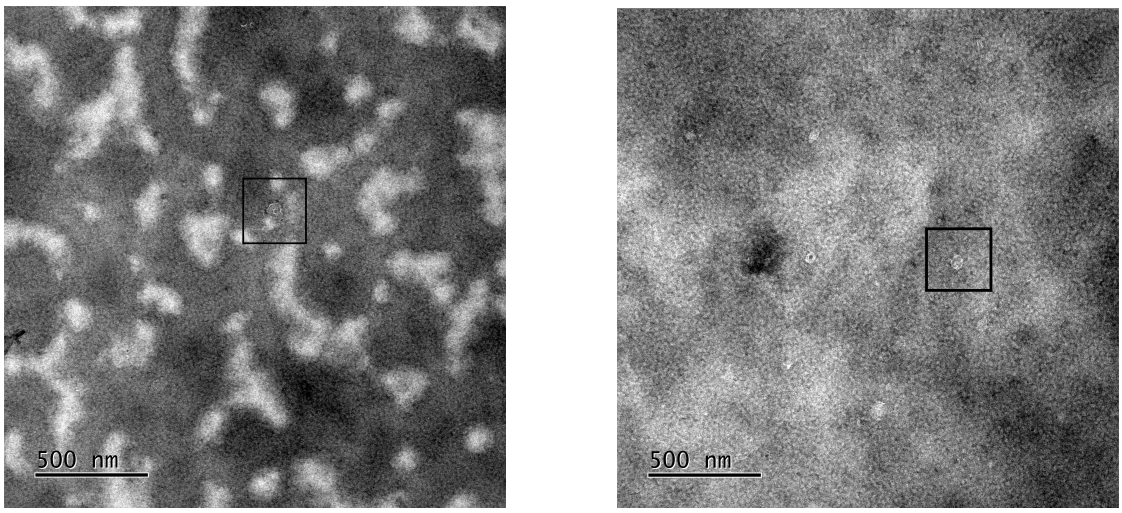


Figure 5.9: **30 minute time interval: K1F TXTL synthesis visualised with EM.**

that is usually more spherical in shape and smaller than a matured capsid. Over the course of the assembly of a viable phage, the procapsid undergoes a dramatic conformational change whilst packaging the genomic DNA, which for phage T7, results in a larger, more angular capsid exhibiting the typical icosahedral structure [212] - this can be expected for phage K1F too. The identified procapsids can be seen within the black boxes in the two EM images displayed in Figure 5.9.

Interestingly, for the 45 minute K1F TXTL synthesis time interval reaction - which corresponds with the first observable K1F phage titer of 10^3 PFU/mL (Figure 5.7), EM analysis reveals that

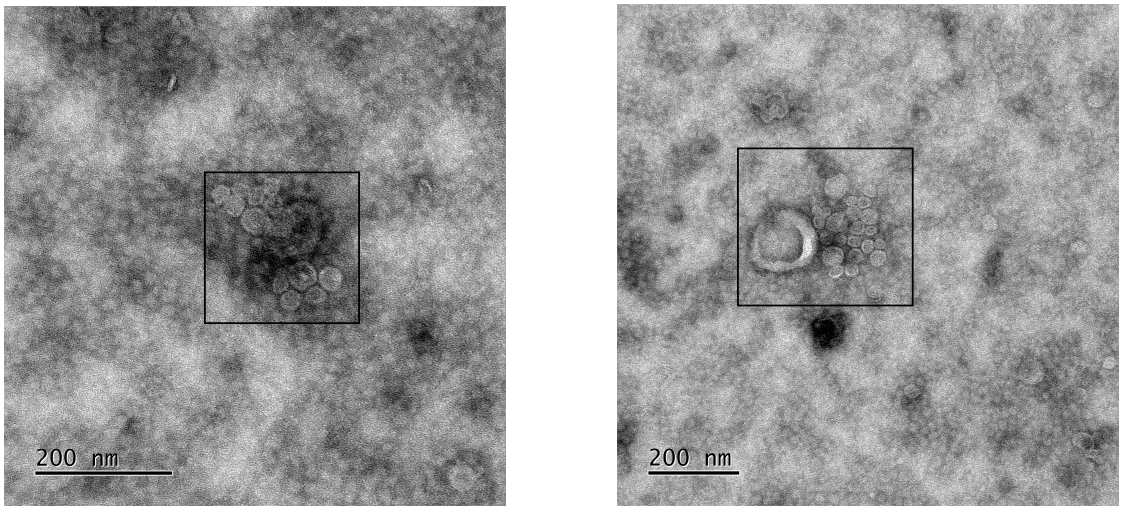


Figure 5.10: **45 minute time interval: K1F TXTL synthesis visualised with EM.**

the procapsids appear to be accumulating together (Figure 5.10). One hypothesis for this procapsid accumulation phenomenon (PAP) is that, following on from phage protein expression and procapsid formation, the procapsids accumulate together whilst DNA packaging is undergone. Perhaps the DNA concatemers that are typical of T7-like phage serve as the rendezvous point for the PAP. Again, the events of interest are highlighted by the black boxes in the images shown in Figure 5.10.

The 1 hour time interval displayed below in Figure 5.11 exhibits a continuation of the proposed PAP. Procapsid accumulation was observed at an increased abundance during EM imaging of these samples compared to the 45 minute samples. Furthermore, the 1 hour K1F TXTL synthesis time interval reaction aligns to a titer of 10^7 PFU/mL (Figure 5.7). One interpretation

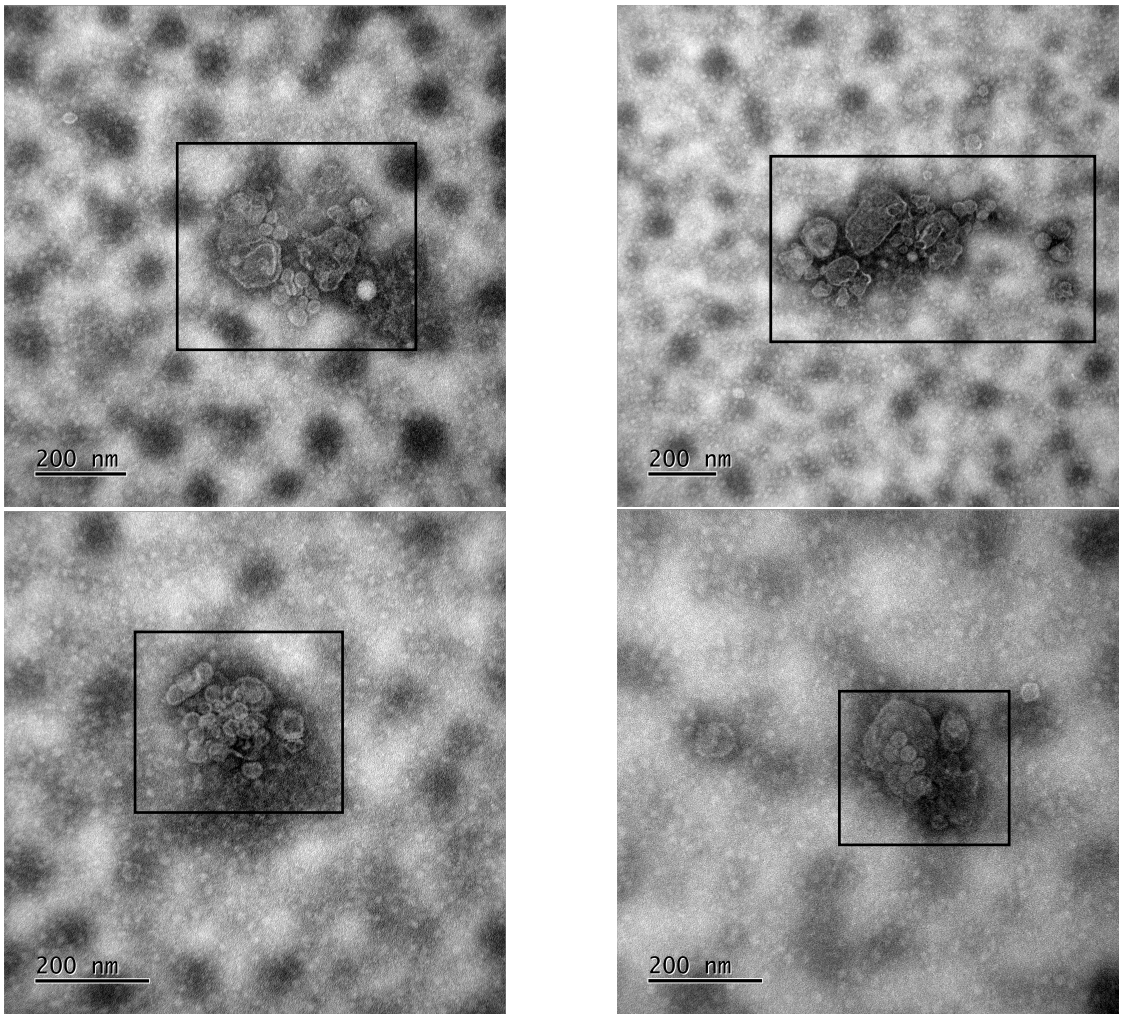


Figure 5.11: **1 hour time interval: K1F TXTL synthesis visualised with EM.**

of this data could suggest that the events displayed in the 45 and 60 minute images are in fact 'DNA packaging' events. This would suggest that premature phage accumulate together in large DNA packaging events as the last stage of their development and after which, they become viable phage. Furthermore, the increased abundancy of the PAP, and therefore possible DNA packaging events, from the 45 minute - 1 hour time interval could represent an acceleration of capsid maturation and subsequently viable phage synthesis. Indeed, the result of procapsid DNA packaging is in fact the generation of viable phage [212]. The rapid shift from 0 PFU/mL at 30 minutes, to 10^3 PFU/mL at 45 minutes, to 10^7 PFU/mL at 60 minutes would support this conclusion.

The EM imaging results for the final K1F TXTL synthesis time interval reaction - 3 hours - display an entirely new characteristic not seen during imaging of the previous time interval

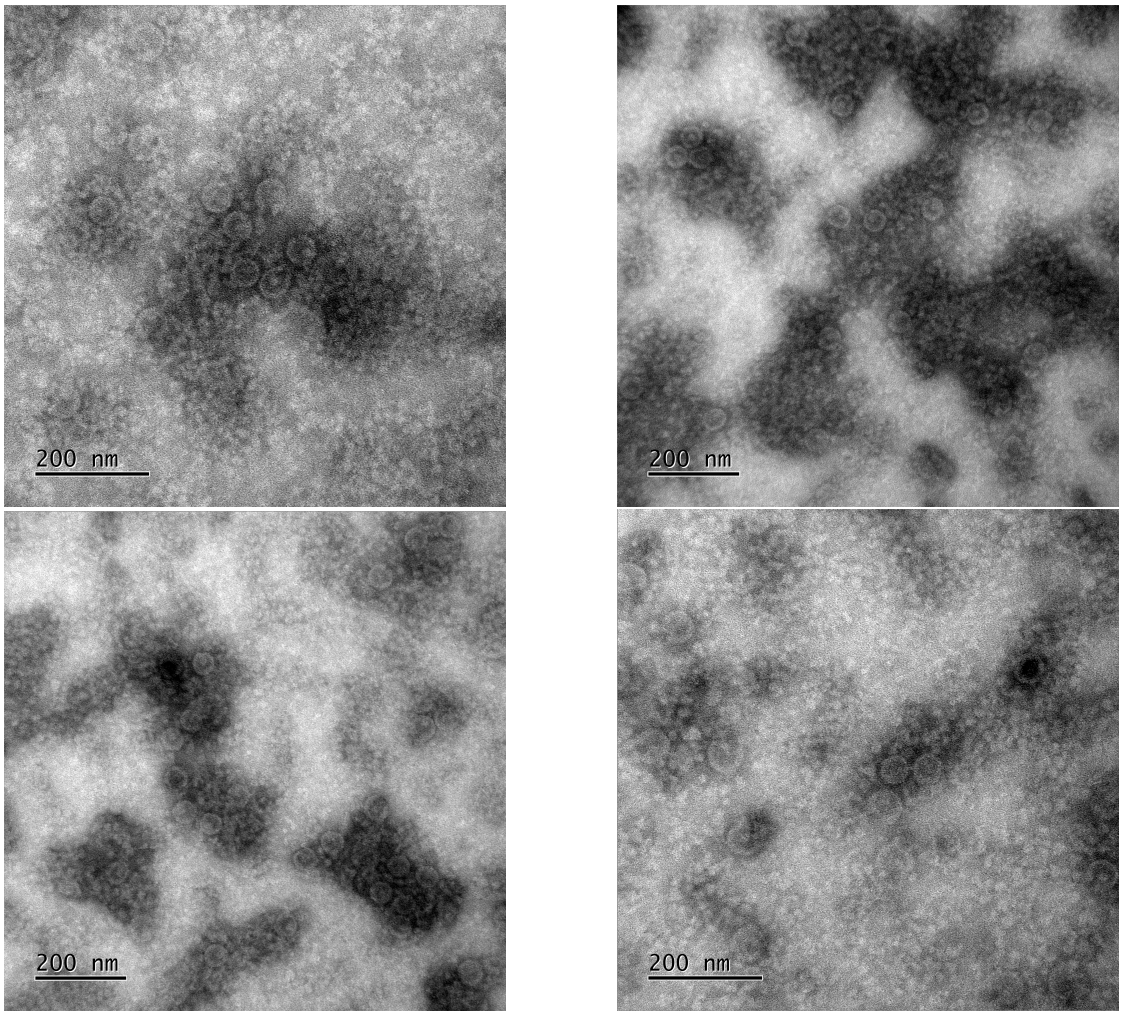


Figure 5.12: **3 hour time interval: K1F TXTL synthesis visualised with EM.**

reactions. The abundancy of phage particles can be seen to have increased, however, they appear to be much less accumulative (Figure 5.12). This suggests that the PAP and DNA packaging events are mostly complete by the 3 hour mark and that almost all phage synthesised in the reaction are viable at this time interval - this interpretation is supported by the fact that the titer observed here is 10^{10} PFU/mL and over the course of the following 5 hours phage synthesis decelerates and plateaus at approximately 10^{12} (Figure 5.7).

Moreover, the phage displayed in the images in Figure 5.12 appear slightly larger (approximately 50 nm in diameter) than at previous time intervals and also exhibit a more uniform and angular conformation, which is typical and would be expected of a mature K1F capsid. The lack of the PAP in these images is not surprising as, at this plateauing stage of K1F phage synthesis within the TXTL reaction, the majority of the phage have already been synthesised and matured and the transcriptional and translational machinery alongside the energy reagents within the cell-free reaction have been exhausted almost to completion. Therefore, it is not expected that a significant amount of new procapsid formation and accumulation would be observable compared to the amount of viable phage already present.

A deep search of the literature revealed that, whilst various mentions of procapsid accumulation are found in an array of decades worth of research papers, they all seem to refer to a process whereby defective capsids which are unable to package their DNA accumulate together. This suggests that the PAP is in fact a known process (albeit, perhaps unsurprisingly, it hasn't previously been referred to specifically as "the PAP"), however, it is not thought to be responsible for that which was hypothesised in the previous paragraph (i.e. DNA packaging) - seemingly quite the opposite. For example, in a publication from the mid 1970s [213], it is described and visually demonstrated that the procapsids of P22, a phage that infects *S. typhimurium* which is actually very similar to T7 and K1F, accumulate amid "vegetative DNA" concatemers in a way that is comparable to that which is shown in Figures 5.10 and 5.11. More recently, it has also been shown that when SPP1, a phage that infects *B. subtilis*, is engineered so that it cannot package DNA into its capsid, the procapsids which are generated accumulate together [114] and for defective ϕ12, a *S. aureus*-infecting phage, where the disarming of DNA packaging machinery, again, leads to procapsid accumulation [214].

However, if the PAP events are an exclusive occurrence for defective phage, then it is surprising that none of these events could be found at the 3-hour mark (Figure 5.12), as there is no reason why they would disappear. One explanation for what is occurring, now hypothesised after considering the results found in the literature alongside the EM/TXTL findings, is that the PAP occurs universally for all phage and it is in fact representative of DNA packaging events (regard-

less of whether those packaging events are successful or not). The reason why it has previously been described as a characteristic of defective phage [114, 213, 214], is because these phage are not capable of progressing past the DNA packaging stage of maturation and therefore remain in the accumulative state. However, it is hypothesised here that "normal" phage also go through the PAP phase whilst they gather around DNA concatemers in a collective (perhaps cooperative? e.g. sharing packaging machinery and/or ATP energy) DNA packaging event. Furthermore, once they have successfully packaged their DNA and become mature phage there is no longer a necessity for accumulation and they progress to the mature phase captured in Figure 5.12.

This concludes this chapter's K1F TXTL synthesis investigative detour. The novel EM/TXTL phage assembly analysis workflow presented offers a new possibility for phage investigation whereby researchers can observe the step-by-step progression of synthesis in a highly controllable environment without disturbing the sample. Future applications could allow for the manipulation of a plethora of parameters (only limited by the researchers imagination) to be visualised in real time and potentially, the uncovering of unexpected findings (such as the hypothesised PAP-mediated DNA packaging events).

Furthermore, the remainder of the results subsequently presented in this chapter utilise this cell-free phage synthesis platform to continue and optimise the ICP::NtNL fusion phage work, in an attempt to conclusively demonstrate host-induced signal release and refine the diagnostic assay.

5.4.5 *in vitro* non-genomic TXTL ICP phage engineering

Whilst the *in vivo* non-genomic engineering approach did produce some initially promising results (namely Figures 4.26 and 5.1), where heat-induced and host-induced signal ejection were at least superficially demonstrated, it was not possible to make any firm conclusions and either accept or reject the hypothesis that the ICP::NtNL fusion can be packaged inside the phage capsid and subsequently be released and detectable upon the induced ejection of the phage internal capsid contents.

Therefore, an optimised non-genomic engineering approach was devised whereby cell-free TXTL is utilised for the expression of the ICP::NtNL fusion *and* phage genome simultaneously, which can conceivably generate a packaged diagnostic phage. This approach is preferable compared to the *in vivo* approach due to the refined and controllable environment of cell-free alongside its superior protein expression capabilities. This new approach is visualised in Figure 5.13. As shown in this figure, there are two possible TXTL phage engineering approaches - the first comprising the simultaneous expression of the phage genome and pBEST-ICP::NtNL plasmid and the second expressing the phage genome whilst supplying pre-TXTL'd ICP::NtNL fusion protein as

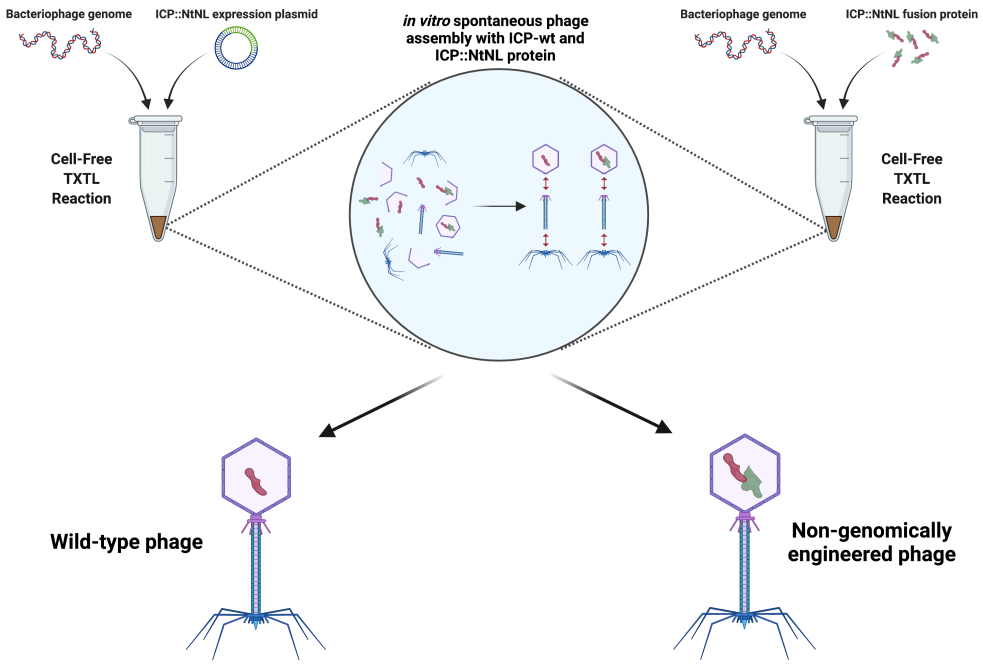
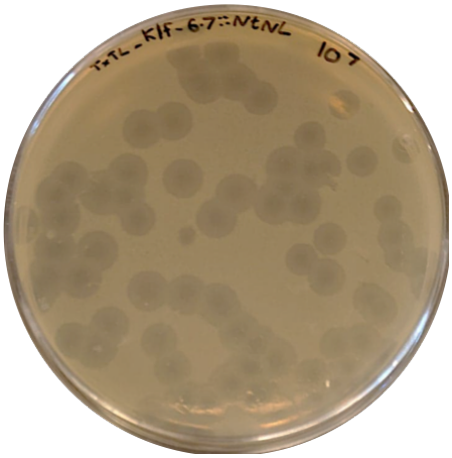


Figure 5.13: **Non-genomic TXTL ICP phage engineering strategy used for packaging the NtNL fusion protein into the phage capsid.** Phage are packaged in a cell-free reaction either with co-expressed fusion protein (endoTXTL) or pre-expressed fusion protein (exoTXTL), generating a mixed population of WT and non-genomically engineered phage.

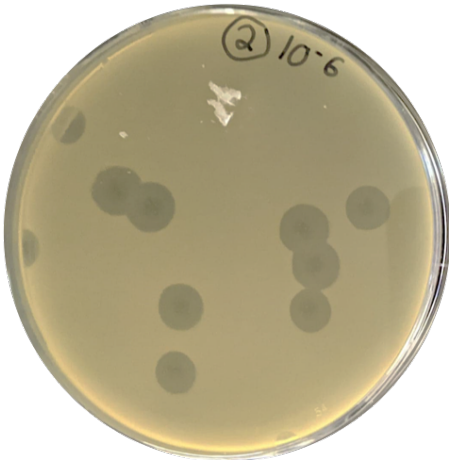
a reaction additive. It was planned for both methods to be attempted in order to find the optimal approach, whilst considering both the packaging efficiency and phage titer generated. The two approaches will be referred to as: endoTXTL (endogenous and simultaneous expression of phage and fusion DNA) and exoTXTL (phage DNA expression supplemented with exogenous fusion protein).

Figure 5.14 displays the plaque assays resulting from plating EV36-wt with either the endoTXTL or exoTXTL engineering reactions for each gene (i.e. g6.7 and g14). Moreover, the plaque assay results show that the endoTXTL engineering attempts successfully yielded viable phage, however, the exoTXTL engineering attempts failed to generate any phage. All subsequent repeats using the exoTXTL method were also unsuccessful.

A possible explanation for this is that the presence of the fusion protein at the start of the exoTXTL reaction may have interfered with the molecular crowding equilibrium maintained by PEG 8k or otherwise destabilised the TXTL ecosystem leading to the failure of K1F synthesis. The PEG 8K-mediated molecular crowding within TXTL is especially important for phage



(a) Plaque assay of endoTXTL engineering attempt for K1F*gp6.7::NtNL*



(b) Plaque assay of endoTXTL engineering attempt for K1F*gp14::NtNL*

(c) Plaque assay of exoTXTL engineering attempt for K1F*gp6.7::NtNL*

(d) Plaque assay of exoTXTL engineering attempt for K1F*gp14::NtNL*

Figure 5.14: **Plaque assay results for EV36-wt with non-genomic endoTXTL and exoTXTL phage engineering attempts.** All TXTL engineering attempts were carried out with the previously optimised myTXTL K1F synthesis platform, either adding 18 nM pBEST-ICP::NtNL DNA or 1 μL of a ICP::NtNL fusion protein TXTL reaction alongside the K1F genomic DNA. TXTL phage engineering reactions were incubated at 29 °C for 16 hours.

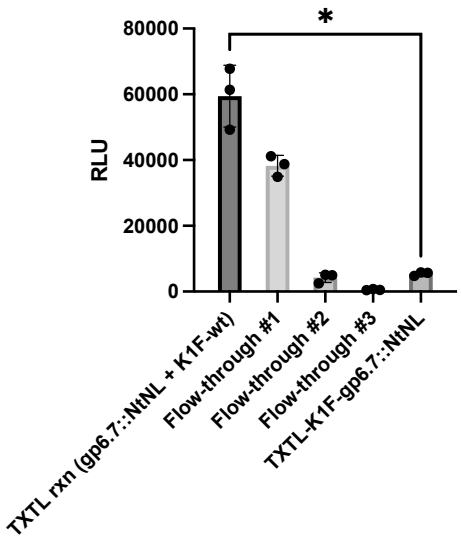
synthesis [30, 31] and so any interference with this is likely to be detrimental for the reaction. Subsequently, rather than spending a significant amount of time on attempting to extensively optimise the exoTXTL approach (e.g. by adding different exogenous protein concentrations and/or purifying the protein before adding it and/or adjusting the PEG 8k concentration), it was decided to proceed with the endoTXTL phage engineering approach.

On the first attempt, the endoTXTL approach yielded a titer of 1.5×10^8 PFU/mL for K1F*gp6.7::NtNL*

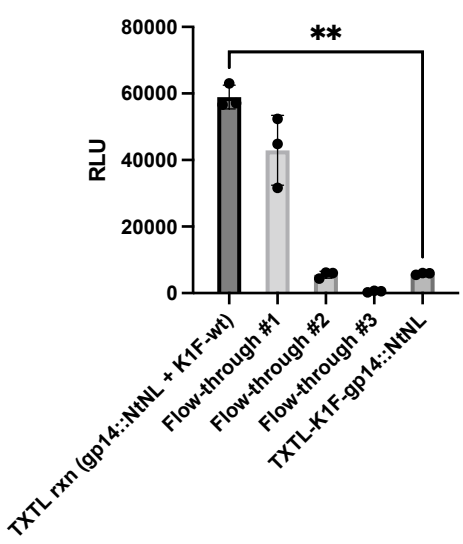
and 2.2×10^6 PFU/mL for K1F*gp14::NtNL* (Figures 5.14a and 5.14b respectively), however, it should be noted that subsequent attempts yielded a titer in the region of 10^9 PFU/mL for both engineered phage. Nevertheless, even at 10^9 PFU/mL, an approximate 1000-fold decrease is observed when comparing to the titer generated from the K1F-wt genome-only myTXTL experiments displayed in Figure 5.5. This considerable difference can be attributed to the increased workload and protein cost that the myTXTL system is burdened with by adding the pBEST plasmid alongside the K1F genome.

Similarly to the *in vivo* non-genomic phage engineering process (Figure 4.23), several aliquots were kept from the different stages of the endoTXTL engineering process and measured for luminescence by combining them with TXTL-expressed CtNL and the Nano-Glo® reagents. This data is shown below in Figure 5.15. The engineered phage generated from endoTXTL reactions will subsequently be referred to as TXTL-K1F*gp6.7::NtNL* and TXTL-K1F*gp14::NtNL* to avoid any confusion with previous phage engineering attempts.

Whilst the pattern that can be seen in both graphs displayed above in Figure 5.15 (i.e. lumi-



(a) K1F*gp6.7::NtNL* non-genomic endoTXTL phage engineering workflow data-points.



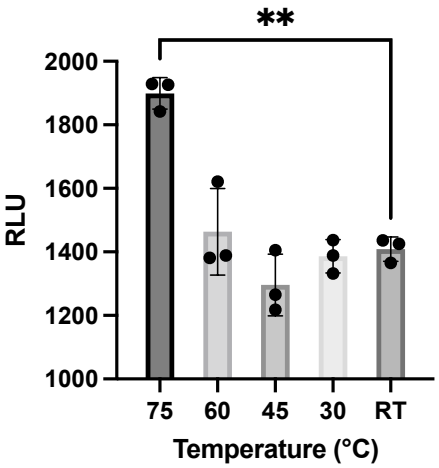
(b) K1F*gp14::NtNL* non-genomic endoTXTL phage engineering workflow data-points.

Figure 5.15: **Non-genomic TXTL engineering workflow data-points.** Luminescence was measured on aliquots taken at several stages during the **a:** K1F*gp6.7::NtNL* and **b:** K1F*gp14::NtNL* non-genomic endoTXTL engineering process. Flow-through samples were taken after passing the phage through a 100 kDa Amicon® Ultra filtration device three times, using a fresh device each time. All samples were combined with an equal volume of a CtNL TXTL reaction and left to incubate for 1 hour at room temperature to allow for spontaneous CtNL::NtNL complementation prior to the Nano-Glo® assay, (+/- SD, n = 3).

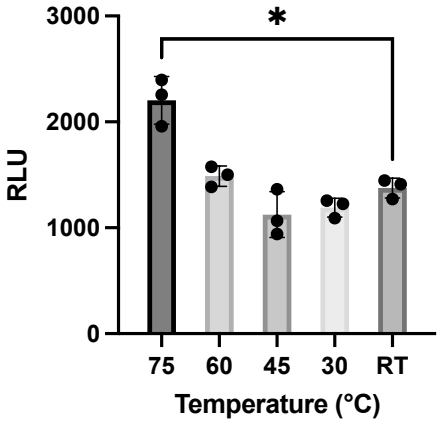
nescence decreases after each filtration stage) was to be expected, one additional interesting observation is that, on a whole, the RLU values are more than 10-fold higher than the corresponding data points from the *in vivo* engineering process (Figure 4.23). This suggests that the TXTL system is capable of producing a significantly higher quantity of ICP::NtNL fusion protein compared to EV36 cells. Furthermore, it is hypothesised that this increase in fusion protein concentration throughout the phage assembly environment will increase the probability of phage particles successfully packaging the fusion protein within their capsid structure. Finally, a background RLU value is again observed in the filtered TXTL-K1F-ICP::NtNL samples, therefore, this background signal must be taken into account during the subsequent diagnostic assay attempts.

5.4.6 Heat-induced signal release by *in vitro* non-genomically TXTL engineered phage

In an attempt to indicate whether the endoTXTL engineering reactions had successfully packaged the ICP::NtNL fusion inside the K1F capsids, the heat-induced signal release assay was carried out. The rationale and experimental conditions from the *in vivo* phage engineering work (Figure 4.26) were implemented here, therefore, these details won't be repeated.



(a) *TXTL-K1F*gp6.7::NtNL *heat-induced load ejection*



(b) *TXTL-K1F*gp14::NtNL *heat-induced load ejection*

Figure 5.16: **Luminescence measurements after heating non-genomically TXTL engineered phage to different temperatures.** All samples were exposed to the desired temperature, then allowed to return to room temperature, then combined with an equal volume of CtNL TXTL reaction and left to incubate for 1 hour at room temperature to allow for spontaneous CtNL::NtNL complementation prior to the Nano-Glo® assay, (+/- SD, n = 3).

The room temperature (RT) values in Figures 5.16a and 5.16b represent the background luminescence present in the external solution surrounding the phage particles for TXTL-K1F*gp6.7::NtNL* and TXTL-K1F*gp14::NtNL* respectively. When exposed to 75 °C heat, both phage exhibited a statistically significant rise in RLU output compared to the RT values - indicating the confirmation of heat-induced release of encapsulated ICP::NtNL fusion from the phage and therefore suggesting the endoTXTL engineering experiments were a success.

For both phage, when heated to 30 °C the RLU output decreased slightly but was comparable to the RT value. Moreover, heating to 45 °C incurred a further observable decrease in signal - suggesting the rising temperature was having a denaturing effect on the ICP::NtNL fusion. When the phage were subjected to an increased temperature of 60 °C, however, the mean RLU values increased to slightly above the background RT level, when it would be expected that the signal would continue to decrease as the temperature increases (please see Figure 4.25 for a reminder on the denaturing effect of temperature on NtNL). This strongly suggests that a significant portion of fusion protein that was previously undetectable (i.e. packaged inside the capsid) had suddenly been released and become available for CtNL complementation, therefore increasing the signal output despite the enhanced denaturing capacity facilitated by the higher temperature. The considerable increase in RLU at the 75 °C temperature, where NtNL should be significantly denatured, adds further confirmation that the fusion had been packaged. Quite remarkably, it would seem that the act of being ejected out of the capsid protects NtNL from the high temperature environment it is ejected into. Hypothetically (and somewhat far-fetched), perhaps the physically demanding process of capsid ejection, in some way, prepares the fusion for the heat stress and prevents extensive degradation from occurring. Alternatively, perhaps the fusion proteins re-associate with empty capsids or other phage proteins which subsequently prevents them from becoming degraded. Regardless of what the mechanism of survival is, the data are still highly indicative of successful fusion packaging.

Next, to further consolidate the acceptance that the fusion proteins had been successful packaged, a SYBR green assay (Thermo Fisher Scientific, Waltham, USA) was carried out to confirm the release of DNA (indicating the simultaneous release of ICPs) at high temperatures. SYBR green is a dsDNA binding dye which, when bound to dsDNA, emits fluorescence [215]. Therefore, the more dsDNA present, the higher the fluorescent signal. The results generated from this assay are displayed in Figure 5.17.

At temperatures ranging from RT to 45 °C, a small consistent RFU value is emitted, which can be considered as the background value. However, when the phage were heated to 60 °C, the fluorescence measurement slightly increased above the background value, potentially indicating that

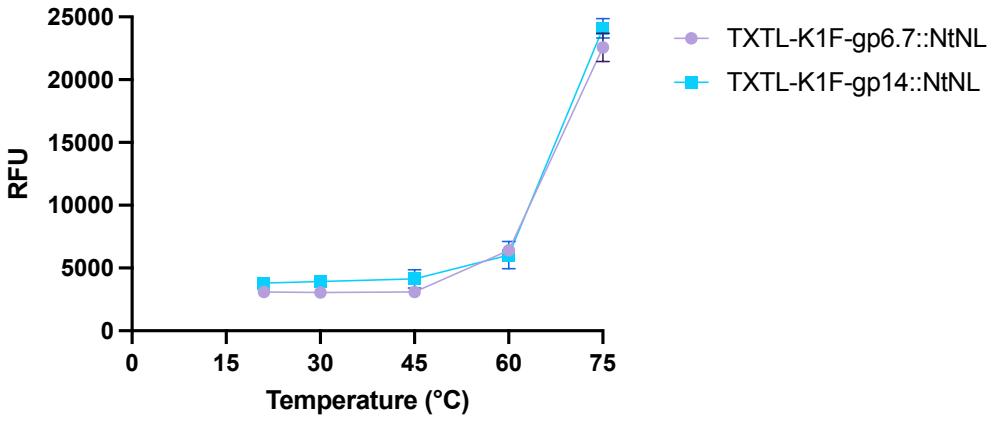


Figure 5.17: **Fluorescence measurements after heating non-genomically TXTL engineered phage to different temperatures.** All samples were exposed to the desired temperature, then allowed to return to room temperature, then mixed with the SYBR green reagent before being measured for fluorescence on a plate reader, (+/- SD, n = 3).

a small amount of DNA had been released from the phage. It is possible that this small amount of DNA represents the preliminary nucleotides which initially get ejected upon host adsorption. Usually, the rest of the genome is "pulled" into the host cytoplasm by the cytoplasmic RNA polymerases [117], however, in this case (i.e. in Figure 5.17) there are no RNA polymerases present as the ejection is heat-induced rather than host-induced. Therefore, it is reasonable to suggest that the SYBR green units are only able to bind to the small amount of preliminarily ejected DNA and subsequently, only a small amount of fluorescence is emitted. Alternatively, this slight increase in RFU could also represent a very small proportion of the phage prematurely ejecting their entire genome. At 75 °C, the RFU value significantly increases and clearly demonstrates the full release of the genomic DNA from a large proportion of the phage population. This was to be expected, as it has previously been shown for phage T7 [183] and was discussed in the previous chapter of this thesis.

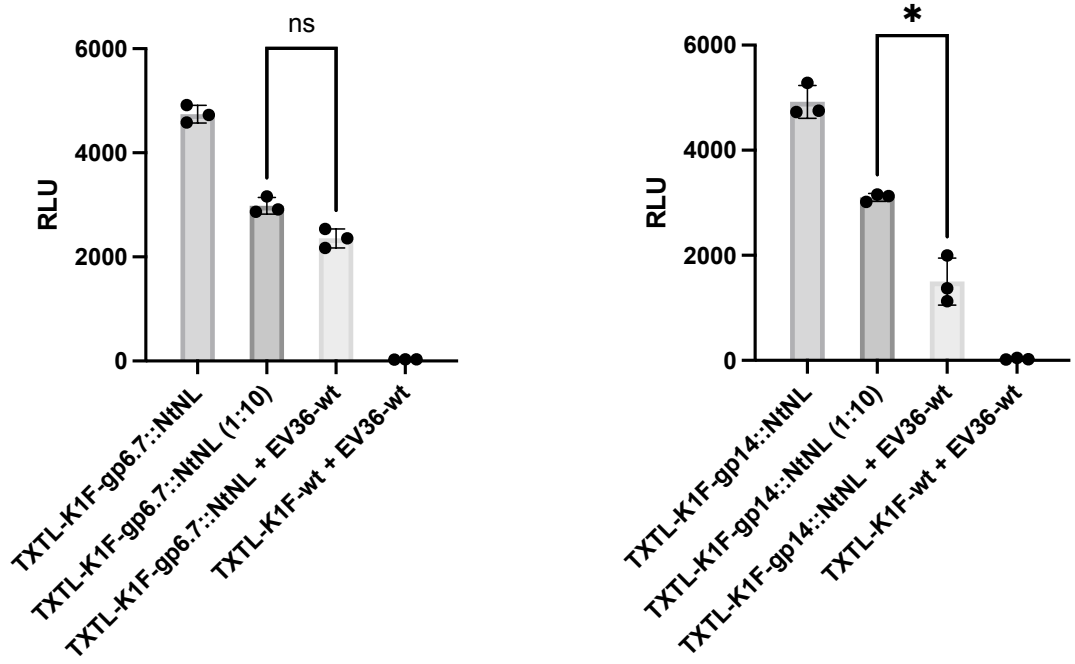
Relating back to the heat-induced fusion protein ejection data shown in Figure 5.16, it can be hypothesised that, at 60 °C, a small proportion of phage are compelled to release their DNA, and therefore their ICPs and NtNL::ICP fusions would precede this. Furthermore, at 75 °C, a large proportion of the phage population are forced to eject their genomic by the high heat pressure and subsequently, a significant amount of the encapsulated fusion is released.

Following on from gaining these initially promising results, it was now possible to approach the host-induced signal release experiments knowing that, at least, it was highly likely that the fusion proteins had been successfully packaged.

5.4.7 Host-induced signal release by *in vitro* non-genomically TXTL engineered phage with externally supplied CtNL

As previously mentioned, it was favourable to move away from the endogenously expressing CtNL host as this was not representative of a real-life scenario and supplying CtNL protein externally would allow for a more controlled and optimised diagnostic assay. Furthermore, the host-induced signal release assays with externally provided CtNL for the most recently generated and verified engineered phage - TXTL-K1F*gp6.7::NtNL* and TXTL-K1F*gp14::NtNL* - commenced.

The results for the first attempt are shown below in Figure 5.18. For this assay, the 'phage + host' samples (10 μL engineered K1F in 100 μL EV36-wt) were incubated using a scaled down version of the standard phage propagation method until clearance was observed, the lysate was then incubated with CtNL at room temperature for 1 hour to allow for NtNL:CtNL complimentation, then the luminescence was measured with the Nano-Glo® reagents. The reason for scaling down



(a) TXTL-K1F*gp6.7::NtNL*

(b) TXTL-K1F*gp14::NtNL*

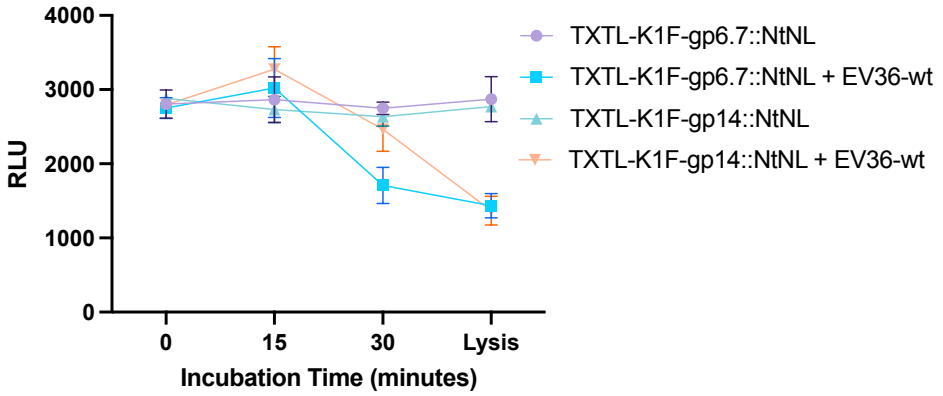
Figure 5.18: **Luminescence measurements after incubating non-genomically TXTL engineered phage with EV36-wt.** Luminescence was measured on the engineered phage only, K1F-wt propagated on EV36-wt and the engineered phage propagated on the EV36-wt To present the opportunity for spontaneous CtNL::NtNL complementation prior to the Nano-Glo® assay, all samples were treated with an equal volume of a CtNL TXTL reaction and left to incubate for 1 hour at room temperature, (+/- SD, n = 3).

was due to the limited volume capacity and cost-restrictions of cell-free TXTL (i.e. 12 μL TXTL reactions compared to 10 mL *in vivo* propagations). The 'phage only' samples were mixed with LB media to the same dilution as the 'phage + host' samples (i.e. 1:10) and all other processing was identical to allow for a equitable comparison.

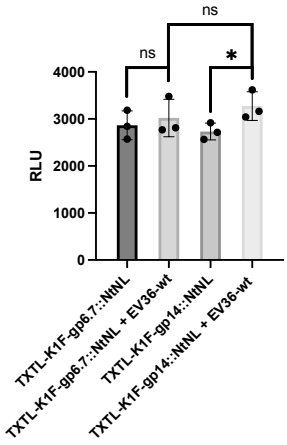
As expected, when combining the 'TXTL-K1F-wt (i.e. a standard K1F-wt TXTL synthesis reaction) + host' samples (i.e. the negative control) with CtNL, no signal was generated - demonstrating the necessity of having both CtNL and NtNL for the luminescence assay to generate light. For TXTL-K1F*gp6.7::NtNL*, no statistically significant difference was found between the 'phage only' and 'phage + host' samples (Figure 5.18a), however there is an observably higher signal output from the 'phage only' sample. Moreover, the 'phage only' sample for TXTL-K1F*gp14::NtNL* exhibited a statistically significant increase in RLU output compared to the 'phage + host' sample. At face value, these results are highly contradictory of a successful diagnostic system, however, they could be attributed to the endogenous protease activity of the host as aforementioned.

In an attempt to overcome this challenge, again as aforementioned, it was hypothesised that if the 'phage + host' samples were stopped at various time intervals and immediately lysed with a lysis buffer + protease inhibitor cocktail then the signal-debilitating action of the endogenous proteases may be avoided or at least partly diminished. With the T7 eclipse time known to be as fast as 10 minutes [210], it is expected that the ICP::NtNL fusion is capable of being ejected within 10 minutes of supplying the host. Furthermore, the second attempt of host-induced signal release assays with externally provided CtNL was carried out with these parameters considered. This time, multiple identical propagations were initiated and then stopped at various time points with the lysis buffer + protease inhibitor cocktail. They were then placed on ice until all the propagations were complete, at which point all samples were then mixed with an equal volume of CtNL and incubated for 1 hour to allow for spontaneous NtNL:CtNL complimentation to occur. The results for this experiment are displayed in Figure 5.19.

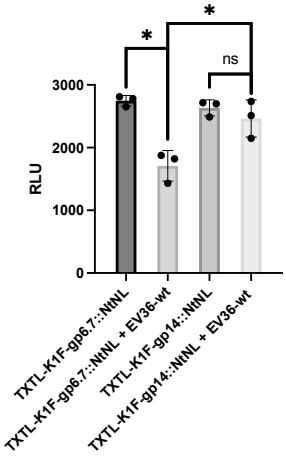
The data are first collectively displayed in a line graph (Figure 5.19a) to allow for a clear observation to be made on the RLU output for each sample over the course of the experiment (i.e. from the start time (0 minutes) up to phage-induced host lysis). For TXTL-K1F*gp6.7::NtNL* and TXTL-K1F*gp14::NtNL*, the 'phage only' samples maintain a stable RLU output over the course of the experiment - this is not surprising, as incubating these samples in LB media for approximately 1 hour (this is usually the time taken for lysis to occur for the 'phage + host' samples and therefore is the amount of time the 'phage only' samples were incubated in LB for) is not expected to impact the luminescent properties of the samples at all.



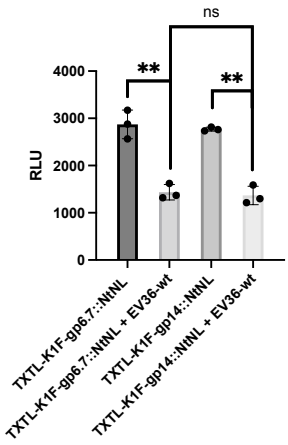
(a) Time-course signal generation with externally provided CtNL



(b) 15 minute time point



(c) 30 minute time point



(d) Lysis time point

Figure 5.19: **Luminescence measurements after incubating non-genomically TXTL engineered phage with EV36-wt for different time periods.** Luminescence was measured on the engineered phage only (diluted in LB media), K1F-wt propagated on EV36-wt and the engineered phage propagated on the EV36-wt at three different time points. At the specified time points, a lysis buffer + protease inhibitor cocktail was added to the samples and they were then immediately placed on ice. The 'Lysis' time point was defined as when the culture was observably cleared by phage lysis - this amount of time was used for the 'phage only' samples. To present the opportunity for spontaneous CtNL::NtNL complementation prior to the Nano-Glo® assay, all samples were treated with an equal volume of a CtNL TXTL reaction and left to incubate for 1 hour at room temperature, (+/- SD, n = 3).

The 'phage + host' samples for both phage begin at a comparable RLU output to the 'phage only' samples, however, at the 15 minute time point a slight observable increase in luminescence is seen - for TXTL-K1F*gp14::NtNL* this increase is higher than for TXTL-K1F*gp6.7::NtNL*. Moreover, Figure 5.19b shows the 15 minute time point data for all four samples, allowing for a statistical comparison to be made. Whilst the TXTL-K1F*gp6.7::NtNL* 'phage only' and 'phage + host' samples are not significantly different (although a small increase in RLU for 'phage

+ host' can be seen), the RLU output of the 'phage + host' TXTL-K1F*gp14::NtNL* sample is higher than its 'phage only' counterpart with statistical significance - potentially indicating that the host-induced release of encapsulated ICP::NtNL fusion protein is responsible for the increase in luminescence. At the 30 minute time point (Figure 5.19c), the TXTL-K1F*gp6.7::NtNL* 'phage + host' sample exhibits a statistically significant and clearly observable drop in luminescence compared to its 'phage only' sample. The RLU output of the TXTL-K1F*gp14::NtNL* 'phage + host' sample is also seen to reduce when comparing to the previous time point, however, when comparing to its 'phage only' counterpart at the 30 minute time point this reduction is not statistically significant. When comparing the two engineered phage with each other, the difference between the RLU output of the 'phage + host' samples at 30 minutes is statistically significant, with TXTL-K1F*gp14::NtNL* emitting a higher amount of luminescence compared to TXTL-K1F*gp6.7::NtNL*. Finally, after approximately 1 hour of incubation when the lysis time point samples were taken (Figure 5.19d), both of the 'phage + host' RLU outputs had decreased again - this time there was negligible difference between the two phage samples and they gave a comparable value of approximately 1300 RLU each.

The general pattern seen here is:

1. After a short period of incubation (similar to the T7 phage eclipse time), 'phage + host' RLU outputs are increased compared to 'phage only' samples

2. As the incubation time prolongs, the 'phage + host' RLU outputs decrease - with TXTL-K1F*gp6.7::NtNL* decreasing much more than TXTL-K1F*gp14::NtNL*

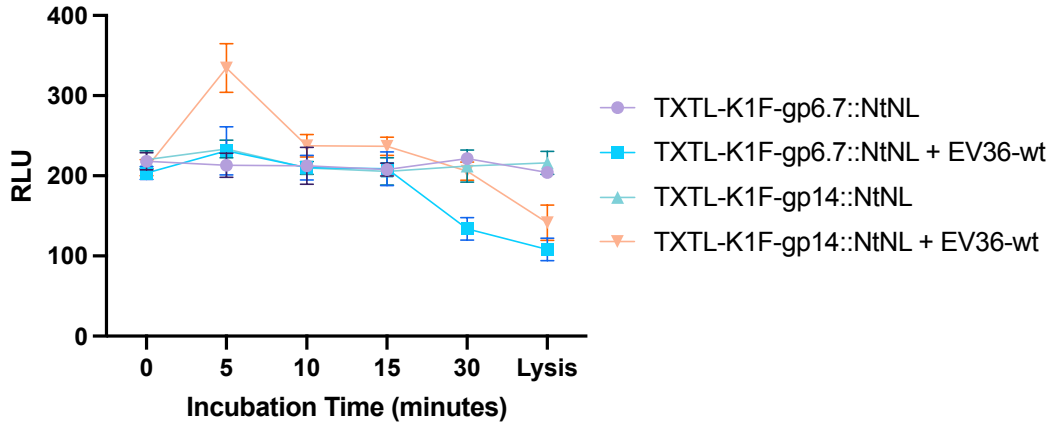
3. Following on from phage-induced lysis, the 'phage + host' RLU outputs decrease even more - with both engineered phage ending up emitting a similar RLU output

It would seem that the attempt to combat the endogenous protease activity with early lysis plus the addition of a protease inhibitor had been successful. The difference in luminescence between the 15 minute and Lysis time points for the 'phage + host' samples for both TXTL-K1F*gp6.7::NtNL* and TXTL-K1F*gp14::NtNL* is clear and this can confidently be attributed to the protease activity being inhibited at 15 minutes but not inhibited by the time phage-induced lysis had occurred (i.e. after approximately 1 hour). This is not surprising, as the longer the period of time the endogenous proteases have with the fusion proteins without the interference of a protease inhibitor, the more likely it is that the fusion proteins will be degraded. An explanation as to why gp6.7::NtNL is significantly degraded at the 30 minute time point whereas gp14::NtNL is not, is that, because gp6.7::NtNL is directly ejected into the cytoplasm, it can immediately start to be degraded by cytoplasmic proteases upon ejection. However, gp14::NtNL is lodged within the host envelope and therefore is hypothetically protected from proteases until cell lysis

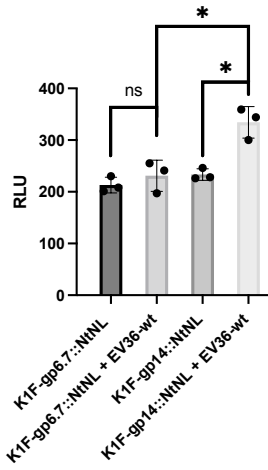
occurs and subsequently the envelope is disassociated and the divulging endogenous proteases (now extracellular) obtain their first opportunity to degrade gp14::NtNL. Continuing with this hypothesis, the slight decrease in luminescence for TXTL-K1F*gp14::NtNL* at 30 minutes suggests that some cell lysis had already occurred and subsequently, a portion of the gp14::NtNL fusion had been degraded.

To investigate this further and in more detail, the experiment was repeated with two extra time points around the T7 eclipse time - 5 minutes and 10 minutes. The data from this repeat are displayed in Figure 5.20. Similarly to the previous experiment, the 'phage only' samples for each engineered phage can be seen to consistently emit a non-fluctuating RLU output throughout the whole incubation time period (Figure 5.20a). In fact, a similar general pattern can be seen for all four samples as described above for the in the previous paragraph - adding a degree of reliability to the key conclusion that a statistically significant increase in luminescence is seen (for TXTL-K1F*gp14::NtNL*) shortly after incubation with the host begins (relating to the T7 eclipse time and ICP::NtNL fusion ejection being responsible for the spike in luminescence). Furthermore, possibly due to the quality of the externally provided CtNL, the RLU values for this experiment were smaller by approximately 10-fold compared to the previous experiment even though the same phage stocks and identical reaction conditions were used. Fortunately though, this isn't detrimental to the primary objective of the experiment - which is to observe the pattern of luminescence over time in comparison to other samples from the same experiment - the logarithm of the relative luminescence value is not relevant as all samples for each experiment are being measured on the same plate and are therefore comparable to each other. General patterns are comparable between different experiments, however, the specific RLU values are not.

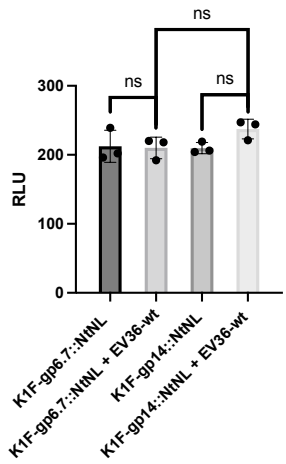
For the 5, 10 and 15 minute time points (Figures 5.20b - 5.20d), the RLU values for both of the TXTL-K1F*gp6.7::NtNL* samples are comparable to each other and no statistically significant difference is found (it is noted though that at the 5 minute mark the mean 'phage + host' RLU output is slightly higher than the 'phage only' value). For TXTL-K1F*gp14::NtNL*, at the 5 minute time point (Figure 5.20b) the RLU output for the 'phage + host' sample is clearly higher than its 'phage only' counterpart with statistical significance. For the next two time points (10 and 15 minutes), the TXTL-K1F*gp14::NtNL* 'phage + host' RLU output remains higher than 'phage only', however, this difference is no longer statistically significant. This possibly suggests that, whilst the increased signal is still observable at he 10 and 15 minute time points (which indicates successful host-induced ejection of the fusion protein), the signal is most potent when it is first released from the phage (i.e. the 5 minute time point). Perhaps, as the wait time increases, the gp14::NtNL fusion becomes more immersed within the host envelope and therefore is less readily available for CtNL complimentation.



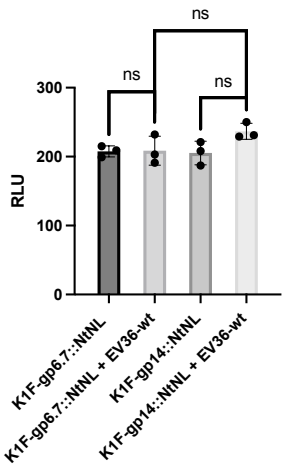
(a) Time-course signal generation with externally provided CtNL



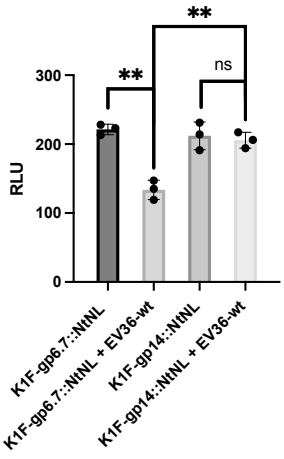
(b) 5 minute time point



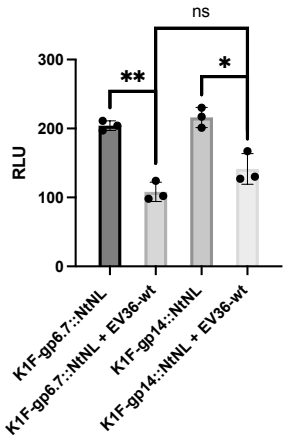
(c) 10 minute time point



(d) 15 minute time point



(e) 30 minute time point



(f) Lysis time point

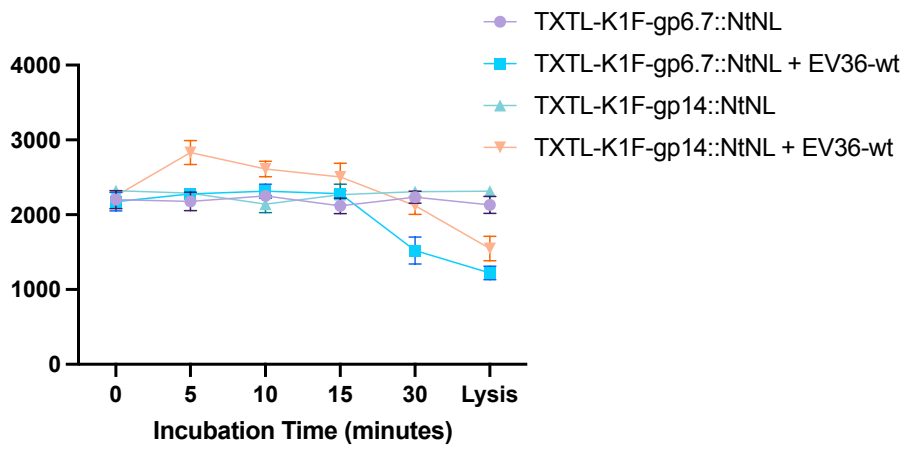
Figure 5.20: **Luminescence measurements after incubating non-genomically TXTL engineered phage with EV36-wt for different time periods (second attempt).** The experiment was repeated identically to Figure 5.19 with the addition of two more time points, (+/- SD, n = 3).

When analysing the 30 minute time point data (Figure 5.20e), both the 'phage only' TXTL-K1F*gp6.7::NtNL* and 'phage + host' TXTL-K1F*gp14::NtNL* samples emit a statistically significant increased amount of luminescence compared to the TXTL-K1F*gp6.7::NtNL* 'phage + host' sample, suggesting high endogenous protease activity on gp6.7::NtNL but not on gp14::NtNL. After phage-induced lysis had occurred (Figure 5.20f), the RLU output of both the TXTL-K1F*gp6.7::NtNL* and TXTL-K1F*gp14::NtNL* 'phage + host' samples had significantly decreased - as observed before in the previous experiment (Figure 5.19). The drop in luminescence is statistically significant compared to the 'phage only' counterparts at the lysis time point.

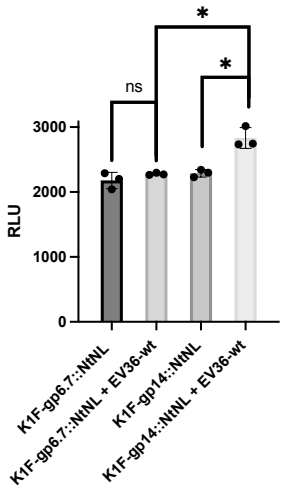
The data collected over the course of the diagnostic assays with *in vitro* non-genomically TXTL engineered phage suggest that the longer the incubation time with the host is (without a lysis buffer + protease inhibitor cocktail intervention), the more the ICP::NtNL fusion proteins are broken down and deactivated. Furthermore, this observation is more fast acting with gp6.7::NtNL compared to gp14::NtNL.

For clarity, the experiment was repeated for a third time with identical conditions to the previous attempt. The data generated from this final attempt are displayed in Figure 5.21a. Moreover, whilst it would seem that the g6.7 engineered phage had completely failed as a diagnostic phage during this attempt, the g14 engineered phage yet again displayed statistically significant characteristics of an effective diagnostic phage. This time, the RLU output emitted by the 'phage + host' TXTL-K1F*gp14::NtNL* sample was higher than the 'phage only' sample with statistical significance for the 5 *and* 10 minute time points (Figures 5.21b and 5.21c). Combining this observation with the finding from the 15 minute time point from Figure 5.19 would suggest that it is possible to detect a potent luminescent output within the first 15 minutes of combining the phage with the host. However, it is preferable to undergo the detection process at the earlier time point (i.e. 5 minutes) to prevent the signal from diminishing. The reason for the signal diminishing within the first 15 minutes is not clear, as it is not expected that a significant amount of host lysis and therefore protease release will have occurred that soon, but it has been suggested that perhaps, as time goes on, the gp14::NtNL fusion becomes increasingly concealed and/or inhibited by the host envelope. Furthermore, following on from the 15 minute time point in Figure 5.21, generally, the same observations can be made as with the previous detection attempts with the TXTL-engineered phage.

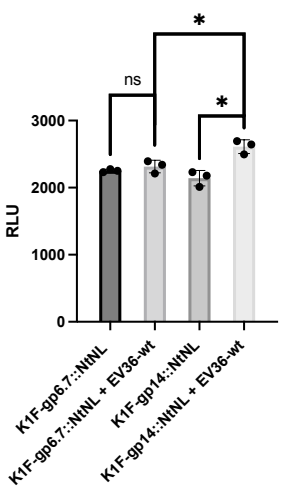
The conclusions that can be made after analysing the data presented across this entire thesis, along with suggestions for future work, are presented in the following chapter.



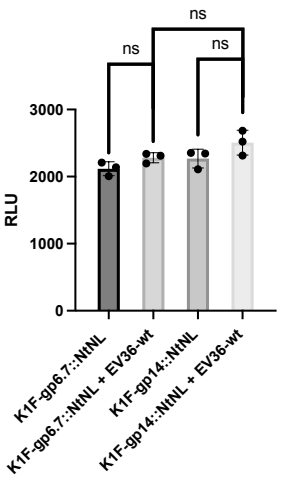
(a) Time-course signal generation with externally provided CtNL



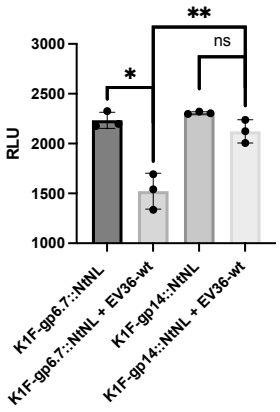
(b) 5 minute time point



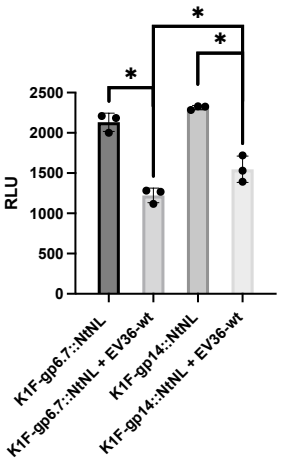
(c) 10 minute time point



(d) 15 minute time point



(e) 30 minute time point



(f) Lysis time point

Figure 5.21: **Luminescence measurements after incubating non-genomically TXTL engineered phage with EV36-wt for different time periods (third attempt).** The experiment was repeated identically to Figure 5.20, (+/- SD, n = 3).

Chapter 6

Conclusions and future directions

6.1 Summary of findings

The key results displayed in this work present two novel K1F ICP fusion constructs, a simple, non-genomic method for packaging inducibly-ejectable proteins inside phage capsids within a cell-free environment and a proof-of-concept diagnostic model where engineered phage emit signal in a host-dependent manner. Secondary results include the cell-free synthesis of phage K1F for the first time, a novel method for visually analysing and tracking phage synthesis using TXTL and EM, a proposed hypothesis for procapsid accumulation-mediated DNA packaging events and the first reported demonstration of heat-induced release of phage ICPs. The amalgamation of these results, alongside a literature review, suggest that the ICP::NtNL fusion is successfully packaged during the non-genomic phage engineering process and subsequently, it is retained within the capsid of the phage until heat- or host-induced ejection occurs - both being events which initiate the ejection of the WT ICPs, genomic DNA and any encapsulated ICP::NtNL fusion protein.

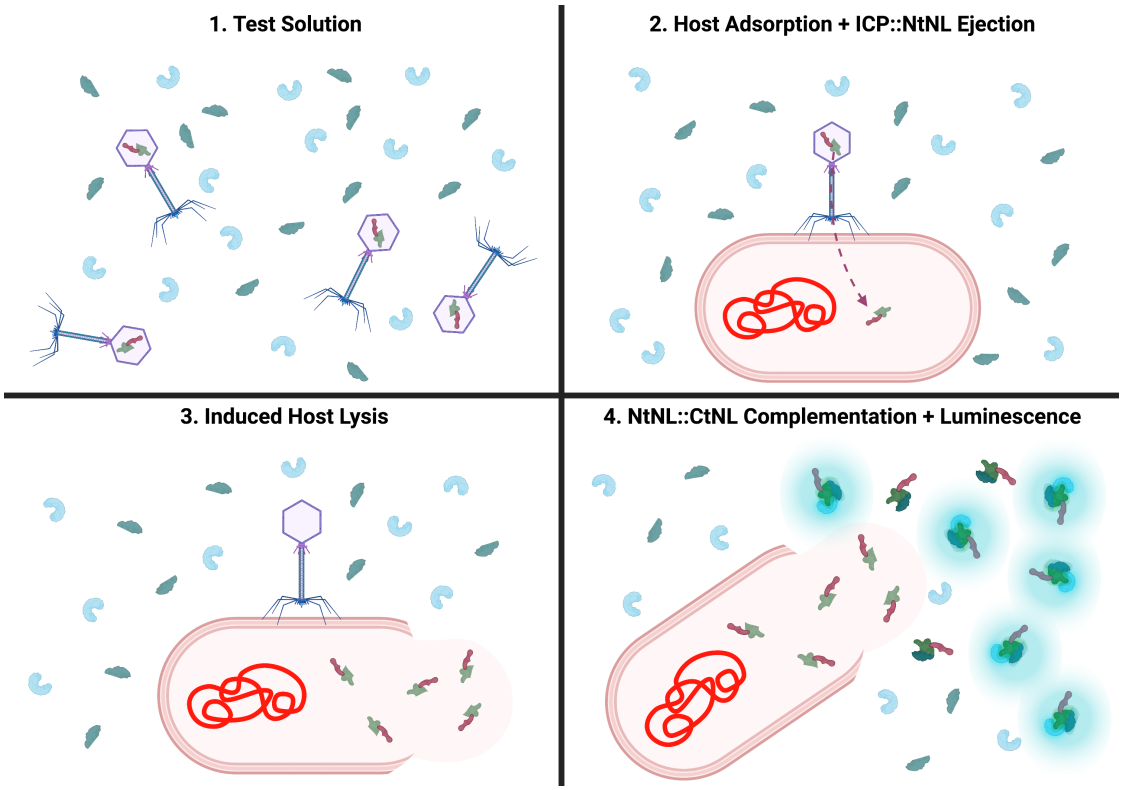


Figure 6.1: **Rapid phage-based pathogen detection model utilising a gp6.7::NtNL fusion.** The test solution consists of: 1). engineered phage (K1F*gp6.7::NtNL*), 2). furimazine substrate (light blue) and 3). CtNL (dark green). Upon the host adsorption, the gp6.7::NtNL fusion is ejected into the cytoplasm, the cells are then lysed with a lysis buffer and then the NtNL fusion is free to complement with the extracellular CtNL and subsequently emit luminescence by reacting with the furimazine substrate.

The overarching goal of this thesis was to present a novel model for the rapid phage-based detection of *E. coli* K1 which does not rely on any host processing and can, theoretically, generate a detectable signal within minutes rather than hours. Whilst the results gained are not entirely conclusive and are far from being at an acceptable level for a legitimate diagnostic test, there is still a clear indication that the rationale, construction work and hypotheses have merit and are worthy of further investigation. Furthermore, the gp14::NtNL fusion presented itself as a considerably more efficient, reliable and robust candidate for a future diagnostic system and should this work be continued and/or potentially commercialised, it is recommended that gp14 becomes the focal point. However, before discussing this in more detail, it is perhaps a fitting interlude here to present the final proposed diagnostic models for rapidly detecting *E. coli* K1 with K1F*gp6.7::NtNL* and K1F*gp14::NtNL*. These models are displayed, respectively, in Figures 6.1 and 6.2.

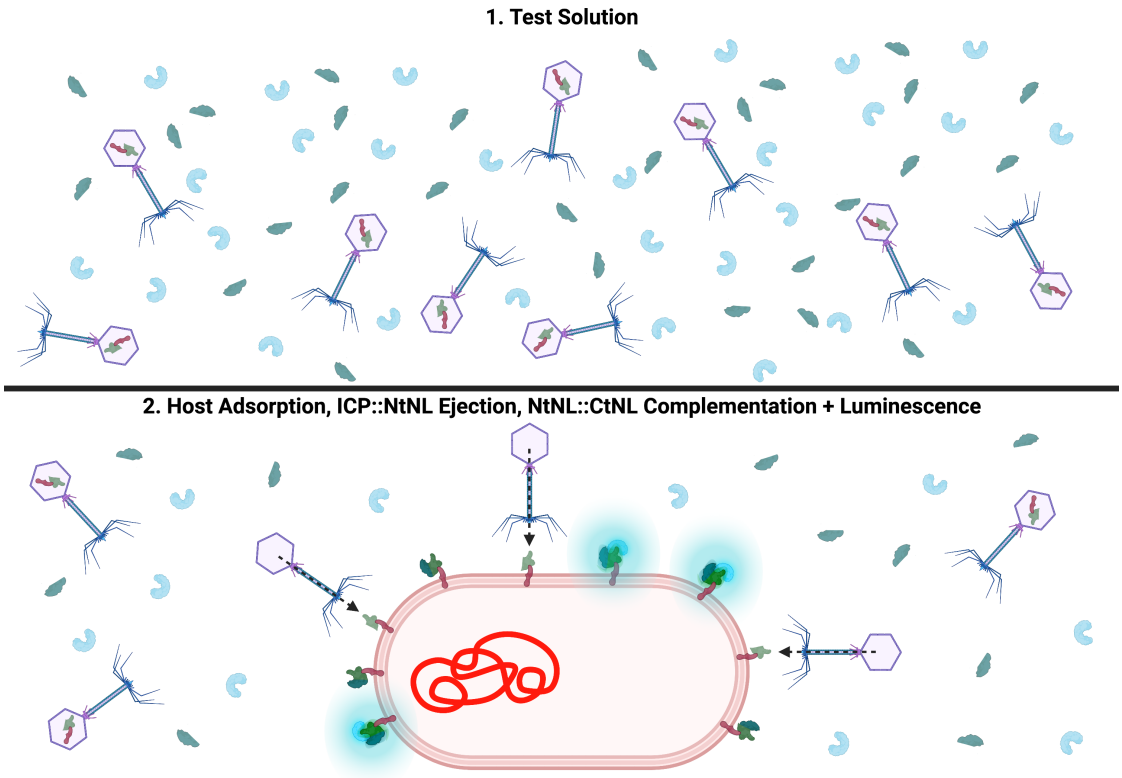


Figure 6.2: **Rapid phage-based pathogen detection model utilising a gp14::NtNL fusion.** The test solution consists of: 1). engineered phage (K1F*gp14::NtNL*), 2). furimazine substrate (light blue) and 3). CtNL (dark green). Upon the host adsorption, the gp14::NtNL fusion is ejected and lodged into the outer membrane, the NtNL domain of the fusion is free to complement with the extracellular CtNL and subsequently emit luminescence by reacting with the furimazine substrate.

6.2 Final considerations and future directions

6.2.1 Revisiting the rationale

The original calculations made at the start of Chapter 3 estimated that a minimum of 60,221,407 NanoLuc copies/mL was necessary in order to generate a detectable luminescent output. Furthermore, the conservative engineering estimations (a 10% engineering success rate - see Chapter 3 for more details) indicated that there would be 18 billion copies of gp6.7::NtNL or 10 billion copies of gp14::NtNL present within a 10^{10} PFU/mL engineered phage stock. As seen in Chapter 5, the best performing engineering method (endoTXTL non-genomic engineering) was capable of producing a 10^{9} PFU/mL yield of engineered phage. Therefore, if applying the same rationale as above, there would be 1.8 billion copies of gp6.7::NtNL or 1 billion copies of gp14::NtNL present in the stock. Finally, if estimating that just 10% of the packaged fusion proteins do successfully complement with CtNL in the diagnostic test, then the final copy number of complimented NanoLuc within the test solution would be 180 million copies of gp6.7::NtNL or 100 million copies of gp14::NtNL - both higher than the minimum viable amount of 60,221,407. Whilst these metrics are highly speculative and the endoTXTL engineered phage titer is indeed reduced after the filtration process (often by an order of magnitude - data not shown), the rationale provided does, at least, indicate that the achievement of the minimum viable NanoLuc copy number is within reason. Moreover, the experimental data obtained during the diagnostic assays and heat-induced ejection tests provide a basis to back this up.

6.2.2 Genetic engineering 2.0

Whilst the entire consortium of genetic engineering attempts (presented in Chapter 4) transpired as a collective failure, it is not expected that genomic ICP fusion engineering is impossible. Rather, it is likely that the problem is merely in need of more time, optimisation and potentially the consideration of additional alternative approaches. Furthermore, it is hypothesised that the most likely reason for the repeated engineering failure in this scenario was due to the inhibiting constraints of ICP modification outweighing any selective pressure that was attempting to force and maintain the engineered genotype.

One option to try and resolve this would be to increase the efficiency of the selective pressures, for example and as previously discussed in Chapter 4, by optimising the CRISPR system so that it effectively purges all WT DNA and therefore super-enriches the engineered population or by carrying out the HR reactions in a more favourable environment (i.e. the λ red system). However, an alternative solution, which perhaps at this stage is the most promising genetic ICP engineering avenue, would be to remove the need for a selective pressure all together by attempt-

ing *in vitro* whole genome assembly and thereby alleviating the concern of phage:host infection propensities. The *in vitro* whole genome assembly workflow has indeed been mapped out for phage T7 recently [37], where the entire genome was split into overlapping fragments and assembled with the NEBuilder HiFi DNA Assembly kit, then transformed into *E. coli* for phage "re-booting". Interestingly, during that study the researchers also used the technique to insert a NanoLuc cassette into a non-structural region of the T7 DNA so that, upon host infection and cytoplasmic processing, the luminescent signal is expressed. This confirms that the technique can be successfully used to engineer phage. Also, this process has been carried out for a smaller phage (AP205 - 4268bp genome) in a cell-free system [206], introducing new possibilities for completely host-free phage engineering *and* synthesis.

Furthermore, by removing the host entirely from the equation, the stability and evolutionary advantages of the phage become less important, which, in theory, means that genetic ICP engineering can occur without selective resistance - a huge advantage for such a disarming adaption. Moreover, even if, upon infection, the selective pressures become instantly overpowering the moment the first engineered genome reaches the host cytoplasm, in the case of the diagnostic phage, the first point of infection (i.e. the ejection of the internal capsid contents) is the only phase of infection that is relevant and therefore, it would not be detrimental if the phage proceeded to lose their engineered features after the initial phase had occurred as the signal would have already been released and become detectable. As a first future experiment, it is recommended that the K1F genome is split into overlapping fragments via PCR (e.g. six 6.6kb fragments) and within the fragments containing g6.7 and g14, the NtNL fusion sequence is added. Once assembled *in vitro*, the engineered genome should also be assembled *in vitro* in a cell-free system, then tested via heat- and host-induced signal ejection. The methods of screening that have been established during this project (i.e. PCR, gel electrophoresis, DNA sequencing and mass spectrometry) can all be used to verify the generation of the successfully engineered phage.

Finally, it may also be interesting to explore g15 and g16 engineering, as these two genes contribute the other two ejection proteins that comprise the K1F internal core and ejectosome [97], therefore, they could also be utilised within an ICP fusion-based rapid diagnostic system. However, it should be noted that, upon infection, gp15 and gp16 are likely to undergo severe unfolding/refolding due to their size [106] and when they reconfigure within the host envelope, they are positioned closer to the cytoplasm rather than the external environment [121]. Subsequently, they may be less suitable and would not benefit from the hypothetical gp14::NtNL ultra-rapid, pre-lysis NtNL:CtNL complimentation displayed in Figure 6.2. Nevertheless, it may still transpire to be a worthwhile investigation and therefore, the sequences required for pBEST-gp15/gp15::NtNL fusion expression/non-genomic TXTL engineering, pUC19-g15/g16::NtNL

HR and pCas9-g15/16 CRISPR selection are provided in Appendix B (Table B.1). Of course, *in vitro* whole genome assembly could also be used to generate K1F*g15::NtNL* and K1F*g16::NtNL* should that method become the preferred/optimal route of engineering.

6.2.3 Considering the diagnostic efficacy

The task of measuring the efficacy of the two diagnostic phage (K1F*gp6.7::NtNL* and K1F*gp14::NtNL*) presented in this thesis is a complex one due to the often inconsistent experimental design used and statistical significance found throughout the various detection attempts and iterations. To simplify matters, it is decided here to draw conclusions only from the TXTL-engineered phage, as the experimental design when using these phage did not fluctuate (i.e. they were always supplied with external CtNL rather than endogenously expressed CtNL and were compared against a representative 'phage only' control with identical conditions other than the host not being provided). Furthermore, when considering the statistical significance of the diagnostic capabilities for these phage only, a more obvious and reliable pattern can be observed (Table 6.1). The significance values used to populate Table 6.1 were taken from the three detection attempts with the TXTL-engineered phage (i.e. Figures 5.19, 5.20 and 5.21). As an example, for the $\leq$15 minute "Attempt 1" time point, the answer for K1F*gp6.7::NtNL* is not significant (i.e. there was no statistically significant difference between the RLU output from the 'phage only' and 'phage + host' samples) and the answer for K1F*gp14::NtNL* is that the 'phage + host' sample emitted a higher RLU output compared to the 'phage only' sample with statistical significance (i.e. with a P value of <0.05).

To begin with and as explained in the previous chapter, the heat-induced signal release assays for both K1F*gp6.7::NtNL* and K1F*gp14::NtNL* clearly indicated that the fusion proteins had been successfully packaged (Figure 5.16), therefore, the assumption was that both phage were entering their detection assays with diagnostic capabilities (i.e. they were able to eject their

Table 6.1: **Engineered phage K1F: diagnostic efficacy**

Detection Attempt	**Incubated for $\leq$15 mins**	**Incubated for 30 mins**	**Incubated until phage-induced lysis**
Attempt 1 (K1F*gp6.7::NtNL*)	ns	phage only = *	phage only = **
Attempt 2 (K1F*gp6.7::NtNL*)	ns	phage only = **	phage only = **
Attempt 3 (K1F*gp6.7::NtNL*)	ns	phage only = *	phage only = *
Attempt 1 (K1F*gp14::NtNL*)	phage + host = *	ns	phage only = **
Attempt 2 (K1F*gp14::NtNL*)	phage + host = *	ns	phage only = *
Attempt 3 (K1F*gp14::NtNL*)	phage + host = *	ns	phage only = *

fusion proteins in an inducible manner).

Moreover, as displayed in Table 6.1, K1F*gp6.7::NtNL* failed to produce any statistically significant results that indicate it is capable of successfully detecting the host and this has been attributed to the host's endogenous protease activity degrading the fusion before enough of it has access to the external CtNL. The likely explanation for the K1F*gp6.7::NtNL* 'phage only' sample emitting a greater RLU output with statistical significance compared to its 'phage + host' counterpart for the 30 minute and phage-induced lysis time points is the endogenous proteases being released via host lysis and subsequently degrading the external, unpackaged background fusion which decreases the baseline luminescence value. At 30 minutes, it can be expected that a small/medium amount of host lysis has occurred and subsequently released a sufficient amount of protease to act on the external fusion.

For K1F*gp14::NtNL*, the 'phage + host' sample emitted an increased luminescent output compared to the 'phage only' sample for the $\leq$15 minute time point during each of the three detection attempts (i.e. the three biological repeats). This adds a level of reliability and credibility to the diagnostic test and thereby takes a significant stride towards confirming that the K1F*gp14::NtNL* diagnostic assay is both viable and sufficiently rapid. At 30 minutes, there is a decrease in 'phage + host' output and no significant difference is found when comparing to the 'phage only' sample. This could suggest that, whilst the small amount of host lysis and subsequent protease release has begun to act on the external background gp14::NtNL fusion, a portion of the ejected fusion that is lodged within the outer membrane is still able to emit signal and maintain a certain degree of signal generation, thereby preventing the luminescent output from decreasing a significant amount (whereas for K1F*gp6.7::NtNL*, the output is significantly decreased at this time point). It is also possible that the previously reported hydrophobic regions of gp14 [119] are capable of associating the external background gp14::NtNL fusion with the host membrane structures and prevent them from being degraded by the divulging extracellular protease activity. Moreover, this is something that would not be possible for gp6.7::NtNL and may explain why it is more readily degraded at the 30 minute time point, where a portion of the host population is lysed but the majority of the host population is still viable. Furthermore, by the time phage-induced lysis occurs, the K1F*gp14::NtNL* performance is identical to K1F*gp6.7::NtNL*, suggesting the host membranes have fully broken down as a symptom of complete lysis and subsequently, the preceding method of gp14::NtNL protection/preservation is abolished.

6.2.4 Developing and commercialising an orthogonal and accessible point-of-care pathogen detection system

Following on from the conception, construction and initial testing of the diagnostic model that this thesis has aimed to provide, the next steps that are necessary to take in order to present this model as a commercially viable rapid bacterial detection tool would be to calculate and fine tune the limit of detection (LoD) and speed of detection (SoD) values. Moreover, in order to rival one of the most cutting edge phage-based solutions [198], a LoD of 10^2 CFU/mL and SoD of 3-7 hours would need to be achieved. Due to the fact that the ICP::fusion system bypasses host-TXTL and is constrained only by the time taken for the phage to adsorb to their host and eject their contents, it is expected that it has the potential to vastly outperform existing phage-based approaches (all of which take hours, not minutes) and potentially rival the fastest *E. coli* detection system that has been published (10 minutes [204]). Indeed, in this thesis it has already been proven that detection is possible within minutes, rather than hours with K1F*gp14::NtNL*. However, for the ICP::fusion system to be able to match the sub-10 CFU/mL LoD displayed in state-of-the-art *E. coli* detection systems [201–204], it is expected that significant optimisations will need to be made. Furthermore, all current sub-10 CFU/mL systems rely on either nucleic amplification or host enrichment steps, which is either not relevant for the ICP::fusion system for the former or would massively increase the SoD for the latter and defeat the point of utilising the ejection proteins.

In a very recently published paper, a NanoLuc-expressing phage-based system being used for detecting *E. coli*-contaminated water is optimised by passing the samples through a device whereby the expressed NanoLuc signal is captured and concentrated on a nitrocellulose membrane [216]. Also, research being done in collaboration with this thesis work is developing a system whereby engineered phage can be directionally immobilised onto a device (publication of this work is impending). Here, the SpyTag/SpyCatcher system [217] utilises covalent interactions to bind the capsids of K1F particles (for which g10b is fused to SpyTag) to a surface rich in SpyCatcher. Within a microfluidic device, this can position the phage so that their tail fibers are facing inwards towards a compact channel and their capsids are bound to the channel walls. This way, if combining the ICP::fusion and SpyTag/SpyCatcher technologies, host cells within a sample that is being passed through the device are more likely to be detected and captured by the immobilised ICP engineered phage compared to in an open system. Subsequently, a lower LoD can hypothetically be achieved by increasing phage:host interactions and potentially, a secondary chamber could be engineered within such a device so that any ejected signal can be concentrated/enriched prior to reading the measurement. It should be noted, though, that any additional processing will increase the SoD time and subsequently, a compromise will need to be made to arrive at a desired LoD/SoD combination. It is also worth noting that the transition to a geneti-

cally engineered phage rather than the current non-genomic alternatives will inevitably become necessary. This is to ensure that very minimal background signal is generated in the absence of the host. Moreover, at the present moment, there is a significant level of unpackaged fusion present within the engineered phage stock solutions and it is crucial that this is sufficiently removed. As filtration methods have already been attempted and subsequently failed to remove the entirety of the excess fusion, the most efficient method of achieving this will be to revert the route of engineering to genomic, thereby removing the possibility of generating any excess fusion protein (i.e. only the relevant copy number of ICP::fusion proteins will be coded for by the engineered genome).

Another point to consider is that, in its current state, the ICP::fusion system requires some form of hardware to measure the bioluminescent readout that is emitted once the host is detected. Furthermore, whilst there have been significant developments in the field to increase the accessibility of measuring luminescence, including smartphone-based detection [218] and a commercial detection kit with the user-friendly EnSURE™ Touch hardware [219], it would be preferable to migrate to a colorimetric enzymatic system whereby a readout is produced that is visible to the naked eye. It is anticipated that only a simple alteration to the fusion construct design is necessary to facilitate this (i.e. replacing the NtNL sequence with an alternative enzyme sequence), and the overall motif and engineering approach of the system can remain unchanged. For example, the aforementioned Cytochrome c Peroxidase (CcP) could be experimented with, however, the size of the protein must be considered as it would be ideal for it to be of similar size to NtNL.

Finally, to explore the orthogonality of the diagnostic model displayed in this work (i.e. the ability to precisely distinguish between different pathogens with high sensitivity and specificity), a similar engineering approach could be applied to the *Salmonella*-infecting phage P22 and Epsilon 15 - both of which have been shown to eject ICPs upon infection in a manner similar to T7 and K1F [220, 221]. In doing so, this would demonstrate that the ICP::fusion system can detect three medically relevant pathogens: *E. coli* K1 with engineered K1F, *Salmonella typhimurium* with engineered P22, and *Salmonella anatum* with engineered Epsilon 15. Moreover, in an alternative approach, the modified K1F capsid could be kept and utilised in a new engineered phage system whereby the tail fibers are repeatedly engineered to recognise different bacterial envelope polysaccharide structures, thereby increasing the target host range of the diagnostic system to several pathogens of interest. In such a system, all phage present would have identical capsids but differing tail structures, allowing for multi-pathogen detection within one test. The idea of expanding the host adsorption range of a phage via tail fiber engineering is not a novel one, however, this still remains a research area in its infancy and is perhaps something to begin investigating alongside other methods of expanding the host range of the diagnostic system.

In future work, if the orthogonality can be demonstrated and the LoD and SoD can be quantified, along with integrating the ICP::fusion technology with a device (e.g. a SpyTag/SpyCatcher microfluidics device), then this diagnostic model could have high potential to become a commercially successful pathogen detection system.

6.3 Concluding remarks

In conclusion, this thesis has made a series of significant novel contributions to the field of bioengineering and regardless of the direction that this proof-of-concept ICP::fusion system is propelled towards, it is anticipated that the key results displayed in this work have potential to play a significant role in the future of truly rapid and accessible bacterial diagnostics.

Appendix A

Nucleotide sequences used throughout this study

Table A.1: **List of plasmids used in this study**

SN	Name
1	pBEST-OR2-OR1-Pr-UTR1-deGFP-T500
2	pBEST-OR2-OR1-Pr-UTR1-gp6.7::NtNL-T500
3	pBEST-OR2-OR1-Pr-UTR1-gp14::NtNL-T500
4	pBEST-OR2-OR1-Pr-UTR1-CtNL-T500
5	pBEST-OR2-OR1-Pr-UTR1-NanoLuc-T500
6	pUC19-g6.7::NtNL
7	pUC19-g14::NtNL
8	pCas9-g6.7
9	pCas9-g14
10	pAD-LyseR

A.1 Sequence for pBEST-OR2-OR1-Pr-UTR1-deGFP-T500

TGAGCTAACACCGTGCGTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTGCAG
CTAGCAATAATTTTGTTTAACTTTAAGAACGAGATATACCATGGAGCTTTTCACTGGCGTTG
TTCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGA
GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA
GCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGC
CGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG

TCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCA
TGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGG
ACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCG
TGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGA
GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCTCTAGAGTGCAC
CACCACCACCATCACGTGTAACTCGAGCAAAGCCCGCCGAAAGGCGGGCTTTTCTGTGTCG
ACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACT
ATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGC
GCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTA
TCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAG
AACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCG
TTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCG
CTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG
TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAG
CTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATC
GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG
GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTAC
GGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAA
AAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGT
TTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCT
ACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTAT
CAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAG
TATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCA
GCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGAT
ACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCG
GCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCT
GCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTC
GCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGT
CGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCC
ATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGG
CCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAACC
GTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC
GGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAA
CTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACC
GCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTA
CTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAA
TAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATT
TATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAA
TAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTAT

CATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTGG
CGAATCCTCTGACCAGCCAGAAAACGACCTTTCTGTGGTGAAACCGGATGCTGCAATTCAG
AGCGCCAGCAAGTGGGGGACAGCAGAAGACCTGACCGCCGCAGAGTGGATGTTTGACATG
GTGAAGACTATCGCACCATCAGCCAGAAAACCGAATTTTGCTGGGTGGGCTAACGATATCC
GCCTGATGCGTGAACGTGACGGACGTAACCACCGCGACATGTGTGTGCTGTTCCGCTGGGC
ATGC

A.2 Sequence for pBEST-OR2-OR1-Pr-UTR1-gp6.7::NtNL-T500

TGAGCTAACACCGTGCGTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTGCAG
CTAGCAATAATTTTGTTTAACTTTAAGAACGAGATATACCATGGGCTGTTTCAGTCCGAAG
ATTAGCACTCCGAAGCCTTCGGTCCAAGCACCTGAACCAGCACCTCTGAGTGAGGAAGTTG
CGTCAGTTGACATCGGGGCTGAATCGGATGTGGACACCAATGAGACCAAAGGTATCAAAG
ACCTTAAGGTCAAGAAGGAGTCTGCACCTAAAGATAAATCGTCAGTTAGCCGCGCCATGCG
AGCCTCTGGCGTCAACATGGGGGGTGGTGGTGGTTCTGGTGGTGGTGGTTCTGTCTTCACA
CTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGACCAAGTCCTTG
AACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAG
GATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAA
TAACTCGAGCAAAGCCCGCCGAAAGGCGGGCTTTTCTGTGTCGACCGATGCCCTTGAGAGC
CTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATG
ACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTTCCGCTTCCTCGC
TCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGC
GGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGG
CCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGC
CCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCT
GCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCT
CACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGA
ACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGT
ATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAAC
AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTAC
GCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCT
AGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTG
GTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTT
CATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATC
TGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCA
ATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCC
ATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC
GCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCA

TTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAG
CGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTC
ATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAACCGTAAGATGCTTTTCTGT
GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCT
TGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCA
TTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTC
GATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAA
TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTC
ATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACAT
TTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAA
AAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTGGCGAATCCTCTGACCAGCC
AGAAAACGACCTTTCTGTGGTGAAACCGGATGCTGCAATTCAGAGCGCCAGCAAGTGGGG
GACAGCAGAAGACCTGACCGCCGCAGAGTGGATGTTTGACATGGTGAAGACTATCGCACC
ATCAGCCAGAAAACCGAATTTTGCTGGGTGGGCTAACGATATCCGCCTGATGCGTGAACGT
GACGGACGTAACCACCGCGACATGTGTGTGCTGTTCCGCTGGGCATGC

A.3 Sequence for pBEST-OR2-OR1-Pr-UTR1-gp14::NtNL-T500

TGAGCTAACACCGTGCGTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTGCAG
CTAGCAATAATTTTGTTTAACTTTAAGAACGAGATATACCATGGGCTGCGAGCCAGTAAGT
ATCGGTCTAGGAATCATGGCTGTAGCAGGGGCCACTATGTCCGCATCTCAACAGGCCAAAG
CTGAGGGTGCTGCTATCGACGCTCAGAACCGACAGGCTCAGGAGATGGTTAAGCAGATGA
ATTACTCTGACGCCAACCTAAGGATGCAGGAGCGAGACCTTAAGGAACAGCAGATGGCTG
AACTGACAGAGACCACGTTAAACGGTATCCGCAATCAGGGCATGGTACGAGCTGCGGTAG
CTGAGTCCGGTCTGGAAGGAAACTCTATGGACAGGATTGAACGTCAGGTAGAGGGAGATA
CAGTCAAGGAGAGAGCAGGGATTACCGAAAGTTACAACCGCGACTATGCGGCTATCTTTG
GGAACCGTATCGCCAACATTGAGAACACCAAGTCTGCTATCCGTGGTCAAGGTAAAATCAT
CAAGACTAGCCCACTGGCTCATGCACTTAATGTTGCTAACGCCGGTATGCAGGGATACGCT
GCTGGTAAGTCAATCTCTGGGGCATCAAGCTCTGGTGGTTCTGCACCGATTAGTGCTGCTA
AAGGCACACCTACAGGTCATAGCGGTGGTGGTGGTTCTGGTGGTGGTGGTTCTGTCTTCAC
ACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGACCAAGTCCTT
GAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAA
GGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGA
ATAACTCGAGCAAAGCCCGCCGAAAGGCGGGCTTTTCTGTGTCGACCGATGCCCTTGAGAG
CCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTAT
GACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTTCCGCTTCCTCG
CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGG
CGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAG
GCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCG
CCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGG
ACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCC

TGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGC
TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACG
AACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCG
GTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAA
CAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC
TTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATT
ACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTC
AGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCAC
CTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGT
TCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGC
AATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTC
CATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTC
ATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAA
GCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACT
CATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAACCGTAAGATGCTTTTCTG
TGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTC
TTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATC
ATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTT
CGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCT
GGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAA
ATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCT
CATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACA
TTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATA
AAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTGGCGAATCCTCTGACCAGC
CAGAAAACGACCTTTCTGTGGTGAAACCGGATGCTGCAATTCAGAGCGCCAGCAAGTGGG
GGACAGCAGAAGACCTGACCGCCGCAGAGTGGATGTTTGACATGGTGAAGACTATCGCAC
CATCAGCCAGAAAACCGAATTTTGCTGGGTGGGCTAACGATATCCGCCTGATGCGTGAACG
TGACGGACGTAACCACCGCGACATGTGTGTGCTGTTCCGCTGGGCATGC

A.4 Sequence for pBEST-OR2-OR1-Pr-UTR1-CtNL-T500

TGAGCTAACACCGTGCGTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTGCAG
CTAGCAATAATTTTGTTTAACTTTAAGAACGAGATATACCATGGGTCTGAGCGGCGACCAA
ATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGG
TGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGATCGACTATTT
CGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGAC
CCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTG
TTCCGAGTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGTAGT

AGCTCGAGCAAAGCCCGCCGAAAGGCGGGCTTTTCTGTGTCGACCGATGCCCTTGAGAGCC
TTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGA
CTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG
GTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGC
CAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCC
CCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC
TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTG
CCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTC
ACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAA
CCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGT
AAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTA
TGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACA
GTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTG
ATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACG
CGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGT
GGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTA
GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGG
TCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTC
ATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCT
GGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAA
TAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCAT
CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGC
AACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATT
CAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCG
GTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCAT
GGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAACCGTAAGATGCTTTTCTGTGA
CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTG
CCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATT
GGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGA
TGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGG
TGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATG
TTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCAT
GAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTT
CCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAA
ATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTGGCGAATCCTCTGACCAGCCAG
AAAACGACCTTTCTGTGGTGAAACCGGATGCTGCAATTCAGAGCGCCAGCAAGTGGGGGA
CAGCAGAAGACCTGACCGCCGCAGAGTGGATGTTTGACATGGTGAAGACTATCGCACCAT
CAGCCAGAAAACCGAATTTTGCTGGGTGGGCTAACGATATCCGCCTGATGCGTGAACGTGA
CGGACGTAACCACCGCGACATGTGTGTGCTGTTCCGCTGGGCATGC

A.5 Sequence for pBEST-OR2-OR1-Pr-UTR1-NanoLuc-T500

TGAGCTAACACCGTGCGTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTGCAG
CTAGCAATAATTTTGTTTAACTTTAAGAACGAGATATACCATGGTCTTCACACTCGAAGATT
TCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGACCAAGTCCTTGAACAGGGAG
GTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTG
AGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCG
GCGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCA
CTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGATC
GACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAA
CAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCT
CCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCT
GGCGTAGTAGCTCGAGCAAAGCCCGCCGAAAGGCGGGCTTTTCTGTGTCGACCGATGCCCT
TGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGC
ACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTTCCGCT
TCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTC
AAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGC
AAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAG
GCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCC
GACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTT
CCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTC
TCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTG
TGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCC
AACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGA
GCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGG
TAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAG
CAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTG
ACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGA
TCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAG
TAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCT
ATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGC
TTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATT
TATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTAT
CCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAA
TAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTA
TGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGC
AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGT
TATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAACCGTAAGATGC
TTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGA
GTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGA

TCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG
CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGA
CACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGT
TATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC
CGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATT
AACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTGGCGAATCCTC
TGACCAGCCAGAAAACGACCTTTCTGTGGTGAAACCGGATGCTGCAATTCAGAGCGCCAGC
AAGTGGGGGACAGCAGAAGACCTGACCGCCGCAGAGTGGATGTTTGACATGGTGAAGACT
ATCGCACCATCAGCCAGAAAACCGAATTTTGCTGGGTGGGCTAACGATATCCGCCTGATGC
GTGAACGTGACGGACGTAACCACCGCGACATGTGTGTGCTGTTCCGCTGGGCATGC

A.6 Sequence for plasmid pUC19-g6.7::NtNL

CCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTT
CGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGC
CAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCG
GGGATCCTCTAGAGGAAGTTGCGTCAGTTGACATCGGGGCTGAATCGGATGTGGACACCA
ATGAGACCAAAGGTATCAAAGACCTTAAGGTCAAGAAGGAGTCTGCACCTAAAGATAAAT
CGTCAGTTAGCCGCGCCATGCGAGCCTCTGGCGTCAACATGGGCGGTGGTGGTGGTTCTGG
TGGTGGTGGTTCTGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGC
TACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGG
TGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACAT
CCATGTCATCATCCCGTATGAATAAGACAATGCTACCATATCTCAACTCACGCGAAGGTCG
CCACATGTGCGCTTGTCGCCTCTGGGAAGACGGGCAGTCTAACTTCAAGTCATTTGAGGAC
TTCAAGGCTCATACTTACCGTATGGCTGACGAGTTCGACGGTGAAGAATATACTAGTAGCG
GCCGCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTG
TTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGT
GCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGG
AAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCG
TATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGC
GAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACG
CAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCG
TTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAA
GTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTT
CGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTT
CGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCG
GTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCAC
TGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTG
GCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTT
ACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTG
GTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTT

GATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTC
ATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAAT
CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGC
ACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGA
TAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAG
AAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAG
TAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTG
TCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGA
AGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGT
CATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAA
TAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCAC
ATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAG
GATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG
CATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAA
AAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTAT
TGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
ATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAAC
CATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGC
GTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTG
TCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGG
GTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATG
CGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCG

A.7 Sequence for pUC19-g14::NtNL

CCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTT
CGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGC
CAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCG
GGGATCCTCTAGAGATCATCAAGACTAGCCCACTGGCTCATGCACTTAATGTTGCTAACGC
CGGTATGCAGGGATACGCTGCTGGTAAGTCAATCTCTGGGGCATCAAGCTCTGGTGGTTCT
GCACCGATTAGTGCTGCTAAAGGCACACCTACAGGTCATAGCGGTGGTGGTGGTTCTGGTG
GTGGTGGTTCTGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTA
CAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTG
TCCGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCC
ATGTCATCATCCCGTATGAATAAGAGGAGGACTAATGGCTAGTAATATTGAATCAGCTCTG
GCTAATCGGACTATGGGTCGTGGCAGAGCGCCGGGTAAAACTATCGCCGTCAACTATCAAG
CAGCCAGCGTTCAGGCTCCAACTGGTGACTCAGGTCTGGCTCGGGCGTTACTAGTAGCGGC
CGCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTA
TCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGC
CTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGA

AACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTA
TTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA
GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCA
GGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCC
CTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCG
GGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCG
CTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGT
AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGC
CTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGG
TTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTG
ATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCA
TGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATC
AATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCA
CCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT
AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCA
CGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGA
AGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGT
AAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTG
TCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGA
AGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGT
CATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAA
TAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCAC
ATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAG
GATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG
CATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAA
AAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTAT
TGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
ATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAAC
CATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGC
GTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTG
TCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGG
GTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATG
CGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCG

A.8 Sequence for plasmid pCas9-g6.7

TTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGT
GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGT
GCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTC
TCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCA
CGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACC
CGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTA
GTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTC
CTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCC
CTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAA
GAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAATTTATCTCTTCAAA
TGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTTGACAGCTTATC
ATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCGT
GTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTAGG
CATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATTACGAAATCATCCTGTGG
AGCTTAGTAGGTTTAGCAAGATGGCAGCGCCTAAATGTAGAATGATAAAAGGATTAAGAG
ATTAATTTCCCTAAAAATGATAAAACAAGCGTTTTGAAAGCGCTTGTTTTTTTGGTTTGCAG
TCAGAGTAGAATAGAAGTATCAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATA
ACGGACTAGCCTTATTTTAACTTGCTATGCTGTTTTGAATGGTTCCAACAAGATTATTTTAT
AACTTTTATAACAAATAATCAAGGAGAAATTCAAAGAAATTTATCAGCCATAAAACAATAC
TTAATACTATAGAATGATAACAAAATAAACTACTTTTTAAAAGAATTTTGTGTTATAATCTA
TTTATTATTAAGTATTGGGTAATATTTTTTGAAGAGATATTTTGAAAAAGAAAAATTAAAG
CATATTAAACTAATTTCGGAGGTCATTAAAACTATTATTGAAATCATCAAACTCATTATGG
ATTTAATTTAAACTTTTTATTTTAGGAGGCAAAAATGGATAAGAAATACTCAATAGGCTTA
GATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTA
AAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGG
CTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAG
AAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATG
GCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACA
AGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAA
ATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTG
CGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGG
AGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTAC
AATCAATTATTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTT
CTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAA
GAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAAT
CAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGA
TTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGA
ATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCT
CCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTTTAAA
AGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAA

AACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCA
AACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAG
ATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGT
GAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTG
AGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGC
AATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTG
AAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGA
TAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACG
GTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTC
TTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAAC
CGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATT
TCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTA
TTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTT
AACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCTCAC
CTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTT
TTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGA
CATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAAC
ATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGT
TGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATG
GCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGA
ATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAAT
ACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATG
TGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACA
AAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGT
GGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGA
CAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAAC
GTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCG
CCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAA
AATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTT
CCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGAT
GCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGG
AGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAA
GAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAA
CAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGG
AAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTC
CATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGA
GTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCA
AAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGG
TGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTA
TGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGA
AGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGT

CGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCA
AGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAG
AAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTAT
TGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTT
CTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATT
CATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAAT
TGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCC
ATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGAAGTATAT
TTTAGATGAAGATTATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATAAATGCAGT
AATACAGGGGCTTTTCAAGACTGAAGTCTAGCTGAGACAAATAGTGCGATTACGAAATTTT
TTAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGAG
CCTCTGGCGTCAACATGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCACA
CTGAGACTTGTTGAGTTCCATGTTTTAGAGCTATGCTGTTTTGAATGGACTCCATTCAACAT
TGCCGATGATAACTTGAGAAAGAGGGTTAATACCAGCAGTCGGATACCTTCCTATTCTTTC
TGTTAAAGCGTTTTCATGTTATAATAGGCAAAAGAAGAGTAGTGTGATCGTCCATTCCGAC
AGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCTATGCGC
ACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTAC
TTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGATCCTCTACGC
CGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCC
GACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCG
TGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACC
ATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAG
GAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCT
TCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAA
CTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGA
GCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGC
CTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATG
GCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCC
CCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCC
AGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGC
CTAACTTCGATCATTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACAT
GGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCG
TCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAAC
GGATTCACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAA
CCAACCCTTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCAT
CTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACC
CGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGA
ACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCTTCG
GTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCCCCTACGTGCTGCTGAAGTTGCCC
GCAACAGAGAGTGGAACCAACCGGTGATACCACGATACTATGACTGAGAGTCAACGCCAT
GAGCGGCCTCATTTCTTATTCTGAGTTACAACAGTCCGCACCGCTGTCCGGTAGCTCCTTCC
GGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTC

GTAGGACAGGTGCCGGCAGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCC
ACGCCGAAACAAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGC
TACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTAACCGTTTTTATCAGGCTCTGGG
AGGCAGAATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTG
GCCTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAAC
CAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATT
TGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCACCAGGCGT
TTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC
TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACCT
GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAAC
GGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAG
GGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTT
CACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTA
TTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGA
ACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATT
CATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACG
GTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTG
ACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCC
AGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATA
CGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACG
TCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTA
TTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGAT
GCTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTT
CTATCAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAAGCACCGC
CGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCA
GTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTG
ATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGC
GAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTA
ACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTT

A.9 Sequence for plasmid pCas9-g14

TTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGT
GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGT
GCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTC
TCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCA
CGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACC
CGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTA
GTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTC
CTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCC
CTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAA
GAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAATTTATCTCTTCAAA

TGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTTGACAGCTTATC
ATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCGT
GTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTAGG
CATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATTACGAAATCATCCTGTGG
AGCTTAGTAGGTTTAGCAAGATGGCAGCGCCTAAATGTAGAATGATAAAAGGATTAAGAG
ATTAATTTCCCTAAAAATGATAAAACAAGCGTTTTGAAAGCGCTTGTTTTTTTGGTTTGCAG
TCAGAGTAGAATAGAAGTATCAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATA
ACGGACTAGCCTTATTTTAACTTGCTATGCTGTTTTGAATGGTTCCAACAAGATTATTTTAT
AACTTTTATAACAAATAATCAAGGAGAAATTCAAAGAAATTTATCAGCCATAAAACAATAC
TTAATACTATAGAATGATAACAAAATAAACTACTTTTTAAAAGAATTTTGTGTTATAATCTA
TTTATTATTAAGTATTGGGTAATATTTTTTGAAGAGATATTTTGAAAAAGAAAAATTAAAG
CATATTAAACTAATTTCGGAGGTCATTAAAACTATTATTGAAATCATCAAACTCATTATGG
ATTTAATTTAAACTTTTTATTTTAGGAGGCAAAAATGGATAAGAAATACTCAATAGGCTTA
GATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTA
AAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGG
CTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAG
AAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATG
GCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACA
AGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAA
ATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTG
CGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGG
AGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTAC
AATCAATTATTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTT
CTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAA
GAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAAT
CAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGA
TTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGA
ATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCT
CCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTTTAAA
AGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAA
AACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCA
AACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAG
ATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGT
GAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTG
AGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGC
AATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTG
AAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGA
TAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACG
GTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTC
TTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAAC
CGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATT
TCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTA

TTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTT
AACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCTCAC
CTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTT
TTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGA
CATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAAC
ATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGT
TGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATG
GCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGA
ATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAAT
ACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATG
TGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACA
AAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGT
GGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGA
CAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAAC
GTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCG
CCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAA
AATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTT
CCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGAT
GCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGG
AGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAA
GAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAA
CAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGG
AAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTC
CATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGA
GTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCA
AAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGG
TGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTA
TGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGA
AGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGT
CGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCA
AGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAG
AAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTAT
TGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTT
CTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATT
CATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAAT
TGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCC
ATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGAAGTATAT
TTTAGATGAAGATTATTTCTTAATAACTAAAAATATGGTATAATACTCTTAATAAATGCAGT
AATACAGGGGCTTTTCAAGACTGAAGTCTAGCTGAGACAAATAGTGCGATTACGAAATTTT
TTAGACAAAAATAGTCTACGAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACACC
TACAGGTCATAGCTAAGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTCAGCAC

ACTGAGACTTGTTGAGTTCCATGTTTTAGAGCTATGCTGTTTTGAATGGACTCCATTCAACA
TTGCCGATGATAACTTGAGAAAGAGGGTTAATACCAGCAGTCGGATACCTTCCTATTCTTT
CTGTTAAAGCGTTTTCATGTTATAATAGGCAAAAGAAGAGTAGTGTGATCGTCCATTCCGA
CAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCTATGCG
CACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTA
CTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGATCCTCTACG
CCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGC
CGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGC
GTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCAC
CATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCA
GGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCC
TTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCA
ACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGG
AGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAG
CCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCAT
GGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTC
CCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTC
CAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAG
CCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACA
TGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGC
GTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAA
CGGATTCACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAA
ACCAACCCTTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCA
TCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGAC
CCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGA
ACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCTTCG
GTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCCCCTACGTGCTGCTGAAGTTGCCC
GCAACAGAGAGTGGAACCAACCGGTGATACCACGATACTATGACTGAGAGTCAACGCCAT
GAGCGGCCTCATTTCTTATTCTGAGTTACAACAGTCCGCACCGCTGTCCGGTAGCTCCTTCC
GGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTC
GTAGGACAGGTGCCGGCAGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCC
ACGCCGAAACAAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGC
TACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTAACCGTTTTTATCAGGCTCTGGG
AGGCAGAATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTG
GCCTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAAC
CAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATT
TGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCACCAGGCGT
TTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC
TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACCT
GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAAC
GGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAG
GGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTT

CACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTA
TTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGA
ACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATT
CATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACG
GTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTG
ACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCC
AGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATA
CGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACG
TCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTA
TTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGAT
GCTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTT
CTATCAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAAGCACCGC
CGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCA
GTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTG
ATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGC
GAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTA
ACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTT

A.10 Sequence for pAD-LyseR

ATCGATAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTT
CTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAA
TATTGAAAAAGGAAGAGTATGGTAGAAATCAATAATCAACGTAAGGCGTTCCTCGATATG
CTGGCGTGGTCGGAGGGAACTGATAACGGACGTCAGAAAACCAGAAATCATGGTTATGAC
GTCATTGTAGGCGGAGAGCTATTTACTGATTACTCCGATCACCCTCGCAAACTTGTCACGCT
AAACCCAAAACTCAAATCAACAGGCGCCGGACGCTACCAGCTTCTTTCCCGTTGGTGGGAT
GCCTACCGCAAGCAGCTTGGCCTGAAAGACTTCTCTCCGAAAAGTCAGGACGCTGTGGCAT
TGCAGCAGATTAAGGAGCGTGGCGCTTTACCTATGATTGATCGTGGTGATATCCGTCAGGC
AATCGACCGTTGCAGCAATATCTGGGCTTCACTGCCGGGCGCTGGTTATGGTCAGTTCGAG
CATAAGGCTGACAGCCTGATTGCAAAATTCAAAGAAGCGGGCGGAACGGTCAGAGAGATT
GATGTATGATAACTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCA
GAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCA
CCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTC
CCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGAC
TGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCC
GGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCC
ATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTT
CTACAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAAT
AACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCG
TGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCT
GGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGA
TCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGC

ACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAAC
TCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAA
GCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGAT
AACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTT
TGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGC
CATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAA
ACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAG
GCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTG
ATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGG
TAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGA
AATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAG
TTTACTCATATATACTTTAGATTGATTTACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGG
TGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTC
GCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGG
GCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTTGG
GTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAG
TCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGG
CTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGA
TTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTAAAAGGATCT
AGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCAC
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGT
AATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAA
GAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTG
TCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATAC
CTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGG
GTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTC
GTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGA
GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGG
CAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTA
TAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGG
GGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGC
CTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAG
CGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCA
CACCGCATAGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG
GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATG
TGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGAAGGAGATGGCGCCCAAC
AGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCG
AAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCAC
CTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGAGCTGCTCATGTTTGAC
AGCTTATC

Table A.2: **List of oligos used in this study**

Name	**Sequence (5' - 3')**
pBESTseq-FWD	GTGAAGACTATCGCACCATCAG
pBESTseq-REV	GATAAAGAAGACAGTCATAAGTGC
pUC19seq-FWD	GCGGCATCAGAGCAGATTGTAC
pUC19seq-REV	CAGCTATGACCATGATTACGC
pCas9seq-FWD	CAGCTAGGAGGTGACTGAAG
pCas9seq-REV	GGACGATCACACTACTCTTC
g6.7-CRISPRa	AAACGAGCCTCTGGCGTCAACATGG
g6.7-CRISPRb	AAAACCATGTTGACGCCAGAGGCTC
g14-CRISPRa	AAACACCTACAGGTCATAGCTAAGG
g14-CRISPRb	AAAACCTTAGCTATGACCTGTAGGT
K1Fg6.7-FWD	CCTTAAGGTCAAGAAGGAGTCTGC
K1Fg7.3-REV	GTCATCAGAGCCAAGACCAAC
K1Fg14-FWD	GCATCAAGCTCTGGTGGTTCTG
K1Fg15-FWD	GCTGGAAGAACTCGTCCGTAG
NtNL-FWD	GATCGACATCCATGTCATCATCC
K1Fg8-REV	CACACGGACGATAGGTAACTGAAGC
K1Fg16-REV	CGCCAGACTCTGGACTATCGAAAG
amp-FWD	CGTCGTGTAGATAACTACGATACG
amp-REV	GGTTACATCGAACTGGATCTC

Appendix B

Nucleotide sequences for the suggested future work

Table B.1: **List of nucleotides for attempting g15 and g16 engineering**

SN	Name
1	gp15::NtNL pBEST insert
2	gp16::NtNL pBEST insert
3	g15::NtNL pUC19 HR cassette
4	g16::NtNL pUC19 HR cassette
5	g15-CRISPRa
6	g15-CRISPRb
7	g16-CRISPRa
8	g16-CRISPRb

B.1 Sequence for gp15::NtNL pBEST insert

CCATGGCTAGTAATATTGAATCAGCTCTGGCTAATCGGACTATGGGTCGTGGCAGA
GCGCCGGGTAAAACTATCGCCGTCAACTATCAAGCAGCCAGCGTTCAGGCTCCAACTGGTG
ACTCAGGTCTGGCTCGGGCGTTAACCAACTTCGTTGAGTCTGGTACTGGATTATACAAGCA
GTTCAAAGACGAGGAGAAGACACGGGCCGACGAGCGGTCTAACGAGATTATCCGTAAGCT
GACACCTCAGCAAAGACGTGAGGCTATCCAGAACGGCACATTGCTGTATCAGGATGACCCT
TACGCTATGGAAGCACTTCGAGTCAAGACTGGTCGTAATGCTGCCTTTGCTGTAGACGACG
AGATTAACGTTAAGATTCAGAACGGTGAGTTCCGTACACGTCAGGACATGGAAGAGTATC
GCCACCAGCGACTTCAGGACGCCGCTAAGTCCTATGCTGAAGAGGCGGGTATTAACCCTAC
GGACGAGTTCTTCCAGCGTGGGTTCAACGATAACATCACAGACCGCAACATCGCTATCTAT

GGGTCTTTCAATAAGTATTTCAGCAAGCAGTCTGAAGAGACAGCAATGTTGAACACCCGTA
TTGAGATGAACTCATTCCTGAACGATGGCGACCTGATGCGTTCACCTGAGTCTGGCAAGAC
ATTCATGGCCTACCTTCGTGACGGACTGACTACTGCTGCTATCCCCTCGGACCAGCGAGCG
CGGGAGGTCATTACCCAGACGGTCCGTGACGCAATCCAGAAGTCAGGCGGCTCAAACTTC
CTACAGCAAGTACGAGGCGAGCGAATCACCCTTAACGGCGTGGATGCTACAGTCGAAGAG
ATTGTAGGACCTGATATCTTCAACGCTGCTATTGTTGAGGCACAAGGAACTGAGTACAAGC
TGGTGGCTAAGTATCAGGAAGACTTAGCGTTAGGCGTTCAGTCTGCGATTCTTCAGGACGA
CCCAACCATCGGTCTGGCCCAGATTCAAAAACTCAAGGAGCAGAACAACCTGCTTCAACCG
GGTGAAGAACTCACACCTCAGCGTCAGATGCTTATTAATGCCGAAGCCAGCTTACTGGAAG
CGGTAAAGCGTAAGTCTGCTGAACAGGCGAAGGAGAACACGAAGTTAATCCAGACCCAGA
ACAAGCAACTGGTCATTGACCAAGTGTATCAGCGACGTCTGGCTGGAGACAACGTGTCCAC
CAACTATGAGGACCTTCCGGTCTCTGAAGCCACAGGAGAGTTCAAGCGTTCAGACATGAAC
AACTATGCGTCTTCTAAGCTACAGCAGATTGACCAGATGGACATACCTGAGGCTGCTAAGG
ACGCCCAGAAGGTGGCATTGTTAAGAGCTGACACTAACAACGGTCCGTTCCGTAATGCCTT
CCAGACGCTTACTCAGGACGCTGCTGGTGAGTGGCAAGCTGCTGTCATCCGTGGACAGTAC
GACCCAGACAAGATGAAACGCTTCGAGTCTCTTCGTCGTGCCTACACTCAGGACCCTTCAA
GTTTCGCTGCTCTGTATCCTGACCAAGCTCAGCTGTTCTCTACGTTCGACCAGATGGACAAG
ATGGGTCTGGACCCTCAGACGATGATTGAAGCTGATAAGCAAGCTGCAAGTCAAAGCCGT
GAGATGCGCATGGAGTCAGACAAGGCGTGGCAGGAGTTGAAGAACGACTCTCGGAATAAG
GACCTTTCGCGTCTTCCTACGTCTCTGGACTCAAGTGCTCGTAAGGTCTGGGACTCATGGTA
CTATCGTACAGGTAACGCTGACGCTGCAACTCAGCAGACTCAACGCTGGCTGAATGAGAA
CACCGTAACGTTCCAGTCTGAGGGTTCTGATGGTAAGTCCATCGGCATGGTGTCCAAACAC
CAGCTTATGGTCGGGGATAACCCAGAGTCGTGGCAGGTGGGTCGAGACATTATAGACACA
GCTCGTCAGCAGCTCATTAAGGCCAACCCTTGGGTAGTGAACTCTCAGTTGTCCGTTGTTGA
ACAGAACGGCTCTATCTTCCTCCAAGACGCTACGGGGACTATTCGTATTCGATACGATAAA
GAACTTGTAGGTAAACTCTACCGCGAACAACAGCAGAAGGCACAAGATGCCGCATATGCT
CAGGCAGAACGTGACGCTAACAAGCGAGCGCGTATCGTCGGGACTAAAGCTGCTGGTGAC
AAACGTCGAGCTGACCGAGAGGCCAACATCGAGAAGCGCGGTGGGATGTACAATGACGTC
TCACTGGAGGGTATCGCAAATGCACTAATCGGTAAGGAGGGTGGTGGTGGTTCTGGTGGTG
GTGGTTCTGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAA
CCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCC
GTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATG
TCATCATCCCGTATGAATAACTCGAG

B.2 Sequence for gp16::NtNL pBEST insert

CCATGGCGACTCGTGGTATTCGCAATAACAATCCGGGGAACATCCGGGTAAGTAAG
GACCAATGGGAAGGAATGACTGGAGATGATGGCGCATTTGTCACTTTCGATAGTCCAGAGT
CTGGCGTCCGAGCTTTAGGTAAAAACCTGCTGTCCTACGGTCGCCAAGGTTATGACTCCAT
CGAGAAGATCATCAACCGATGGGCGCCTCCTAATGAGAATGACACTCAGGCTTATATTGAC
TCAGTGGTGGCTGCGACTGGTATCCCAGCTACCCAGAGTCTAGACCTCTCGAACCCGGACA
CCCTGTCTGCTCTGGCGCAAGCTATCAGCTTCCATGAGACAGGTTCCCGGTACAACCCTGA

AGTATACCAACAGGGGGTCGCAAGAGCACTCAACGGTATCAGCCCAAAGACTCCACCAGT
AAGCGCTAACGTATTTGACGCACTCACGGAAGGACTCAAGGCTAAACCTAAAGTAGCTCTG
GGTGAGAACCTTCCGACCGCTGCTGGTCTGAACATTGAGGGTCAAGCACCTGAAGCTCCCA
ACGAATCGTTCGGTGAGATGTTCTATAAGGCTACTGGAGAGACCATGCAGGAACGAGAGG
ATCGCTCTACGTGGTTCGGTTTCGGTACGGCTACAGAAGCTGAAGTGAAGAACTCTATGGT
CGGCGTGGCTATCCGCGCTGGTCAGACTGAGGACTCACTGGATGTCATTGGCGATGTGTTC
AACCCAACCCGATGGAATAACCATAAGTGGTCTCGTGAGGAGCTGGACCAGATTCGTAAC
GCTGGGGTTCTGCCTCAGTATTACGGGGTCATTACTGGTGGCTCCCCTCAGAACCTGACCG
AGCTTATTAACTTGGCGCTTGAGAACCAGAAGTTTGACCAAGAGAAGGCCAAGGCCGGGA
CTGGCGCTCAACTGGCGGCTGGTGTGATTGGTGCTGGCGTGGACCCTCTGACCTACGTTCCT
ATTGCTGGGCAGGTAGGTAAAGGTGGGAAGCTAGTCAACAAGATGTTCACTGTGGCTGCTC
AGTCTGGTGCTCTGGCTGGGGCATCCGAGATTGCTCGAACCTCAGTTGCTGGTGGGGACGC
TCATGTGGCTGAGGCAATTATGGGCGGTGCTCTCTTCGGTGGGGGAATGACAGCTATCGGG
GACGCTATCGGCAAGGCCTTAGGCAAGTCAACCAACGAGTTCGCCGGGCCAGCCACACGT
CTGGAAGCCCGTGAGACTGCACGTAATGTTGATGGTCAGGACCTGTCGCGTATGCCAATCC
GTGAAGGAGAGGAGACCTTTAGTCACCAAGGCGTTAAGTTCGCTGACGTGCCGAATGAGC
CGGGAAGTGTACGATTAGAAGACGGTTCAATCCTGATTGGTGAGAACCCTCTGAACCCTAA
GACACGTCAAGTCTTTGACGAAGTGATTGAGCCAGAACGTGCCGCTGCTGGTGTGAACCTT
GGCGGACTTACCGAGATTGGTCTGAAGCTGCTTCGGTCTGAGAACCCGGAGATTCGTGGTG
TAGCTACTGACTTAGTGCGTTCACCGACTGGTATGCAGTCTGGGGCCTCAGGTAAAATCGG
GACCACTGCGTCTGACGTATTCGAGAGACTTCGTGCTGTAGACCATCGGTTCTACAACGAC
ATCGACGATGCTGTTACTGAGGCACTCAAGGACCCTTACTTCCAGACAGCATTCTGGCGAG
ACTCTGGCGCATTCCGTCAAGACATCTATCAGCGTGTGTCTATGGCTATCGAAGATGGTAG
TGGGAACCTGAAGGCTGAACTGACTCCGGGAGAACTGAAAGTCTATGACCTGCTGAAGAA
CCAGTTCGACGCCAAGCGTGAGATGATGGAGAACCCAGCTATGTTTGGTCGGCCAGACGCT
CAGTCTATCTTTCCGGGCAGCCGCTTCAAGGGAACCTACGTCCCGCATGTGTATAGCAAAC
AGATGAAGGAGCTGTACATCAAGGAGCTTGGGAGTCCAGAGGCGTTGCAGGAGGCCATCA
AGAAGTCATGGTTGACCAGCTATGCGTCTCGACCTGAAGTCAAGAAACGCGTGGACGAGG
CACTCTTAGAGGCTGACCCTACGTTGACCCCAGAAGGACTTGCTGCTGCGGTCGATAAGTA
CGCCAACGATAAGGCTTACGGTATCTCTCACACCGAGCAGTTCGAACGTTCATCCGTAATG
GAAGAGAACATCAACGGTCTGGTTGGTCTGGAGAACAACAGCTTCCTTGAGGCTCGTAACC
TGTTCGATAGCGATATGTCAATCGTCCTACCTAACGGTCAGACCTTCAGTGTCAACAACCT
GCGTGAGTGGGACATGGACAAGATTGTCCCGGCCTACAACCGTCGAATTAATGGCGATATT
GCTATCATGGCTGGTACAGGCAAAACCACGAAGGACATGAAGGACTTGGTTGAGACCATG
ATGAACAAGGCTGGGGATGACGGTAAGTTGAAAGGTGAAGTATCTACCTTACGTGACACC
TTGAAGATTCTAACTGGTCGTGCTCGACGTGATGGTGCTGATGATGCAGCCTTCGCTACCG
TGATGCGCACAATGACAGACCTATCGTTCTTCGCTAAGAATGCCTACATGGGTGTTCAGAA
CTTAACGGAGATTGGTGGTATGCTGGCTCGTGGCAACGTTCGTGCAATGCTGCATGGAGTC
CCAATTTTCCGTGACCTAGCCTTCCGTAACAAGAAGGTTGGGGCCTCAGAGATTAAGGACC
TGCACAATGTTATCTTCGGTAAGGAACTGGATGACTCAATCCGTCCGTCTAAACAGGATGT
CATTGACCGTCTGCGATCTTACAGTGACCTAGGTCGTGGTACAGCTACAGCTCTGGGGACT
GCCAAGTATTACACTGGCGAACTTGCAGTACGCTCTCCGTTCACTAAAGTCCTCAACGGTA

CGACCAACTACCTGTTAGATGCTGGACGTCAGGGCTTCCTGTCTGACATCGTGGAGCATAG
CCTGACTGGTAGTAAGCGTAAGTTCGATGACCGCTGGCTGAAGACCGCTGGTATCTCTGAC
GAGCAGTGGAAGGGCATTAAGTCCCTCATCCGTGAGTCAGTGACGCGTGGTCCAGACGGG
AAGTACACCATCAAGGATAAGAAGGCGTTCAGTCAGGACCCAAGGGCTATGGACCTGTGG
CGTATGGGTGACACCATCGCTGATGAAACGTTACTCCGTCCTCATAAGCTGTCCAACATGG
ACGCCAAGGCTTATGGTCCTATCGCTAAGACTGTCCTTCAGTTTAAGAACTTCGTCATCAAG
TCCATCAATGGGCGAACCATGCGTACCTTCTATAACGCCACGAAGAACAACCGAGCGATTG
ACGCTGCACTATCGACCGTGATGTCTATGGGTCTGGCTGGTATCTACTACATGGCTCAGGC
GCACGTCAAGGCTTATGCTATGCAGGATGGTCGAGACCGTGACTACCTTAAGCAAGCTCTG
GACCCGACGATGATTGGTTATGCGGCTCTGTCCCGTAGTTCACATCTGGGTGGCCCACTTG
GGGTAGCTAACATTCTAGGTGGCATCGCTGGGTATGAGGACACTAAGATGCTCCGTTCGTC
TATCCTACCTCGTTCGCCTACAGAGAAGCCTGAACGTGCCATCACGTTTGGTGCAGCTACA
AGTGACCCTGTGATGAATGTTGTTGGTAACTTCTTGGAGCAGGTTCCAGCTTTCGGATATGC
TGCTAACGTTGGCGTTTCTGCTTACAACTTGGCTGGCTACCTCAAGGCTGATACTCGTGTCA
ACGAGCGTGACTACATGACCGGGATGTATAATACGTTCCGTGAACTGGTTCCGAACGACCC
CATTACCCAGAAGCTGTTGCTTGGAATGTTTGAGGAGCAAGGCATCCACATCAAGGACGGT
GGTGGTGGTTCTGGTGGTGGTGGTTCTGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCG
ACAGACAGCCGGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTT
CAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGC
TGAAGATCGACATCCATGTCATCATCCCGTATGAATAACTCGAG

B.3 Sequence for g15::NtNL pUC19 HR cassette

TCTAGAGCGTGACGCTAACAAGCGAGCGCGTATCGTCGGGACTAAAGCTGCTGGTG
ACAAACGTCGAGCTGACCGAGAGGCCAACATCGAGAAGCGCGGTGGGATGTACAATGACG
TCTCACTGGAGGGTATCGCAAATGCACTAATCGGTAAGGAGGGTGGTGGTGGTTCTGGTGG
TGGTGGTTCTGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTAC
AACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGT
CCGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCA
TGTCATCATCCCGTATGAATAATAGCATAATGGCGACTCGTGGTATTCGCAATAACAATCC
GGGGAACATCCGGGTAAGTAAGGACCAATGGGAAGGAATGACTGGAGATGATGGCGCATT
TGTCACTTTCGATAGTCCAGAGTCTGGCGTCCGAGCTTTAGGTAAAAACCTTACTAGTAGC
GGCCGCTGCAG

B.4 Sequence for g16::NtNL pUC19 HR cassette

TCTAGAGGCTGGCTACCTCAAGGCTGATACTCGTGTCAACGAGCGTGACTACATGA
CCGGGATGTATAATACGTTCCGTGAACTGGTTCCGAACGACCCCATTACCCAGAAGCTGTT
GCTTGGAATGTTTGAGGAGCAAGGCATCCACATTAAGGACGGTGGTGGTGGTTCTGGTGGT
GGTGGTTCTGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACA
ACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTC

CGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCAT
GTCATCATCCCGTATGAATAAACTATCACTATAGGAAACGGGAGGCGCTACCATAGGTCTC
CGCTTAAATCACAAAGGAGGCATAATGTCCACGATTACACAATTCCCTTCAGGAAACACTC
AGTACAGGATTGAGTTCGACTACCTAGCCAGAACGTTTGTTGTTGTTACTAGTAGCGGCCG
CTGCAG

B.5 Sequence for g15-CRISPRa oligo

AAACATCGGTAAGGAGTAACATAAG

B.6 Sequence for the g15-CRISPRb oligo

AAAACTTATGTTACTCCTTACCGAT

B.7 Sequence for the g16-CRISPRa oligo

AAACTGATAGTTTAGTCCTTGATGG

B.8 Sequence for the g16-CRISPRb oligo

AAAACCATCAAGGACTAAACTATCA

Bibliography

[1] F. Short, T. Blower, and G. Salmond. A promiscuous antitoxin of bacteriophage T4 ensures successful viral replication. *Molecular Microbiology*, 83(4):665–668, 2012.

[2] I. Yosef, M. Manor, R. Kiro, and U. Qimron. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *PNAS*, 112(23):7267–7272, 2015.

[3] E. Keen. A century of phage research: Bacteriophages and the shaping of modern biology. *BioEssays*, 37(1):6–9, 2014.

[4] P. Leiman and M Schneider. Contractile tail machines of bacteriophages. *Adv Exp Med Biol*, 726(11):93–114, 2012.

[5] A. Ahmed, J. Rushworth, N. Hirst, and P. Millner. Biosensors for whole-cell bacterial detection. *Clin. Microbiol. Rev.*, 27:631–646, 2014.

[6] I. Sorokulova, E. Olsen, and V. Vodyanoy. Bacteriophage biosensors for antibiotic-resistant bacteria. *Expert Rev Medical Devices*, 11(2):175–186, 2014.

[7] B. Chan and S. Abedon. Phage therapy pharmacology: Phage cocktails. *Advanced Applied Microbiology*, 78:1–23, 2012.

[8] S. Santos, C. Carvalho, J. Azeredo, and E. Ferreira. Population dynamics of a salmonella lytic phage and its host: implications of the host bacterial growth rate in modelling. *PLOS One*, 10(8):e0136007, 2014.

[9] S. Abedon. Phage therapy of pulmonary infections. *Bacteriophage*, 5:e1020260, 2015.

[10] J. et al Pirnay. Quality and safety requirements for sustainable phage therapy products. *Pharm Res*, 32:2173–2179, 2015.

[11] M. Catalao, F. Gil, J. Moniz-Pereira, C. Sao-Jose, and M. Pimentel. Diversity in bacterial lysis systems: bacteriophages show the way. *FEMS Microbiology Review*, 37:554–571, 2013.

[12] A. Campbell. The future of bacteriophage biology. *Nature Reviews Genetics*, 4(6):471–477, 2003.

[13] G. Bertani. Lysogeny at mid-twentieth century: P1, p2, and other experimental systems. *Journal of Bacteriology*, 186(3):595–600, 2004.

[14] Cortez M. Dushoff J. Li, G. and J. Weitz. When to be temperate: on the fitness benefits of lysis vs. lysogeny. *Virus Evolution*, 6(2), 2020.

[15] Lion S. Buckling A. Westra E. Bruce, J. and S. Gandon. Regulation of prophage induction and lysogenization by phage communication systems. *Current Biology*, 31(22):5046–5051, 2021.

[16] F. Twort. An investigation on the nature of ultra-microscopic viruses. *The Lancet*, 186(4814):1241–1243, 1915.

[17] F. d'Herelle. Sur un microbe invisible antagoniste des bacillus dysentérique. *Acad Sci Paris*, 165:373–375, 1917.

[18] A. Fleming. On the antibacterial action of cultures of a penicillium with special reference to their use in the isolation of b. influenza. *Br J Exp Pathol.*, 10:226–236, 1929.

[19] R. Rappuoli and S. Black. Deploy vaccines to fight superbugs. *Nature*, 552(7684):165–167, 2017.

[20] J. O'Neill. Antimicrobial resistance: Tackling a crisis for the health and wealth of nations. *AMR Review*, 2014.

[21] C. Llor and L. Bjerrum. Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem, journal = Therapeutic advances in drug safety, volume = 5, number = 6, pages = 229-241, year = 2014.

[22] J. Wheatley, S. Liyanagedera, R. Amaee, A. Sagona, and V. Kulkarni. *Advances in Synthetic Biology*. Springer, 2020.

[23] B. Allegranzi. Report on the burden of endemic health care-associated infection worldwide. *World Health Organisation*, 2011.

[24] J. Cimiotti, L. Aiken, D. Sloane, and E. Wu. Nurse staffing, burnout, and health care-associated infection. *American Journal on Infection Control*, 40:486–490, 2012.

[25] L. et al Richter. Recent advances in bacteriophage-based methods for bacteria detection. *Drug Discovery Today*, 23:448–455, 2018.

[26] M. Schmelcher and M. Loessner. Application of bacteriophages for detection of foodborne pathogens. *Bacteriophage*, 4:e28137, 2014.

[27] L. et al Hardala. Biotechnological applications of bacteriophages: State of the art. *Microbial Research*, 212:38–58, 2018.

[28] A. Singh, S. Poshtiban, and S. Evoy. Recent advances in bacteriophage based biosensors for food-borne pathogen detection. *Sensors (Basel)*, 13(2):1763–1786, 2013.

[29] J. Shin, P. Jardin, and V. Noireaux. Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction. *ACS Synthetic Biology*, 1(9):408–413, 2012.

[30] M. Rustad, A. Eastlund, R. Marshall, P. Jardine, and V. Noireaux. Synthesis of infectious bacteriophages in an e. coli-based cell-free expression system. *Synthetic Biology*, (126):56144, 2017.

[31] M. Rustad, A. Eastlund, P. Jardine, and V. Noireaux. Cell-free txtl synthesis of infectious bacteriophage T4 in a single test tube reaction. *Synthetic Biology*, 3(1):ysy002, 2018.

[32] J. Rees and J. Barr. Detection of methicillin-resistant staphylococcus aureus using phage amplification combined with matrix-assisted laser desorption/ionization mass spectrometry. *Anal Bioanal Chem*, 409(7684):1379–1386, 2017.

[33] C. Nguyen, R. Makkar, N. Sharp, M. Page, I. Molineux, and D. Schofield. Detection of bacillus anthracis spores from environmental water using bioluminescent reporter phage. *Journal of Applied Microbiology*, 123(5):1184–1193, 2017.

[34] N. Bharadwaj, S. Bharadwaj, J. Mehta, K. Kim, and A. Deep. Mof-bacteriophage biosensor for highly sensitive and specific detection of staphylococcus aureus. *ACS Appl Mater Interfaces*, 9:33589–33598, 2017.

[35] Y. He, M. Wang, E. Fan, H. Ouyang, H. Yue, X. Su, G. Liao, L. Wang, S. Lu, and Z Fu. Highly specific bacteriophage-affinity strategy for rapid separation and sensitive detection of viable pseudomonas aeruginosa. *Analytical Chemistry*, 89:1916–1921, 2017.

[36] C. England, E. Ehlerding, and W. Cai. Nanoluc: A small luciferase is brightening up the field of bioluminescence. *Bioconjugation Chemistry*, 27:1175–1187, 2016.

[37] E. Pulkkinen, T. Hinkley, and S. Nugen. Utilizing in vitro DNA assembly to engineer a synthetic T7 nanoluc reporter phage for escherichia coli detection. *Molecular Microbiology*, 11(3):63–68, 2019.

[38] H. Hoang and T. Dien. Rapid and simple colorimetric detection of escherichia coli O157-H7 in apple juice using a novel recombinant bacteriophage-based method. *Biocontrol Science*, 20(2):99–103, 2015.

[39] Y. Tanji, C. Furukawa, S. Na, T. Hijikata, K. Miyanaga, and H. Unno. Escherichia colidetection by GFP-labeledlysozyme-inactivated T4 bacteriophage. *Journal of Biotechnology*, 114(1-2):11–20, 2004.

[40] M. Oda, M. Morita, H. Unno, and Y. Tanji. Rapid detection of escherichia coli O157:H7 by using green fluorescent protein-labeled PP01 bacteriophage. *Appl Environ Microbiol*, 70:527–534, 2004.

[41] A. Smartt, T. Xu, P. Jegier, J. Carswell, S. Blount, G. Sayler, and S. Ripp. Pathogen detection using engineered bacteriophages. *Analytical and Bioanalytical Chemistry*, 402(10):3127–3146, 2012.

[42] K. Sergueev, Filippov A., and M. Nikolich. Highly sensitive bacteriophage-based detection of brucella abortus in mixed culture and spiked blood. *Viruses*, 9(6):144, 2017.

[43] S. Alcaine, D. Pacitto, D. Sela, and S. Nugen. Phage and phosphatase: a novel phage-based probe for rapid, multi-platform detection of bacteria. *Analyst*, 140(22):7629–7636, 2015.

[44] S. Hagens, T. De Wouters, P. Vollenweider, and M. Loessner. Reporter bacteriophage a511::celb transduces a hyperthermostable glycosidase from pyrococcus furiosus for rapid and simple detection of viable listeria cells. *Bacteriophage*, 1:143–151, 2011.

[45] J. Monod and F. Jacob. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol*, 26:389–401, 1961.

[46] T. Sampson and D. Weiss. Exploiting CRISPR/Cas systems for biotechnology. *Bioassays*, 36(1):34–38, 2014.

[47] R. Sorek, V. Kunin, and P. Hugenholtz. CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology*, 6:181–186, 2008.

[48] S. Shmakov, V. Sitnik, K. Makarova, Y. Wolf, K. Severinov, and E. Koonin. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *mBio.*, 8(5):e01397–17, 2017.

[49] A. Bolotin, B. Quinquis, A. Sorokin, and S. Ehrlich. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151:2551–2561, 2005.

[50] R. Barrangou, C. Fremaux, H. Deveu, D. Romero, and P. Horvath. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315:1709–1712, 2007.

[51] P. et al Horvath. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *Journal of Bacteriology*, 190(4):1401–1412, 2008.

[52] A. Stern, L. Keren, O. Wurtzel, G. Amitai, and R. Sorek. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends in Genetics*, 26(8):335–340, 2010.

[53] G. Goldberg, W. Jiang, D. Bikar, and L. Marraffini. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature*, 514(5):633–637, 2014.

[54] Z. Karim, A. Ahmadi, A. Najafi, and R. Ranjbar. Bacterial CRISPR regions: General features and their potential for epidemiological molecular typing studies. *The Open Microbiology Journal*, 12:59–70, 2018.

[55] C. Pourcel, G. Salvignol, and G. Vergnaud. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151:653–663, 2005.

[56] H. Deveu, J. Garneau, and S. Moineaus. CRISPR/Cassystem and its role in phage-bacteria interactions. *Annual Review of Microbiology*, 64:475–493, 2010.

[57] R. Lillestol, P. Redder, R. Garret, and K. Brugger. A putative viral defence mechanism in archaeal cells. *Arachea*, 2(1):59–72, 2006.

[58] U. Pul, R. Wurm, Z. Arslan, R. Geissen, N. Hofmann, and R. Wagner. Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Molecular Microbiology*, 75:1495–1512, 2010.

[59] R. Lillestol, S. Shah, R. Garret, K. Brugger, P. Redder, H. Phan, J. Christiansen, and R. Garret. CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Molecular Microbiology*, 72(1):259–272, 2009.

[60] P. Horvath and R. Barrangou. CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 327(5962):167–170, 2010.

[61] R. Barrangou. CRISPR-Cas systems and RNA-guided interference. *Wiley Interdisciplinary Review of RNA*, 4:267–278, 2013.

[62] R. Barrangou and L. Marraffini. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Molecular Cell*, 54:234–244, 2014.

[63] L. Xiao-Jie, X. Hui-Ying, C. Zun-Ping, K. Jin-Lian, and J. Li-Juan. CRISPR-Cas9: a new and promising player in gene therapy. *Journal of Medical Genetics*, 52:289–296, 2015.

[64] D. Siegal-Gaskins, Z. Tuza, J. Kim, V. Noireaux, and R. Murray. Gene circuit performance characterization and resource usage in a cell-free "breadboard". *ACS Synthetic Biology*, 3(6):416–425, 2014.

[65] M. et al Takahashi. Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems. *ACS Synthetic Biology*, 4(5):503–515, 2015.

[66] J. Garamella, R. Marshall, M. Rustad, and V. Noireaux. The all e. coli tx-tl toolbox 2.0: A platform for cell-free synthetic biology. *ACS Synthetic Biology*, 5:344–355, 2016.

[67] Q. Dudley, A. Karim, and M. Jewett. Cell-free metabolic engineering: biomanufacturing beyond the cell. *Journal of Biotechnology*, 10:69–82, 2015.

[68] B. Bundy, J. Hunt, M. Jewett, J Swartz, D. Wood, D Frey, and G. Rado. Cell-free biomanufacturing. *Current Opinion in Chemical Engineering*, 22:177–183, 2018.

[69] C. You and Y. Zhang. Cell-free biosystems for biomanufacturing. *Adv Biochem Eng Biotechnol*, (131):89–119, 2013.

[70] P. Ng, M. Jia, K. Patel, J. Brody, J. Swartz, S. Levy, and R. Levy. A vaccine directed to B cells and produced by cell-free protein synthesis generates potent antilymphoma immunity. *PNAS*, 109(36):14526–14531, 2012.

[71] O. Ogonah, K. Polizzi, and D. Bracewell. Cell free protein synthesis: a viable option for stratified medicines manufacturing? *Current Opinion in Chemical Engineering*, 18:77–83, 2017.

[72] Z. Sun, C. Hayes, J. Shin, F. Caschera, R. Murray, and V. Noireaux. Protocols for implementing an escherichia coli based TX-TL cell-free expression system for synthetic biology. *Trends in Genetics*, 79:e50762, 2013.

[73] A. Didovyk, T. Tanooka, L. Tsimring, and J. Hasty. Rapid and scalable preparation of bacterial lysates for cell-free gene expression. *ACS Synthetic Biology*, 6:2198–2208, 2017.

[74] D. Garenne, S. Thompson, A. Brisson, A. Khakimzhan, and V. Noireaux. The all-e. coli TXTL toolbox 3.0: new capabilities of a cell-free synthetic biology platform. *Synthetic Biology*, 6(1), 2021.

[75] Y. Wang, K. Wei, and C. Smolke. Synthetic biology: advancing the design of diverse genetic systems. *Annual review of chemical and biomolecular engineering*, 4:69–102, 2013.

[76] K. et al Pardee. Rapid, low-cost detection of zika virus using programmable biomolecular components. *Cell*, 165:1255–1266, 2016.

[77] D. Ma, L. Shen, K. Wu, C. Deihnelt, and A. Green. Low-cost detection of norovirus using paper-based cell-free systems and synbody-based viral enrichment. *Synthetic Biology Oxford*, 3(1), 2018.

[78] A. et al Grawe. A paper-based, cell-free biosensor system for the detection of heavy metals and date rape drugs. *PLOS One*, 14(3):e0210940, 2019.

[79] C. Xu, S. Hu, and X. Chen. Artificial cells: from basic science to applications. *Materials Today*, 19(9):516–532, 2016.

[80] K. Adamala, D. Martin-Alarcon, K Guthrie-Honea, and E. Boyden. Engineering genetic circuit interactions within and between synthetic minimal cells. *Nat. Chem.*, 9:431–439, 2017.

[81] G. Rampioni, F. D'Angelo, M. Messina, A. Zennaro, Y. Kuruma, D. Tofani, L. Leoni, and P. Stano. Synthetic cells produce a quorum sensing chemical signal perceived by Pseudomonas aeruginosa. *Chemical Communications (Cambridge)*, 54(17):2090–2093, 2018.

[82] T. Tang, B. An, Y. Huang, S. Vasikaran, Y. Wang, X. Jiang, T. Lu, and C. Zhong. Materials design by synthetic biology. *Nature reviews materials*, 6:332–350, 2021.

[83] J. Merritt, A. Ollis, A Fisher, and M. DeLisa. Glycans-by-design: Engineering bacteria for the biosynthesis of complex glycans and glycoconjugates. *Biotechnol. Bioeng.*, 110:1550–1564, 2013.

[84] M. Yanagisawa and R. Yu. The expression and functions of glycoconjugates in neural stem cells. *Glycobiology*, 17(7):57–74, 2007.

[85] S. Sachdeva, R. Palur, K. Sudhakar, and T. Rathinavelan. E. coli group 1 capsular polysaccharide exportation nanomachinary as a plausible antivirulence target in the perspective of emerging antimicrobial resistance. *Frontiers in Microbiology*, 8, 2017.

[86] U. Mamat, K. Wilke, D. Bramhill, A. Schromm, B. Lindner, T. Kohl, J. Corchero, A. Villaverde, L. Schaffer, S. Head, C. Souvignier, T. Meredith, and R. Woodard. Detoxifying escherichia coli for endotoxin-free production of recombinant proteins. *Microb Cell Fact*, 14(57), 2015.

[87] G. Nelson and M. Greene. *Principles and Practice of Infectious Diseases*. Elsevier, 2020.

[88] J. Angelin and M. Kavitha. Exopolysaccharides from probiotic bacteria and their health potential. *International journal of biological macromolecules*, 162:853–865, 2020.

[89] B. Kunduru, S. Nair, and T. Rathinavelan. Ek3d: an e. coli k antigen 3-dimensional structure database. *Nucleic acids research*, 44:675–681, 2016.

[90] J. King, H. Aal Owaif, J. Jia, and I. Roberts. Phenotypic heterogeneity in expression of the K1 polysaccharide capsule of uropathogenic escherichia coli and downregulation of the capsule genes during growth in urine. *Infection and immunity*, 83(7):2605–2613, 2015.

[91] C. Meier, T. Oelschlaeger, H. Merkert, T. Korhonen, and J. Hacker. Ability of escherichia coli isolates that cause meningitis in newborns to invade epithelial and endothelial cells. *Infection and immunity*, 64(7):2391–2399, 1996.

[92] S. Huang, Y. Chen, Q. Fu, M. Stins, Y. Wang, C. Wass, and K. Kim. Identification and characterization of an escherichia coli invasion gene locus, ibeb, required for penetration of brain microvascular endothelial cells. *Infection and Immunity*, 67(5):2103–2109, 1999.

[93] K. Kim, S. Elliott, F. Di Cello, M. Stins, and K. Kim. The K1 capsule modulates trafficking of e. coli-containing vacuoles and enhances intracellular bacterial survival in human brain microvascular endothelial cells. *Cellular microbiology*, 5(4):245–252, 2003.

[94] C. Moller-Olsen, S. Ho, R. Shukla, T. Feher, and A. Sagona. Engineered K1F bacteriophages kill intracellular escherichia coli K1 in human epithelial cells. *Scientific Reports*, 8:17559, 2018.

[95] E. Moxon and J. Kroll. *Bacterial Capsules*. Springer, 1990.

[96] S. Metkar, S. Awasthi, E. Denamur, K. Kim, S. Gangloff, S. Teichberg, A. Haziot, J Silver, and S. Goyert. Role of cd14 in responses to clinical isolates of escherichia coli : Effects of K1 capsule expression. *Infection and Immunity*, 75(11), 2007.

[97] P. Kemp, L. Garcia, and I. Molineux. Changes in bacteriophage T7 virion structure at the initiation of infection. *Virology*, 340(2):307–317, 2005.

[98] I. Molineux. *Encyclopedia of Virology*. Elsevier, 1999.

[99] A. King, M. Adams, E. Carstens, and E. Lefkowitz. *Virus taxonomy*. Elsevier, 2011.

[100] D. Scholl and C. Merril. The genome of bacteriophage K1F, a T7-like phage that has acquired the ability to replicate on K1 strains of escherichia coli. *Journal of Bacteriology*, 187(24), 2005.

[101] G. Petter and E. Vimr. Complete nucleotide sequence of the bacteriophage K1F tail gene encoding endo-n-acylneuraminidase (endo-n) and comparison to an endo-n homolog in bacteriophage PK1E. *J. Bacteriol.*, 175:4354–4363, 1993.

[102] H. Yue, Y. Li, M. Yang, and C. Mao. T7 phage as an emerging nanobiomaterial with genetically tunable target specificity. *Adv Sci*, 9(4), 2022.

[103] L. Garcia and I. Molineux. Rate of translocation of bacteriophage T7 DNA across the membranes of escherichia coli. *Journal of Bacteriology*, 177(14):4066–4076, 1995.

[104] X. Zhang and W. Studier. Multiple roles of T7 RNA polymerase and T7 lysozyme during bacteriophage T7 infection. *Journal of Molecular Biology*, 340(4):707–730, 2004.

[105] b. Cámara, M. Liu, J. Reynolds, A. Shadrin, B. Liu, K. Kwok, P. Simpson, R. Weinzierl, K. Severinov, E. Cota, S. Matthews, and S. Wigneshweraraj. T7 phage protein gp2 inhibits the escherichia coli RNA polymerase by antagonizing stable DNA strand separation near the transcription start site. *PNAS*, 107(5):2247–2252, 2010.

[106] S. Leptihn, J. Gottschalk, and A. Kuhn. T7 ejectosome assembly: A story unfolds. *Bacteriophage*, 6(1):e1128513, 2016.

[107] E. Vimr and F. Troy. Regulation of sialic acid metabolism in escherichia coli: role of n-acylneuraminate pyruvate-lyase. *Journal of Bacteriology*, 164(2):854–860, 1985.

[108] J. Shin and V. Noireaux. Efficient cell-free expression with the endogenous e. coli RNA polymerase and sigma factor 70. *Journal of Biological Engineering*, 4(8):1754–1611, 2010.

[109] W. Jiang, D. Bikard, D. Cox, F. Zhang, and L. Marraffini. RNA-guided editing of bacterial genomes using CRISPR-cas systems. *Nature Biotechnology*, 31:233–239, 2013.

[110] T. Hinkley, S. Garing, S. Singh, A. Le Ny, K. Nichols, J. Peters, J. Talbert, and S. Nugen. Reporter bacteriophage T7NLC utilizes a novel nanoluc::CBM fusion for the ultrasensitive detection of escherichia coli in water. *Analyst*, 143(17):4074–4082, 2018.

[111] A. Aksyuk, P. Leiman, L. Kurochkina, M. Shneider, V. Kostyuchenko, V. Mesyanzhinov, and M. Rossmann. The tail sheath structure of bacteriophage t4: a molecular machine for infecting bacteria. *The EMBO Journal*, 28(7):821–829, 2009.

[112] W. Chen, H. Xiao, X. Wang, S. Song, Z. Han, X. Li, F. Yang, J. Song, H. Liu, and L. Cheng. Structural changes of a bacteriophage upon DNA packaging and maturation. *Protein Cell*, 11:374–379, 2020.

[113] I. Molineux and D. Panja. Popping the cork: mechanisms of phage genome ejection. *Nat Rev Microbiol*, 11:194–204, 2013.

[114] L. Oliveira, P. Tavaresa, and C. Alonso. Headful DNA packaging: Bacteriophage SPP1 as a model system. *Virus Research*, 173(2):247–259, 2013.

[115] P. Leiman, F. Arisaka, M. van Raaij, V. Kostyuchenko, A. Aksyuk, S. Kanamaru, and M. Rossmann. Morphogenesis of the t4 tail and tail fibers. *Virol J*, 7(355), 2010.

[116] L. Garcia and I. Molineux. Transcription-independent DNA translocation of bacteriophage T7 DNA into escherichia coli. *Journal of Bacteriology*, 178(23):6921–6929, 1996.

[117] I. Molineux. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Molecular microbiology*, 40(1):1–8, 2001.

[118] M. Pérez-Ruiz, M. Pulido-Cid, J. Luque-Ortega, J. Valpuesta, A. Cuervo, and J. Carrascosa. Assisted assembly of bacteriophage T7 core components for genome translocation across the bacterial envelope. *BioRxiv*, 2021.

[119] W. Chen, H. Xiao, L. Wang, X. Wang, Z. Tan, Z. Han, X. Li, F. Yang, Z. Liu, J. Song, H. Liu, and L. Cheng. Structural changes in bacteriophage T7 upon receptor-induced genome ejection. *PNAS*, 118(37), 2021.

[120] A. Cuervo, M. Fàbrega-Ferrer, C. Machón, J. Conesa, F. Fernandez, R. Perez-Luque, J. Pous, M. Vega, J. Carrascosa, and M. Coll. Structures of T7 bacteriophage portal and tail suggest a viral DNA retention and ejection mechanism. *Nat Commun*, 40:3746, 2019.

[121] N. Swanson, R. Lokareddy, F. Li, C. Hou, S. Leptihn, M. Pavlenok, M. Niederweis, R. Pumroy, V. Moiseenkova-Bell, and G. Cingolani. Cryo-em structure of the periplasmic tunnel of T7 DNA-ejectosome at 2.7 Å resolution. *Molecular Cell*, 81(15):3145–3159, 2021.

[122] X. Agirrezabala, J. Martín-Benito, J. Castón, R. Miranda, J. Valpuesta, and J. Carrascosa. Maturation of phage T7 involves structural modification of both shell and inner core components. *The EMBO Journal*, 24:3820–3829, 2005.

[123] E. Widder. Bioluminescence in the ocean: origins of biological, chemical, and ecological diversity. *Science*, 328:704–708, 2010.

[124] M. Hall, J. Unch, B. Binkowski, M. Valley, B. Butler, M. Wood, P. Otto, K. Zimmerman, G. Vidugiris, T. Machleidt, M. Robers, H. Benink, C. Eggers, M. Slater, P. Meisenheimer, D. Klaubert, F. Fan, L. Encell, and K. Wood. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chemical Biology*, 7(11):1848–1857, 2012.

[125] T. Riss. Nanoluc®: A smaller, brighter, and more versatile luciferase reporter. https://www.promega.com/~/media/files/promega%20worldwide/europe/promega%20uk/webinars%20and%20events/cell%20analysis%20seminar%20tour/terry-riss-02.pdf, Accessed: 2022-02-28.

[126] J. Zhao, T. Nelson, Q. Vu, T. Truong, and C. Stains. Self-assembling nanoluc luciferase fragments as probes for protein aggregation in living cells. *ACS Chemical Biology*, 11(1):132–138, 2015.

[127] B. Li, Y. Li, D. Bai, X. Zhang, H. Yang, J. Wang, G. Liu, J. Yue, Y. Ling, D. Zhou, and H. Chen. Whole-cell biotransformation systems for reduction of prochiral carbonyl compounds to chiral alcohol in escherichia coli. *Sci Rep*, 4:6750, 2014.

[128] T. Nelson, J. Zhao, and C. Stains. Utilizing split-nanoluc luciferase fragments as luminescent probes for protein solubility in living cells. *Methods in enzymology*, 622(55-66), 2019.

[129] Benchling. Analyze as a translation. https://help.benchling.com/en/articles/4832352-create-view-and-analyze-translations#:~:text=Create%20a%20translation,-Open%20a%20DNA&text=Click%20Create%20at%20the%20top,translation%20will%20automatically%20be%20indexed., Accessed: 2022-05-15.

[130] A. Chenal, P. Nizard, V. Forge, M. Pugnière, M. Roy, J. Mani, F. Guillain, and D. Gillet. Does fusion of domains from unrelated proteins affect their folding pathways and the structural changes involved in their function? a case study with the diphtheria toxin t domain. *Protein Engineering*, 15(5):383–391, 2002.

[131] Calctool. Protein size. https://www.calctool.org/CALC/prof/bio/protein_size, Accessed: 2022-05-15.

[132] S. Kowalczykowski, D. Dixon, A. Eggleston, S. Lauder, and W. Rehrauer. Biochemistry of homologous recombination in escherichia coli. *Microbiological Reviews*, 58(3):401–465, 1994.

[133] NEB. puc19 vector. https://international.neb.com/products/n3041-puc19-vector#Product%20Information, Accessed: 2022-03-16.

[134] Addgene. Crispr guide. https://www.addgene.org/guides/crispr/, Accessed: 2022-03-16.

[135] C. Moller-Olsen, T. Ross, K. Leppard, V. Foisor, C. Smith, D. Grammatopoulos, and A. Sagona. Bacteriophage K1F targets escherichia coli K1 in cerebral endothelial cells and influences the barrier function. *Scientific Reports*, 10(1), 2020.

[136] I. Shachrai, A. Zaslaver, U. Alon, and E. Dekel. Cost of unneeded proteins in e. coli is reduced after several generations in exponential growth. *Molecular Cell*, 38(5):758–767, 2010.

[137] Y. Kwon and M. Jewett. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Scientific Reports*, 5(1), 2015.

[138] P. Shrestha, T. Holland, and B. Bundy. Streamlined extract preparation for escherichia coli-based cell-free protein synthesis by sonication or bead vortex mixing. *Biotechniques*, 53(3), 2012.

[139] J. Kim, C. Copeland, S. Padumane, and Y. Kwon. A crude extract preparation and optimization from a genomically engineered escherichia coli for the cell-free protein synthesis system: Practical laboratory guideline. methods and protocols. *Methods and Protocols*, 2(3):68, 2019.

[140] M. Levine, N. Gregorio, M. Jewett, K. Watts, and J. Oza. Escherichia coli-based cell-free protein synthesis: Protocols for a robust, flexible, and accessible platform technology. *Journal of Visualized Experiments*, 144, 2019.

[141] Arbor Biosciences. Cell free protein expression. https://arborbiosci.com/synthetic-biology/cell-free-protein-expression/mytxtl/, Accessed: 2022-05-16.

[142] T. Kigawa, T. Yabuki, N. Matsuda, T. Matsuda, R. Nakajima, A. Tanaka, and S. Yokoyama. Preparation of escherichia coli cell extract for highly productive cell-free protein expression. *Journal of Structural and Functional Genomics*, 5:63–68, 2004.

[143] V. Noireaux, R. Bar-Ziv, and A. Libchaber. Principles of cell-free genetic circuit assembly. *Proceedings of the National Academy of Sciences*, 100(22):12672–12677, 2003.

[144] D. Liu, J. Zawada, and J. Swartz. Streamlining escherichia coli s30 extract preparation for economical cell-free protein synthesis. *Biotechnology Progress*, 21(2):460–465, 2005.

[145] A. Silverman, N. Kelley-Loughnane, J. Lucks, and M. Jewett. Deconstructing cell-free extract preparation for in vitro activation of transcriptional genetic circuitry. *ACS Synthetic Biology*, 8(2):403–414, 2019.

[146] J. Garrett, R. Fusselman, J. Hise, L. Chiou, D. Smith-Grillo, J. Schulz, and R. Young. Cell lysis by induction of cloned lambda lysis genes. *Mol Gen Genet.*, 182(2), 1981.

[147] S. Crabtree and J. Cronan. Facile and gentle method for quantitative lysis of escherichia coli and salmonella typhimurium. *Journal of Bacteriology*, 158:354–356, 1984.

[148] E. Falgenhauer, S. Schönberg, C. Meng, A. Mückl, K. Vogele, Q. Emslander, C. Ludwig, and F. Simmel. Evaluation of an e. coli cell extract prepared by lysozyme-assisted sonication via gene expression, phage assembly and proteomics. *ChemBioChem*, 22(18):2805–2813, 2021.

[149] K. Fujiwara and N. Doi. Biochemical preparation of cell extract for cell-free protein synthesis without physical disruption. *PLoS ONE*, 11(4), 2016.

[150] J. Chatton. *Reference Module in Biomedical Sciences*. Elsevier, 2015.

[151] X. Guo, Y. Zhu, L. Bai, and D. Yang. The protection role of magnesium ions on coupled transcription and translation in lyophilized cell-free system. *ACS Synthetic Biology*, 9(4):856–863, 2020.

[152] D. Garenne, M. Haines, E. Romantseva, P. Freemont, E. Strychalski, and V. Noireaux. Cell-free gene expression. *Nat Rev Methods Primers*, 1(49), 2021.

[153] M. Winkler, A. Glieder, and N. Klempier. Enzyme stabilizer dtt catalyzes nitrilase analogue hydrolysis of nitriles. *ChemInform*, 37(30), 2006.

[154] B. Akabayov, S. Akabayov, S. Lee, G. Wagner, and C. Richardson. Impact of macromolecular crowding on DNA replication. *Nat Commun*, 4(1615), 2013.

[155] M. McSweeney and M. Styczynski. Effective use of linear DNA in cell-free expression systems. *Frontiers in Bioengineering and Biotechnology*, 9, 2021.

[156] R. Marshall, C. Maxwell, S. Collins, C. Beisel, and V. Noireaux. Short DNA containing sites enhances DNA stability and gene expression in e. coli cell-free transcription-translation systems. *Biotechnology and bioengineering*, 114(9):2137–2141, 2017.

[157] A. Dixon, M. Schwinn, M. Hall, K. Zimmerman, P. Otto, T. Lubben, B. Butler, B. Binkowski, T. Machleidt, T. Kirkland, M. Wood, C. Eggers, L. Encell, and K. Wood. Nanoluc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chemical Biology*, 11(2):400–408, 2016.

[158] R. Rozbeh and K. Forchhammer. Split nanoluc technology allows quantitation of interactions between pii protein and its receptors with unprecedented sensitivity and reveals transient interactions. *Sci Rep*, 11(12535), 2021.

[159] Y. Chen, H. Batra, J. Dong, C. Chen, V. Rao, and P. Tao. Genetic engineering of bacteriophages against infectious diseases. *Frontiers in Microbiology*, 10, 2019.

[160] G. Kellenberger, M. Zichichi, and J. Weigle. Exchange of DNA in the recombination of bacteriophage . *Proceedings of the National Academy of Sciences*, 47(6):869–878, 1961.

[161] G. Sarkis, W. Jacobs, and G. Hatfulll. L5 luciferase reporter mycobacteriophages: a sensitive tool for the detection and assay of live mycobacteria. *Molecular Microbiology*, 15(6):1055–1067, 1995.

[162] T. Fehér, I. Karcagi, F. Blattner, and G. Pósfai. Bacteriophage recombineering in the lytic state using the lambda red recombinases. *Microbial Biotechnology*, 5(4):466–476, 2011.

[163] A. Poteete. What makes the bacteriophage red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiology Letters*, 201(1):9–14, 2001.

[164] P. Tao, X. Wu, W. Tang, J. Zhu, and V. Rao. Engineering of bacteriophage T4 genome using CRISPR-Cas9. *Journal of Biotechnology*, 6(10):1952–1961, 2017.

[165] J. Shen, J. Zhou, G. Chen, and Z. Xiu. Efficient genome engineering of a virulent klebsiella bacteriophage using CRISPR-Cas9. *Journal of Virology*, 92(17):e00534–1, 2018.

[166] B. Martel and S. Moineau. CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. *Nucleic Acid Research*, 42(14):9504–9513, 2014.

[167] E. Miller, E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. RUGER. Bacteriophage T4 genome. *Microbiol Mol Biol Rev*, 67:86–156, 2003.

[168] P. Mohanraju, K. Makarova, B. Zetsche, F. Zhang, E. Koonin, and J. van der Oost. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science*, 353(6299):aad5147, 2016.

[169] R. Kiro, D. Shitrit, and U. Qimron. Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA Biology*, 11(1):42–44, 2014.

[170] M. Duong, C. Carmody, Q. Ma, J. Peters, and S. Nugen. Optimization of t4 phage engineering via crispr/cas9. *Scientific Reports*, 10(1), 2020.

[171] S. Abedon and E. Bartom. *Brenner's Encyclopedia of Genetics (Second Edition)*. Elsevier, 2013.

[172] A. Box, M. McGuffie, B. O'Hara, and K. Seed. Functional analysis of bacteriophage immunity through a type I-E CRISPR-cas system in vibrio cholerae and its application in bacteriophage genome engineering. *Journal of Bacteriology*, 198(3):578–590, 2016.

[173] S. Bari, F. Walker, K. Cater, B. Aslan, and A. Hatoum-Aslan. Strategies for editing virulent staphylococcal phages using CRISPR-cas10. *ACS Synthetic Biology*, 6(12):2316–2325, 2017.

[174] N. Ács, M. Gambino, and L. Brøndsted. Bacteriophage enumeration and detection methods. *Frontiers in Microbiology*, 11, 2020.

[175] P. González-Torres, F. Rodríguez-Mateos, J. Antón, and T. Gabaldón. Impact of homologous recombination on the evolution of prokaryotic core genomes. *mBio*, 10(1), 2019.

[176] R. Smyth, T. Schlub, A. Grimm, V. Venturi, A. Chopra, S. Mallal, M. Davenport, and J. Mak. Reducing chimera formation during pcr amplification to ensure accurate genotyping. *Gene*, 469(1):45–51, 2010.

[177] E. Vandermarliere, M. Mueller, and L. Martens. Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrometry Reviews*, 32(6):453–465, 2013.

[178] W. Qian, T. Liu, M. Monroe, E. Strittmatter, J. Jacobs, L. Kangas, K. Petritis, D. Camp, and R. Smith. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and sequest analysis: The human proteome. *Journal of Proteome Research*, 4(1):53–62, 2004.

[179] J. Vrana, J. Gamez, B. Madden, J. Theis, H. Bergen, and A. Dogan. Classification of amyloidosis by laser microdissection and mass spectrometry–based proteomic analysis in clinical biopsy specimens. *Blood*, 114(24):4957–4959, 2009.

[180] W. Bowen, J. Calvo, and A. Hernández. Steps of membrane blocking in flux decline during protein microfiltration. *Journal of Membrane Science*, 101(1-2):153–165, 1995.

[181] O. Birrenbach, F. Faust, M. Ebrahimi, R. Fan, and P. Czermak. Recovery and purification of protein aggregates from cell lysates using ceramic membranes: Fouling analysis and modeling of ultrafiltration. *Frontiers in Chemical Engineering*, 3, 2021.

[182] A. Cole, S. Tran, and A. Ellington. Heat adaptation of phage T7 under an extended genetic code. *Virus Evolution*, 7(2), 2021.

[183] Z. Vörös, G. Sevcsik, G. Csík, L. Herényi, and M. Kellermayer. Temperature-dependent nanomechanics and topography of bacteriophage T7. *Biophysical Journal*, 114(3):354, 2018.

[184] M. Pyne, M. Moo-Young, D. Chung, C. Chou, and M. Kivisaar. Coupling the CRISPR/cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in escherichia coli. *Applied and Environmental Microbiology*, 81(15):5103–5114, 2015.

[185] A. Steven, B. Trus, J. Maizel, M. Unser, D. Parry, J. Wall, J. Hainfeld, and F. Studier. Molecular substructure of a viral receptor-recognition protein. the gp17 tail-fiber of bacteriophage T7. *Mol. Biol.*, 200:351–365, 1988.

[186] H. Jones, C. Shield, and B. Swift. The application of bacteriophage diagnostics for bacterial pathogens in the agricultural supply chain: From farm-to-fork. *PHAGE*, 1(4):176–188, 2020.

[187] J. Verma, S. Saxena, and S. Babu. *Analyzing Microbes*. Springer, 2013.

[188] M. Verma, J. Rogowski, L. Jones, and F. Gu. Colorimetric biosensing of pathogens using gold nanoparticles. *Biotechnology Advances*, 33(6):666–680, 2015.

[189] Matea C.-Pop T. Mosteanu O. Buzoianu A. Puia C. Iancu C. Mocan, T. and L. Mocan. Development of nanoparticle-based optical sensors for pathogenic bacterial detection. *Journal of Nanobiotechnology*, 15(1), 2017.

[190] Y. Yamamoto. PCR in diagnosis of infection: Detection of bacteria in cerebrospinal fluids. *Clinical and Vaccine Immunology*, 9(3):508–514, 2002.

[191] Malek L., Sooknanan R., and Compton J. *Protocols for Nucleic Acid Analysis by Nonradioactive Probes*. Humana Press, 1994.

[192] T. Notomi. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Research*, 28(12):63e, 2000.

[193] R. Daher, G. Stewart, M. Boissinot, and M. Bergeron. Recombinase polymerase amplification for diagnostic applications. *Clinical Chemistry*, 62(7):947–958, 2016.

[194] S. Kim and S. Kim. Bacterial pathogen detection by conventional culture-based and recent alternative (polymerase chain reaction, isothermal amplification, enzyme linked immunosorbent assay, bacteriophage amplification, and gold nanoparticle aggregation) methods in food samples: A review. *Journal of Food Safety*, 41(1), 2020.

[195] A. Deisingh and M. Thompson. Strategies for the detection of escherichia coli o157:h7 in foods. *Journal of Applied Microbiology*, 96(3):419–429, 2004.

[196] Coleparmer. Rapid e. coli o157 and shiga toxin antigen detection test kits - cole-parmer. https://www.coleparmer.com/p/rapid-e-coli-o157-and-shiga-toxin-antigen-detection-test-kits/15361, Accessed: 2022-02-22.

[197] Bioassayworks. E. coli o157:h7 rapid detection kit. https://bioassayworks.com/?product=e-coli-o157h7-rapid-detection-kit-20-tests, Accessed: 2022-02-22.

[198] D. Wang, J. Chen, and S. Nugen. Electrochemical detection of escherichia coli from aqueous samples using engineered phages. *Analytical Chemistry*, 89(3):1650–1657, 2017.

[199] Y. Zhao, D. Zeng, C. Yan, W. Chen, J. Ren, Y. Jiang, L. Jiang, F. Xue, D. Ji, F. Tang, M. Zhou, and J. Dai. Rapid and accurate detection of escherichia coli o157:h7 in beef using microfluidic wax-printed paper-based ELISA. *Analyst*, 145(8):3106–3115, 2020.

[200] L. Zheng, G. Cai, S. Wang, M. Liao, Y. Li, and J. Lin. A microfluidic colorimetric biosensor for rapid detection of escherichia coli o157:h7 using gold nanoparticle aggregation and smart phone imaging. *Biosensors and Bioelectronics*, 124-125:143–149, 2019.

[201] J. Kim and S. Oh. Rapid detection of e. coli o157:h7 by a novel access with combination of improved sample preparation and real-time PCR. *Food Science and Biotechnology*, 29(8):1149–1157, 2020.

[202] L. Heijnen and G. Medema. Method for rapid detection of viable escherichia coli in water using real-time NASBA. *Water Research*, 43(12):3124–3132, 2009.

[203] X. Xia, B. Zhang, J. Wang, B. Li, K. He, and X. Zhang. Rapid detection of escherichia coli o157:h7 by loop-mediated isothermal amplification coupled with a lateral flow assay targeting the z3276 genetic marker. *Food Analytical Methods*, 2021.

[204] J. Hu, Y. Wang, H. Su, H. Ding, X. Sun, H. Gao, Y. Geng, and Z. Wang. Rapid analysis of escherichia coli o157:h7 using isothermal recombinase polymerase amplification combined with triple-labeled nucleotide probes. *Molecular and Cellular Probes*, 50:101501, 2020.

[205] D. Garenne, S. Bowden, and V. Noireaux. Cell-free expression and synthesis of viruses and bacteriophages: applications to medicine and nanotechnology. *Current Opinion in Systems Biology*, 28, 2021.

[206] H. Yeom, T. Ryu, A. Lee, J. Noh, H. Lee, Y. Choi, N. Kim, and S. Kwon. Cell-free bacteriophage genome synthesis using low-cost sequence-verified array-synthesized oligonucleotides. *ACS Synth. Biol.*, 9(6):1376–1384, 2020.

[207] Q. Emslander, K. Vogele, P. Braun, J. Stender, M. Joppich, M. Abele, C. Meng, C. Ludwig, J. Bugert, C. Willy, F. Simmel, and G. Westmeyer. Personalized production, non-genomic engineering, and time-resolved proteomics of therapeutic phages for biohazardous and multidrug-resistant bacteria. *Cell Press*, 2022.

[208] C. Chang, P. Kemp, and J. Molineux. Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. *Virology*, 398(2):176–186, 2010.

[209] G. Gopal and A. Kumar. Strategies for the production of recombinant protein in escherichia coli. *The Protein Journal*, 32(6):419–425, 2013.

[210] Suthers P. You, L. and J. Yin. Effects of escherichia coli physiology on growth of phage T7 in vivo and in silico. *Journal of Bacteriology*, 184(7):1888–1894, 2002.

[211] K. Vogel, E. Falgenhauer, S. Schönberg, F. Simmel, and T. Pirzer. Small antisense DNA-based gene silencing enables cell-free bacteriophage manipulation and genome replication. *ACS Synthetic Biology*, 10(3):459–465, 2021.

[212] F. Guo, Z. Liu, P. Fang, Q. Zhang, E. Wright, W. Wu, C. Zhang, F. Vago, Y. Ren, J. Jakana, W. Chiu, P. Serwer, and W. Jiang. Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. *Proceedings of the National Academy of Sciences*, 111(43), 2014.

[213] E. Lenk, S. Casjens, J. Weeks, and J. King. Intracellular visualization of precursor capsids in phage p22 mutant infected cells. *Virology*, 68(1):182–199, 1975.

[214] N. Hawkins, J. Kizziah, J. Penadés, and T. Dokland. Shape shifter: redirection of prolate phage capsid assembly by staphylococcal pathogenicity islands. *Nature Communications*, 12(1), 2021.

[215] N. Marmiroli and E. Maestri. *Food Toxicants Analysis*. Elsevier, 2007.

[216] L. Alonzo, P. Jain, T. Hinkley, N. Clute-Reinig, S. Garing, E. Spencer, V. Dinh, D. Bell, S. Nugen, K. Nichols, and A. Ny. Rapid, sensitive, and low-cost detection of escherichia coli bacteria in contaminated water samples using a phage-based assay. *Sci Rep*, 12(7741), 2022.

[217] D. Hatlem, T. Trunk, D. Linke, and J. Leo. Catching a SPY: Using the spycatcher-spytag and related systems for labeling and localizing bacterial proteins. *International journal of molecular sciences*, 20(9):2129, 2019.

[218] H. Kim, Y. Jung, I. Doh, R. Lozano-Mahecha, B. Applegate, and E. Bae. Smartphone-based low light detection for bioluminescence application. *Scientific Reports*, 7(1), 2017.

[219] Hygiena. Ensure™ touch. https://www.hygiena.com/food-safety-solutions/atp-monitoring/ensure-touch/, Accessed: 2022-03-03.

[220] J. Chang, M. Schmid, C. Haase-Pettingell, P. Weigele, J. King, and W. Chiu. Visualizing the structural changes of bacteriophage epsilon15 and its salmonella host during infection. *Journal of Molecular Biology*, 402(4):731–740, 2010.

[221] C. Wang, J. Tu, J. Liu, and I. Molineux. Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. *Nature Microbiology*, 4(6):1049–1056, 2019.