Development of the PHEARLESS system and its application for directed protein mutagenesis and characterisation of putative endolysin activity

Hannah Rose Bonham

Supervised by Dr Keith Shearwin

Co-Supervised by Dr Nan Hao

Co-Supervised by Dr Fiona Whelan

THE UNIVERSITY of ADELAIDE

A THESIS SUBMITTED IN FULFILMENT OF THE

REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

University of Adelaide

Department of Molecular and Cellular Biology,

School of Biological Sciences, South Australia, Australia

November 2023

Contents

List of Tables....3

List of Figures....4

List of Appendix Figures....7

Abbreviations....8

Abstract....9

Declaration....11

Acknowledgments....12

Chapter 1: Introduction....13

1.a Antimicrobial Resistance - The Next Global Health Issue....13

1.b Antibiotic Resistance - The End of Antibiotics?....14

1.c Bacteriophage - A Therapeutic Alternative to Antibiotics?....20

1.d The Scope of this Thesis....33

Chapter 2: Materials and Methods....36

2.a Materials....36

2.b Basic Methods....52

2.c Cloning Strategy....54

2.d Bacteriophage Integration for making 186 lysogens....57

2.e Bacteriophage production and Integration....58

2.f Recombineering Strategy....59

2.g Integration of Plasmids at Phage Attachment Sites....59

2.h PHEARLESS Protein Screening Protocols....61

2.i PHEARLESS Protein Library Screening Protocols....62

Chapter 3: Design and construction of the PHEARLESS system....64

3.a Generating the PHEARLESS Expression Strain....64

3.b PHEARLESS V1 Expression Strain Background and variations....73

3.c Testing PHEARLESS Version 1, Using ClyF....77

3.d Engineered Bacteriophage 186 Functional test....85

Chapter 4: PHEARLESS V2 Assay....90

4.a PHEARLESS version 2 Background....90

4.b The Propagation Strain for Engineered 186 Propagation....98

4.c PHEARLESS V2.1 Expression Strain Development and Optimization....101

4.d PHEARLESS V2.1 Mutant Library Screening Proof of Principle 108
4.e Generation of an Inactive ClyF Protein 110
4.f Analysis of the Sensitivity of the PHEARLESS Protein Library Screening Assay 114
4.g PHEARLESS V2.2; Improved Version of the Expression Strain 120
4.h PHEARLESS Mutant Protein Library Screening using the PHEARLESS V2.2 Expression Strain, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G)) 137
Chapter 5: Bioinformatic Pipeline for Putative Antimicrobial Phage Protein Discovery 151
5.a PHASTER: Web-based Phage Search Tool 151
5.b Data Formatting and Domain Search 153
5.c *S. aureus* Genomes Collected from Patients with Chronic Rhinosinusitis Provided by Basil Hetzel Institute ENT Surgery Group 156
5.d *S. aureus* Genomes Collected from Patients with Fibrosis Lung Infections 157
5.e Domains Detected in the Bioinformatics 157
Chapter 6: Selected Proteins Tested for Antimicrobial Activity 172
6.a Proteins Investigated using the PHEARLESS System 172
6.b Putative Endolysin Genes Provided by the Basil Hetzel Institute 198
6.c Testing of chimeric endolysins 216
Chapter 7: Experimental Summary and Conclusions 228
7.a Discussion and Conclusion of the generation of PHEARLESS V1 expression strain and the Generation PHEARLESS Protein Screening Assay 228
7.b Discussion and Conclusion – Chapter 4. Generation of PHEARLESS Version 2.2 Expression Strain and its Application in Library Screening 229
7.c Discussion and Conclusion Chapter 5 - Bioinformatics Search and Screening of Putative Lytic Proteins in the PHEARLESS Assay System 232
Chapter 8: Future Directions 238
8.a.i Adapting the PHEARLESS Assay System for Liquid Culture Screening. 238
8.a.ii Adapting the PHEARLESS Assay System for Environmental Library Screening 239
Chapter 9: Appendix Data 241
References: 249

List of Tables

Table 1: Bacteria and Bacteriophage. 36
Table 2: Plasmid Construct Table of Content. 45
Table 3: Primers Used During the Course of this Thesis. 49
Table 4: Summary of the Different PHEARLESS V1 variants tested. 85
Table 5: Keyword Domain Name List from the domain search script. 154
Table 6: Basel Hetzel Institute - Phage Predictions and Protein database V3. 156
Table 7: Cystic Fibrosis Lung Infections - Phage Predictions and Protein database V2. 157

List of Figures

Figure 1: The Arms Race of Antibiotic Development and Antibiotic Resistance. 16
Figure 2: Snap Shot of the Driving Forces for Development and Spread of Antibiotic Resistance. ...19
Figure 3: General architecture exhibited by a phage that belongs to the Caudoviradae order of phage - Podoviridae (T7), Myoviridae (T4, 186) and Siphoviridae (SPP1, λ) phage. 23
Figure 4: The General Outline of the 186 Bacteriophage Lytic and Lysogenic Life Cycle.....25
Figure 5: Gram-Positive Bacterial Cell Wall and Peptidoglycan Layer With Locations of Catalytic Domain Targets. 27
Figure 6: Bacteriophage Endolysin Architecture. 31
Figure 7: PHEARLESS Expression Strain Modules. 65
Figure 8: Activation of PHEARLESS V1 endolysin expression and cell lysis (Plasmid Version). 68
Figure 9: Map of the Wild type 186 Prophage genome and Engineered 186 Prophage Genome.....69
Figure 10: Empty tum Expression Stain (HB66) Liquid Culture Lysis Assay. 71
Figure 11: Tum72 Expression Stain (HB67) Liquid Cultures Lysis Assay. 72
Figure 12: Tum Wild Type Expression Strain (HB70) Liquid Cultures Cell Lysis Assay. 73
Figure 13: Plasmid Maps of Protein Expression Plasmids.....76
Figure 14: PHEARLESS Protein Screening Assay: Comparing Promoters in the Tum Wild Type Expression Strain on Low-Copy Number Expression Plasmids. 78
Figure 15: PHEARLESS Protein Screening Assay: Comparing Promoter Types in the Tum72 Expression Strain with Low-Copy Expression Plasmids. 80
Figure 16: PHEARLESS Protein Screening Assay: Comparing Low Copy Number and High Copy Number 186pJ Expression Plasmids in the Tum wild type lysis strain. 82
Figure 17: PHEARLESS Protein Screening Assay: Comparing the Low Copy and High Copy Number 186pJ Expression Plasmids in the Tum72 lysis strain.....83
Figure 18: Testing for Production of Functional Phage Particles.....87
Figure 19: Testing for Production of Functional Phage Particles from Engineered 186 cos^{+} Prophage 88
Figure 20: PHEARLESS version 2 design, part 1. 93
Figure 21: PHEARLESS version 2 design part 2. 96
Figure 22: PHEARLESS version 2 design, continuous expression though phage propagation. 96
Figure 23: Plasmid Maps for the Three Variants of the Protein Integration plasmid (pIT4). 98
Figure 24: Determining Optimal Propagation Strain to Target Strain Ratios.....100
Figure 25: Phage functional tests after integration of the pIT4 integration plasmids into the Tum wild type and Tum72 lysis strains.....102
Figure 26: PHEARLESS Endolysin Screening Assay: Comparing Promoters in the Tum Wild Type Lysis Strains with integrated pIT4 ClyF Plasmid.....103
Figure 27: PHEARLESS Endolysin Screening Assay: Comparing Promoters in the Tum72 Wild Type Lysis Strains with integrated pIT4 ClyF Plasmid. 105
Figure 28: PHEARLESS Protein Screening Assay: Comparing the Tum lysis and Promoter-ClyF Expression Strain Variants on Plates Containing both Target and Propagation Strains.....107
Figure 29: ClyF Inactive Mutants Constructs Sequences and 3D Structure Modelling of ClyF. 111
Figure 30: PHEARLESS Protein Screening Assay in the PHEARLESS V1 Expression System: ClyF Mutants Tested for Antimicrobial Activity. 113
Figure 31: PHEARLESS Protein Screening Assay in the PHEARLESS V1 Expression System: ClyF(C36G) Mutant Tested for Antimicrobial Activity. 114
Figure 32: Sensitivity of the PHEARLESS Protein Screening assay.....119
Figure 33: Assay Sensitivity Estimations for the PHEARLESS Protein Library Screening Assay using the PHEARLESS V2.1 Expression Strain.....119

Figure 34: Map of modified pIT4 plasmids that contain the 186 cos sequence and the integration map for the 186 genome. 122
Figure 35: Recovery of Phage Production by Re-introducing the 186 cos Sequence into the PHEARLESS Expression Strain. 124
Figure 36: Preliminary Gibson Assembly Assay: Testing Number of Fragments used in Gibson assembly for Protein Library Screening Protocols. 126
Figure 37: Transformation and Integration Efficiency Assay for each step. 130
Figure 38: Maps for pIT4 plasmids used to increase the 186 genome size. 133
Figure 39: Phage Functional Assay: Testing Phage Propagation for Phage with Increased 186 Genome Size. 135
Figure 40: Production of Functional ClyF Protein From Strains With Increased 186 Prophage Genome Size. 137
Figure 41: Drip Assay Showing Transformation and Integration Efficiency of the ClyF(C36G) Gibson Assembly Library. 140
Figure 42: PHEARLESS Mutant Protein Library Screening using the PHEARLESS V2.2 Expression Strain, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G)). 141
Figure 43: PHEARLESS Mutant Protein Library Screening, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G)), Phage Isolation Plate 1. 142
Figure 44: PHEARLESS Mutant Protein Library Screening, PCR amplification of the mutant ClyF gene from phage stocks and Sanger Sequencing results. 143
Figure 45: Repeated and Optimised PHEARLESS Mutant Protein Library Screening using the PHEARLESS V2.2 Expression Strain, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G)). 145
Figure 46: Repeated and Optimised PHEARLESS Mutant Protein Library Screening, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G)), Phage Isolation Plates. 148
Figure 47: Repeated and Optimised PHEARLESS Mutant Protein Library Screening, Sanger Sequencing results. 150
Figure 48: PHASTER (PHAge Search Tool Enhanced Release) (168,169)**; Web based application for identification and annotation of prophage sequences within bacterial genomes and plasmids**. ...152
Figure 49: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_BHI_P1, chosen from the BHI bioinformatics search. 174
Figure 50: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_BHI_P2, chosen from the BHI bioinformatics search. 176
Figure 51: Putative phage protein (P3) with predicted antimicrobial activity against S. aureus, PE_BHI_P3, chosen from the BHI bioinformatics search. 180
Figure 52: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_GDT_P4, chosen from a separate bioinformatics search by Oliveria and Sampaio (3). 182
Figure 53: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_GDT_P5, chosen from a bioinformatics search by Oliveria and Sampaio (3). 185
Figure 54: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_GDT_P6, chosen from a separate bioinformatics search by Oliveria and Sampaio (3). 187
Figure 55: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_ CFLI_P7, chosen from the CF bioinformatics search. 190
Figure 56: Twist Biosciences pTwist Cm High copy number Expression Plasmid Containing the PE_GDT_P5 gene controlled by the 186pJ promoter. 193
Figure 57: PHEARLESS Protein Screening Assay using the PHEARLESS V2.2 Expression Strain, Testing Seven Putative Proteins for Antimicrobial Activity Against S. aureus Strain RN4220. 195
Figure 58: PHEARLESS Protein Screening Assay using the PHEARLESS V2.2 Expression Strain, Testing Seven Putative Proteins for Antimicrobial Activity Against S. aureus Strain HER1049. 196

Figure 59: Times Lapse of PHEARLESS Protein Screening Assay using the PHEARLESS V2.2 Expression Strain, Testing PE_GDT_P5 for Antimicrobial Activity Against S. aureus Strain RN4220.....198
Figure 60: Domain Maps for Endolysin Proteins Discovered by the ENT surgery group at the Basil Hetzel Institute. 201
Figure 61: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain C319. 206
Figure 62: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain ATCC 25923......208
Figure 63: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain C43. 210
Figure 64: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain 259. 212
Figure 65: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain C330. 214
Figure 66: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain C244. 216
Figure 67: Chimeric Endolysin Protein Maps......218
Figure 68: Crystal Structures and Secondary Structure Predictions of Staphylococcal phage 2638A endolysin Peptidase_M23 and SH3 domains. 219
Figure 69: Secondary Structural predictions for the Linker Sequences of the Parent Proteins and Chimeric Protein......221
Figure 70: PHEARLESS Protein Screening Assay using the PHEARLESS V2.2 Expression Strain, Testing the Chimeric Endolysins and Catalytic Domains for Antimicrobial Activity Against S. aureus Strain RN4220......223
Figure 71: PHEARLESS Protein Screening Assay using the PHEARLESS V2.2 Expression Strain, Testing the Chimeric Endolysins and Catalytic Domains for Antimicrobial Activity Against S. aureus Strain RN4220......226

List of Appendix Figures

Appendix Fig. 1: G-Block version 1, 186 recombineering fragment……241
Appendix Fig. 2: G-Block version 2, 186 recombineering fragment……243
Appendix Fig. 3: Results for the PCR amplification of correct recombineering fragment present in the 186 genome, strain HB61……244
Appendix Fig. 4: Results for the PCR amplification of correct pIT4 plasmid integration into the 186 genome, for expression strains HB143, HB149 and HB152……245
Appendix Fig. 5: Results of the PCR amplification of pIT4 plasmid integration into the 186 genome, for each of the putative proteins investigated in Chapter 6.a……247

Abbreviations

RDF – Recombination Directionality Factor

S. aureus – *Staphylococcus aureus*

CWB – Cell Wall Binding

TSS – Transformation & Storage Solution

SF-GFP – Superfolder Green Fluorescent Protein

PCR – Polymerase Chain Reaction

CTH – C-terminal Histidine Tag

NTH – N-terminal Histidine Tag

Amp – Ampicillin

Cm – Chloramphenicol

Kan – Kanamycin

Spec – Spectinomycin

bp – Base Pair

Kbp – Kilobase Pair

MW – Molecular Weight

ECC – Electrocompetent Cells

TSS – Transformation and Storage Solution

AGRF – Australian Genome Research Facility

cfu – Colony-Forming Unit

CF – Cystic Fibrosis

SH3 – Src Homology-3

PG – Peptidoglycan

PHEARLESS – Phage-based Expression, Amplification and Release of Lytic Enzyme Species.

MRSA – Methicillin-resistant *Staphylococcus aureus*

MSSA – Methicillin-sensitive *Staphylococcus aureus*

TSB – Tryptic Soy Broth (TSB)

BHI – Basil Hetzel Institute

TM – Phage Storage Solution

Abstract

Bacteriophages (phage) are considered a primary resource for developing new therapeutic treatments for bacterial infections. Phage will likely be used either as an alternatives to antibiotics or in cooperative treatment with antibiotics. However, a mixed blessing of phage and phage proteins is that they are highly selective for their target bacteria. The advantage of this is that the natural bacterial fauna of the patient is undisturbed. Future therapeutic phage and phage-derived protein treatment for bacterial infections will likely be administered as a cocktail, of several phage or phage-derived proteins. More phage and phage-derived proteins with high antimicrobial activity and suitable pharmacokinetics must be identified to develop these cocktails.

Current methods for identifying phage-derived proteins include their purification and testing against target bacteria. However, many phage-derived proteins have properties that make them difficult to purify, including limited solubility. The inability to efficiently and cost-effectively purify these proteins means that many potentially effective antimicrobial proteins are overlooked. The **PHEARLESS** (**Ph**age-based **E**xpression, **A**mplification and **R**elease of **L**ytic **E**nzyme **S**pecie**s**) assay system developed here consists of a bacterial expression strain where both the expression of the protein of interest and cell lysis to release that protein are controlled via a simple chemical induction. What makes this system unique is the coordination of two modules, one controlling expression of the protein of interest and the second a biological switch modified for cell lysis. The system can test any protein expressed in *E. coli* against any target bacterial species that can be co-plated with *E. coli*, without a requirement for protein purification. The system provides for a simple and rapid method to test for antimicrobial activity.

The *E. coli* expression strain carries a bacteriophage 186 prophage that has been genetically engineered to take control of its lytic cycle via a separate chemically-inducible module integrated elsewhere in the bacterial genome. Cumate (*p*-isopropyl benzoate) was used as the chemical inducer, and allowed tight control of the activation of the prophage lytic cycle. The first version of the PHEARLESS system was developed using a multicopy plasmid to express the antimicrobial protein of interest. Expression of the protein of interest was placed under control of a bacteriophage 186 late promoter, thus linking gene expression and cell lysis. This

plasmid-based system allows rapid assessment of antimicrobial activity of any gene cloned into the expression plasmid.

A second variant of the system was developed to facilitate recovery of the gene of interest following cell lysis. In this case, the plasmid carrying the gene of interest has been site-specifically integrated into the phage 186 prophage genome, such that cumate-induced cell lysis also leads to phage production, where every phage particle carries a copy of the gene. Continued progeny phage propagation through the inclusion of a sacrificial propagation strain ensures continued protein expression. This version of the PHEARLESS assay system it is also suitable for mutant library screening. A proof of principle test confirmed that the PHEARLESS assay system is capable of high throughput screening of a library of mutants, using a mutated endolysin protein, and screening for revertants with restored activity against *S. aureus*.

Bioinformatic searches of *S. aureus* genomic sequences for putative phage-associated endolysins provided a list of possible anti-microbial enzymes, and a small sample of these were tested for activity using the PHEARLESS system. Similarly, the system can be used for testing engineered variants of known endolysins and the impact of protein linker sequences was examined.

In summary, the tools developed here will help discover new proteins for anti-microbial therapeutics and can be employed to optimize and investigate their chemical properties.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Hannah Bonham 10th of November 2023

Acknowledgments

My thesis and achieving a PhD would have never become a reality if not for the continued support from many different people in my life. In my professional life, I had a massive amount of support from my supervisor, Dr Keith Shearwin, who gave me much encouragement and compassion during my PhD. As well as my co-supervisor, Dr Nan Hao and Dr Fiona Whelan, and all my colleagues in the Shearwin laboratory and colleagues at Adelaide University. These people provided me guidance and friendship throughout my academic journey. In my personal life, I had support from my family: my father, Robert Bonham; mother, Megan Bonham; my sister, Alexandra Bonham; and partner at the time, Blake Luetkens. They provided me with the support that made it feel like I wasn't walking this path alone.

Throughout my PhD, I have gained confidence through my achievements and self-reflection from my shortcomings. While I never conquered my dyslexia, I feel as if I have reached a point where it will not inhibit me from attempting to reach my goals. I may still fail, but it is better to try and fail than never try to begin with, and okay if I fail.

In addition to all the skills I gained throughout my PhD, I am also grateful for the maturity I gained, not just from my PhD but from the life events that occurred during. I survived the death of a parent, sudden serious health complications of a loved one, dealing with depression and anxiety and the end of a long-term relationship. I went through all these things, and while I can still improve myself in the future, I managed to survive somehow.

I survived a PhD, and I can literally do anything now.

Chapter 1: Introduction

1.a Antimicrobial Resistance - The Next Global Health Issue

The greatest evidence of evolution can be observed in microorganisms. Microorganisms, including parasites, fungi, bacteria and viruses can all become resistant to drug treatments following repeated exposure. These microorganisms can then grow and spread uninhibited by the medications that they were once susceptible to. To make matters worse, they can share genes responsible for this resistance. Antimicrobial resistance has led to an increase in economic costs for the individual, and for healthcare systems as well as major loss of life (1).

Some of the first noticeable costs caused by antimicrobial resistance include extended periods of illness, with longer hospital admissions due to failed treatments (2). If the initial antibiotic/antimicrobial treatments fail to clear the infection, second-line and last-resort antibiotics become required. This can introduce additional risks to the patients, since many of these antibiotics, such as vancomycin and teicoplanin, have poor safety profiles and require careful monitoring for severe side effects (3). Vancomycin is largely considered a last-resort antibiotic for multidrug-resistant, Gram-positive bacterial infections. Although created over 50 years ago, exact dosing methods have still not been determined for this drug (4,5). With potential severe side effects such as damage to kidneys, vision loss and hearing loss, treatment must be closely monitored (4,5).

The costs of increased hospital stays and failed treatments create a major economic burden for countries fighting antimicrobial resistance, in both developed and underdeveloped countries. In Europe alone, it has been estimated that an estimated nine billion € has been spent since 2014 to fight antimicrobial resistance (6). The USA alone has approximately two million people infected with antibiotic resistant bacteria annually, resulting in an estimated 23,000 deaths (1,7). The inclusion of antimicrobial resistant fungi increases the numbers of people infected annually to three million, resulting in an estimated 35,000 deaths (8). It is difficult estimate the world wide cost of all antimicrobial resistance. Several studies estimate that it could cost anywhere from three hundred billion to over one trillion USD annually by 2050 (2). The estimated number of deaths caused by antimicrobial resistance is also estimated to rise to over ten million people annually by 2050 (2,9).

With expectations of increasing morbidity and mortality, vulnerable individuals with already compromised immune systems are at exceptional risk (1). Cancer patients are at particular risk of antibiotic resistant bacteria, due to cancer treatments weakening the immune system (9,10). Current infections also delay the time of cancer treatment, since any infection, resistant or not, has to be cleared before treatment can start (11).

Transplant patients are another group that are at high risk from antibiotic resistant bacteria. Bacterial infections are the leading cause of morbidity and mortality in solid organ transplant recipients (12). Hence, transplant recipients are highly likely to be prescribed a broad-spectrum empiric antibiotic therapy, which increases the risk of development of antibiotic resistance strains.

1.b Antibiotic Resistance - The End of Antibiotics?

1.b.i A Brief History of Antibiotics and the Rise of Antibiotic Resistance

Bacteria are particularly able to adapt to their environment to improve their odds of survival. Antibiotic and antimicrobial resistance is not a phenomenon caused by developments in medicine but rather is simply a naturally occurring response to selective pressure. Bacteria can gain antibiotic resistance in two ways (13–16); gaining access to new genetic material that can be utilised by the bacterial genome via addition to the genome or cell, or by mutations to the pre-existing genome, though base changes to protein coding genes or regulation sequences. Mutations to the genome occur when bacteria have increased exposure to agents that cause DNA damage. While the majority of the bacteria exposed to antibiotics will perish, a strong selection pressure is created for bacteria to adapt in some way to overcome the presence of the chemical.

The history of antibiotics is a well-documented one. Penicillin was discovered fortuitously by Sir Alexander Fleming in 1928 and released to the public in 1945 (17). The discovery of penicillin started a push to discover more antibiotic drugs. By 1937, new antimicrobials, specifically sulfonamides, were developed for therapeutic use. However, the discovery of these antibiotic therapeutics inevitably also pushed the development of antibiotic resistance in bacteria. Before the release of penicillin to the

public had even occurred, bacteria carrying a resistance gene (penicillinase) were identified (18).

Penicillin is classified as a β-lactam antibiotic, which is a class of antibiotics that contain a β-lactam ring. β-lactam antibiotics are one of the most commonly used class of antibiotics, which function by inhibiting peptidoglycan synthesis (19). Used to treat both Gram-positive and Gram-negative bacteria, the mechanism of action is via interruption of the terminal transpeptidation process by inhibiting the catalytic activity of bacterial transpeptidases (20). This results in the loss of the cross-linking peptides that form the peptidoglycan, weakening the structure and leading to cell lysis (19,20).

Penicillin was the first β-lactam to be discovered and was quickly followed by many others, including; aminopenicillins, carboxypenicillins, ureidopenicillins, carbapenems, cephems, cephalosporins, cephamycins, carbapenems many other derivatives (Figure 1). β-lactamase is an enzyme discovered in bacteria that a provides resistance to β-lactam antibiotics, via splitting the amide bond of the β-lactam ring. Many variants of β-lactamase genes have been identified since penicillin was first released. Between 1970 and 2015, over 1000 unique β-lactamase enzymes have been described (18).

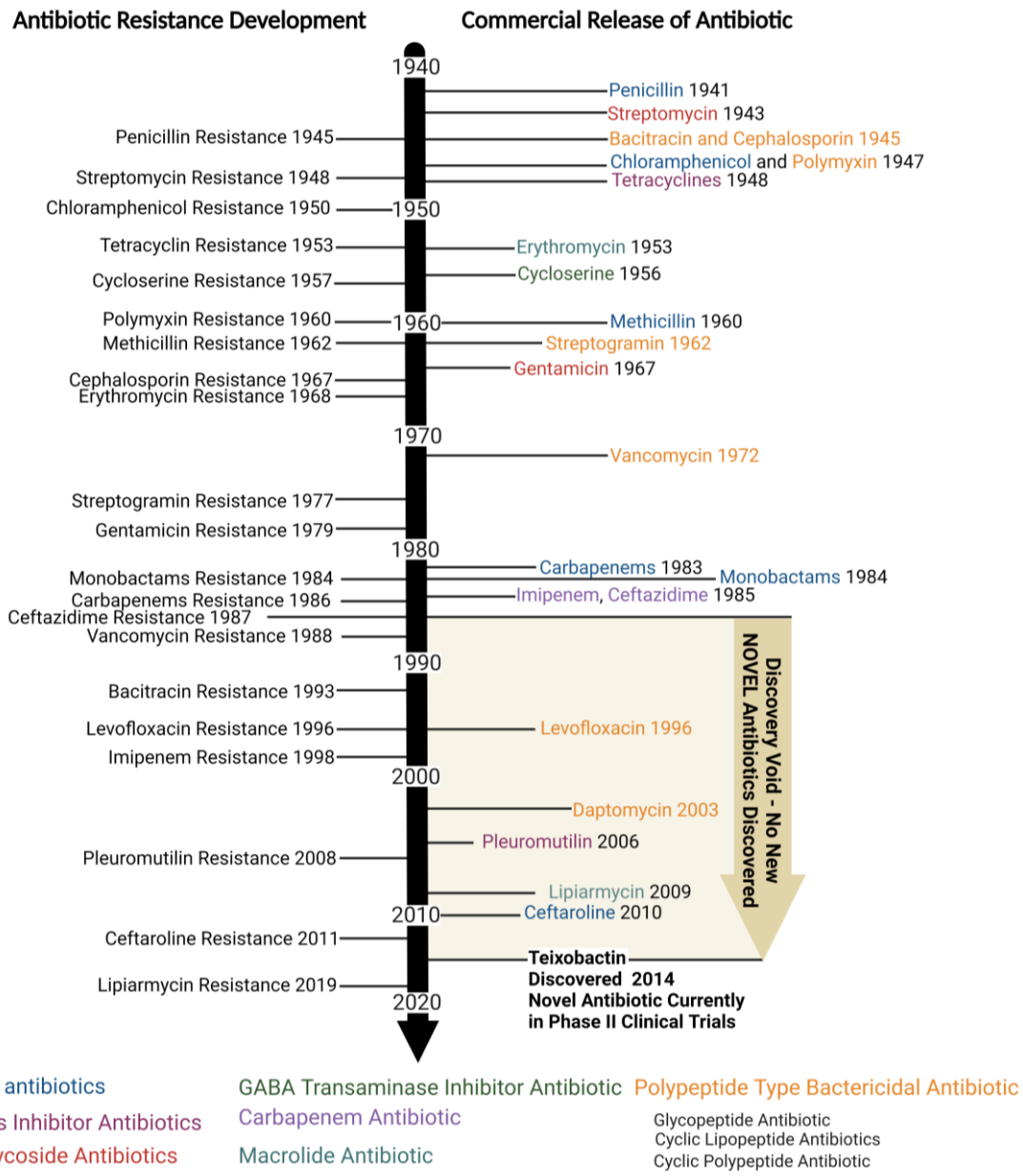


Figure 1: The Arms Race of Antibiotic Development and Antibiotic Resistance. *A general timeline for the commercial release of a number of different antibiotics and the corresponding development of bacterial antibiotic resistance. The general class of antibiotics is located at the bottom of the figure with the antibiotics colour coded. The commercial release date and the date of the first recorded example of resistance are shown. The time frame of the discovery void is also indicated (3,21–23). Teixobactin is a new class of antibiotic currently still currently in clinical trials (24). Figure adapted from; Figure 1 from ANTIMICROBIAL RESISTANCE Global Report on Surveillance (3), Table 1 from Derivation of a Precise and Consistent Timeline for Antibiotic Development (23) and Figure 1 from Antibiotics: past, present and future (25), Image was created with clipart from BioRender.com.*

1.b.ii Drivers for Antibiotic Resistance Development

The strongest drivers for the development of antibiotic resistance are the overuse and consumption of antibiotics in our healthcare system and in agriculture (26) (Figure 2). Overuse and misuse of prescription antibiotics in the medical profession is a key driver for developing antibiotic resistance and tolerance in disease-causing bacteria. The inappropriate use of antibiotics includes the prescription of antibiotics for viral infections, failure of the patient to complete the antibiotic regimen, and the prescription of inappropriate antibiotics. Coordination of management and reporting the identification of antibiotic resistance differs from country to country. The lack of a universal system makes it difficult to track the rates of development and the trends of the emergence of antibiotic resistant bacteria.

The second major driver for the development and transfer of antibiotic resistance is the overuse of antibiotics in agriculture and animal husbandry in some countries. Antibiotics are commonly used in both the developed and developing world as growth supplements in livestock (21). In 2015, 80% of antibiotics sold in the USA were used in livestock (27). Efforts have since been made to decrease the total amount of all antibiotics used in animal husbandry, leading to a general decrease of up to 36% by 2019 (28). However the same report reveals that the distribution of specific medically important antimicrobials for livestock increased by 3% from 2018 to 2019 (28).

Overuse also creates the issue of antibiotics approved for livestock becoming ineffective, thus promoting more antibiotic use and increasing humans' direct exposure to antibiotics in the food chain, and also to bacteria that are resistant to antibiotics. Antibiotics are further released into the environment through the waste products from intensive animal husbandry. It has been estimated that 90% of the antibiotics given to livestock end up in surface runoff, groundwater and in fertilizers produced from the waste products (29). Humans are also exposed to antibiotics from other aspects of agriculture, in global crop production. Antibiotic resistance genes have been identified in a number of plasmids found in plant pathogens (30). At the beginning of 2019, antibiotic use on crops in first world countries has been prohibited or tightly regulated due to the risk of antibiotic resistance development. However, in extreme cases large quantities of antibiotics have been approved for use, this includes as recently as 2019 where the US Environmental Protection Agency temporarily approved for a single season the use of 6.3 tons of aminoglycosides (broad-spectrum antibiotics) for 764,000 acres in farms located in citrus-producing states like California and Texas (31).

Another environment giving rise to selection pressure for antibiotic resistance development is one that is perhaps less obvious - the inappropriate disposal of human waste containing antibiotic drugs. Antibiotics that humans consume at home eventually end up in the wastewater stream. Depending on the local regulations, this waste may end up at a wastewater reservoir or treatment facility. Such locations are the perfect environment to drive mutation and horizontal gene transfer (32). They represent diverse and complex microbial communities including biofilms. These microbial “soups” are also under tremendous selection pressure from many types of pollutants including antibiotics, pharmaceuticals, chemical compounds and heavy metals (32). Similar environments and selection pressure can be found in other locations such as landfills (33), where unused or expired antibiotics may be disposed of.

The last important contributor of antibiotic resistance, while not a direct driving force, is the lack of development of antibiotics with novel mechanisms of action (1,34). A large portion of new antibiotics released for use are simply derivatives of current antibiotics (35). This results in a number of bacteria already being resistance or partially resistant against these new antibiotics, even before they have been released. Unfortunately, for large pharmaceutical companies antibiotic development is not as lucrative as other therapeutics, leaving research pipelines to smaller groups that do not have the capital for long term research and development (R&D) (1,36). While R&D is expensive, even more so when searching for and developing drugs with novel mechanisms, companies often spend more on marketing than R&D (36). Even with support from different government or charitable groups, such as the CDC or The Pew Trust, the antibiotic development pipeline still deals with a number of setbacks. The Pew Charitable Trust (37) estimated that in 2017 there was at least 39 antibiotics currently progressing through clinical trials, in Phases I to III. A more recent analysis showed that of these 39 potential drugs, only 13 (all derivatives of existing antibiotics) have made it to market (38).

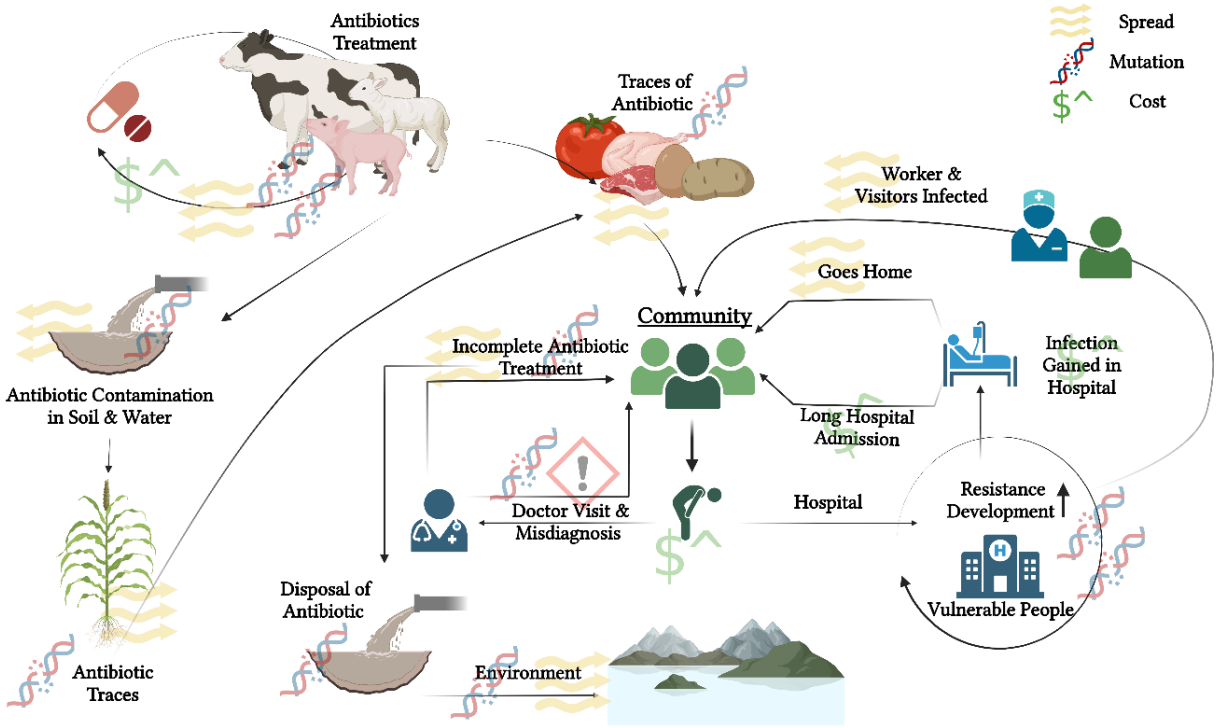


Figure 2: Snap Shot of the Driving Forces for Development and Spread of Antibiotic Resistance.

Figure shows a general flow chart of how bacteria can develop resistance to antibiotics and factors which allow the spread of that resistance. Image was created with clipart from BioRender.com

1.b.iii Mechanisms of Antibiotic Resistance

Gene/s that code for resistance can be complex. Thus, while bacteria may develop resistance while a patient is taking antibiotics, it is far more common for bacteria to have acquired pre-existing resistance genes from a secondary source (13,39). Where the pathways already exist in nature, and it is the transfer of these genes into pathogenic bacteria that makes them dangerous (40). Bacteria have evolved multiple mechanisms for acquiring new DNA from the environment and from other bacteria (horizontal transmission). Collectively referred to as mobile genetic elements, these elements include: plasmids, bacteriophage (phage), transposons, integrons, gene cassettes and pathogenicity islands (15,18,41).

Horizontal gene transfer is the transfer new genetic material sourced from extracellular DNA fragments, phage and plasmids, and occurs through three primary mechanisms; natural transformation, conjugation and transduction (42).

Natural transformation is the process of a bacteria taking up extracellular DNA and incorporating as part of the host genome (43). Transformation does not require actions from the donor DNA, but simply access to free DNA, with the recipient cell

already containing all the required proteins for transformation (43). Single or double stranded DNA can be taken up, converted to single stranded DNA which is then assessed for homologous sequences (44). If homology with the single stranded DNA is present, it can be inserted into the genome using homologous recombination (43). The mechanisms differ between Gram-positive and Gram-negative bacteria, with many details of the processing machinery and conditions for DNA uptake being unknown. A number of pathogenic bacterial species have been shown to be adept at taking up foreign DNA, including *Staphylococcus*, *Streptococcus*, *Acinetobacter*, *Neisseria* and *Pseudomonas* (1,43).

Conjugation is another major mechanism of DNA transfer between bacteria, with DNA transfer between cells via direct physical contact (45,46). The mechanisms for DNA transfer are complex and can differ between bacterial strains (46). One mechanism of DNA transfer in Gram negative bacteria is mediated by the type IV secretion system, a secretion system involved in substrate transport and pilus biogenesis (14,46). Plasmids are transferred from one bacteria to another using conjugation and play a major role in the transfer of antibiotic resistance genes (47).

The third mechanism of DNA transfer, transduction, is the transfer of viral (phage) DNA, often also carrying bacterial genes, to a bacterial host. Phage are viruses capable of integrating their genome into their bacterial host, using the host cell machinery for reproduction (48–50). While phage can be viewed as parasitic in nature, in some cases phage infection can be beneficial for its bacterial host. Phage are capable of incorporating bacterial genes from their previous host and transferring them to new host, which can be beneficial to the new bacterial host (51). Clinical strains of *S. aureus* have been shown to gain antibiotic resistance from the transfer of gene through phage transduction (42,52).

1.c Bacteriophage - A Therapeutic Alternative to Antibiotics?

With an increased awareness of the threat antibiotic resistance presents to human health, steps are being taken to combat its development and spread (1,11). But these measures may only slow, rather than prevent, the transmission and spread of antibiotic resistance in bacteria. This is because novel antibiotic resistance mechanisms will continue to evolve. Compounded by the fact that the number of antibiotics being released with novel mechanisms of action has been declining, we are getting very close to the end of the antibiotic golden age.

This realisation has pushed the search for alternative treatments for bacterial infections. There are a number of different approaches currently moving through clinical trials (53). These alternatives include; antibodies, probiotics, lysins, immune stimulation, vaccines, bacteriophage and bacteriophage products. In this thesis I will be focussing on phage and phage-based products as an alternative to conventional antibiotics.

The first evidence for the existence of phage and their antimicrobial properties was uncovered by British bacteriologist, Ernest Hankin, in the waters of the Ganges and Jumna rivers in India in 1896, where activity against *Vibrio cholera* was observed (54). The first "official" published discovery/rediscovery of phage did not occur until 1910 by Felix d'Herelle, a microbiologist working at the Institute Pasteur in Paris. Not long afterwards, in 1919, phage samples were used as a therapeutic to treat a 12 year old boy for dysentery at Hospital des Enfants-Malades (54). The treatment was successful with the patient recovering after a single dose of d'Herelle's antidysentery phage. After the successful trial, three additional patients were treated for dysentery, each showing improvement within 24 hours of being treated. Unfortunately, these results were not published immediately, therefore the first official recorded application of phage as a therapeutic treatment was in 1921 by Richard Bruynoghe and Joseph Maisin, who used phage to treat *Staphylococcal* skin disease (54). Just before the commercialization of antibiotics in 1945, therapeutic phage where commercially produced by a company in the United States (54). Several different products were produced to treat different bacterial infections including *Escherichia coli, Streptococci* and *Staphylococci*. However, after the production of penicillin, phage therapeutic treatment where largely discontinued in the West. Although there was a decline in research of phage therapy and phage therapy safety in western countries, countries like Georgia, Poland and Russia have continued their research (55,56). The Hirszfeld Institute of Immunology and Experimental Therapy at the Polish Academy of Sciences has continued in investigating phage therapy since its founding in 1949. While a lack of funding has prevented progression through clinical trials great steps have been made at demonstrating the safety of phage therapy and generating a library of phage candidates for treatments (56).

1.c.i What Are Bacteriophage and how can they be used?

What makes phage so attractive as an alternative to antibiotics? Phage are viruses that can infect bacteria and kill them as part of their reproductive life cycle. They are the most abundant entities on earth (estimated numbers of 10^{31}). Phage are incredibly diverse, using either single or double stranded DNA or RNA, with circular or linear genomes, ranging in genome sizes from the small 14.2 kb *Rhodococcus* Phage RRH1 (57) to *Lysinibacillus* Phage G 626 kb (58).

The type of phage that are of most interest for development into therapeutic treatments, are the lytic, tailed phage that belong to the *Caudovirales* order. The *Caudovirales* class is made up of 13 different families, including the tailed phage that belong to the *Podoviridae*, *Siphoviridae* and *Myoviridae* families (59–61) (Figure 3). This classification scheme will change in the future since as of 2023 the morphology-based taxa used to classify phage has been abolished and along with the *Caudovirales* order (62). The families belonging to that order will be classified as undefined until further classifications using the new guidelines are developed. Since the change/modification to the classification of phage species has only been recently announced, this thesis will be continuing to use the previous classification for the order of *Caudovirales.*

Phage morphology and phage structural proteins has been studied extensively by both electron microscopy and X-ray crystallography approaches, providing a number of well-defined three-dimensional structures (63–65). Depending on their classification, phage particles of the *Caudovirales* order can be considered to have three regions; head (1), tail (2) and tail tip (3) (Figure 3). For the purposes of this study only a general explanation of the overall structure of the phage is going to be addressed.

The head region (1) consists of an icosahedral capsid head structure that can range from 40 nm to 180nm in dimeter (66). The capsid is a hollow structure responsible for containing the phage genomic material along with any proteins required for infection. The head structure is connected to the tail structure via a head to tail connector (2). The tail section commonly contains a core tube in the *Myoviridae* and *Siphoviridae* family, surround by a sheath in the *Myoviridae* family. Depending on the phage species the tail structure can also contain a baseplate. The tail length can vary from short, to medium, to long for the *Podoviridae, Myoviridae* and *Siphoviridae* respectively (67,68) (Figure 3). Assembly of the tail structure will either occur after

head formation in the *Podoviridae* family, or as separate structures in the *Myoviridae* and *Siphoviridae* families. When the tail and head are assembled as separate structures, the tail is attached via the neck proteins to the head following DNA packaging into the head (68). Finally, the tail tip region is located at the base of the tail (3). This region contains proteins that can include; distal tail proteins, receptor binding protein, tail fibres (69,70). This region of the phage is responsible for recognizing and attaching to the host bacteria. After recognition, the phage genome is injected into the host bacteria to begin life cycle (48). The complexity of the tail tip of a phage structure can differ greatly depending on the family the phage belongs to. Further explanation of the *Myoviridae* family structural components is given below.

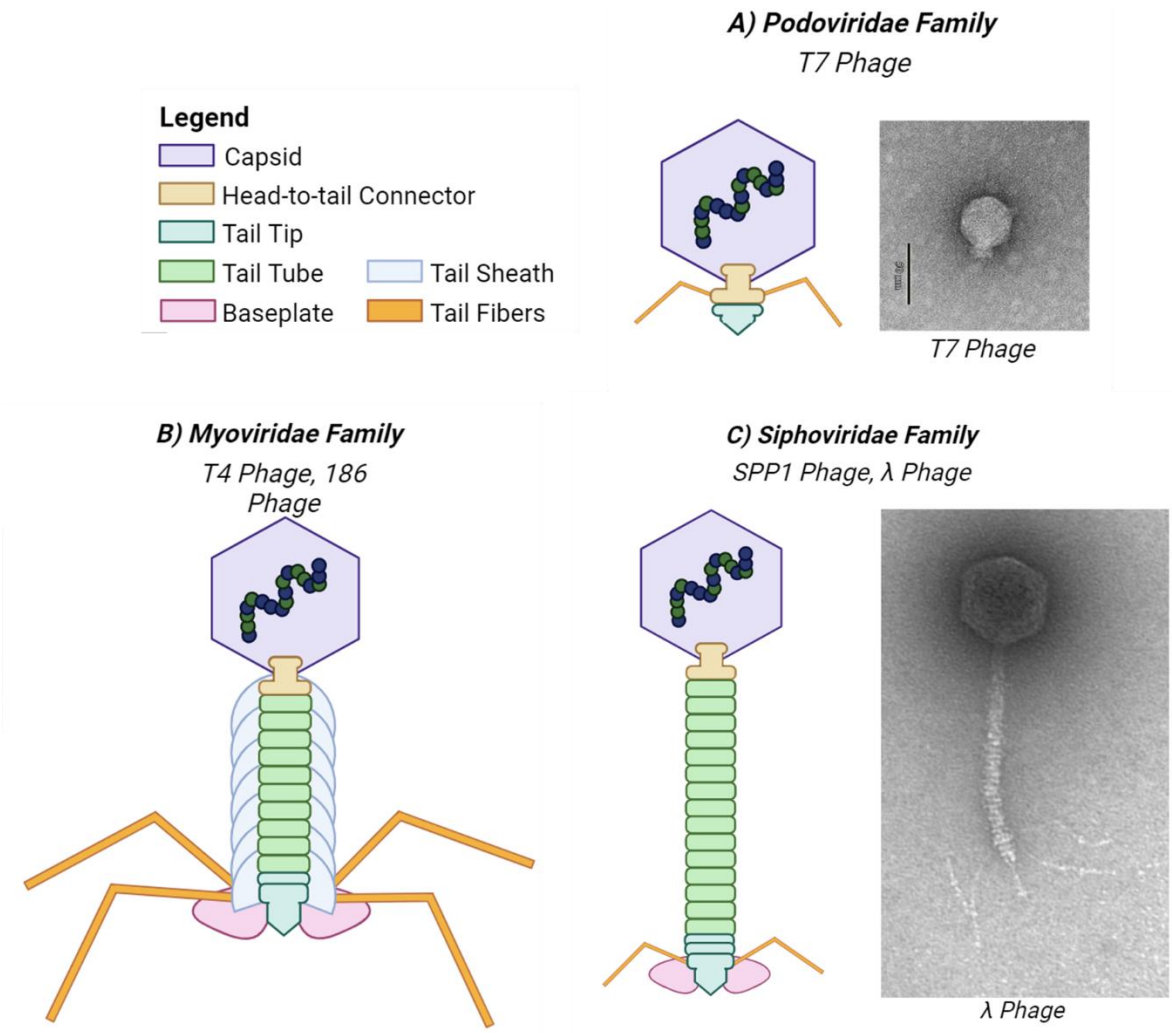


Figure 3: General architecture exhibited by a phage that belongs to the Caudoviradae order of phage - Podoviridae (T7), Myoviridae (T4, 186) and Siphoviridae (SPP1, λ) phage. *A) Phage belonging to the Podoviridae family like the T7 phage consist of a capsid (purple) connected a tail tip (aqua) via a head to tail connector (wheat colour). Phage of the Podoviridae family can also carry small tail fibers (orange). Second image is the transmission electron micrograph image of T7-family phage called SFPH2. EM mage is soured from* (71)*. B) Phage belonging to the Myoviridae family, for*

example T4 and 186, also consist of a capsid structure (purple) but between the head-to-tail connector (wheat colour) and tail tip (aqua) is a tail tube (green) structure. This structure can also be surround by a tail sheath (light blue). At the bottom of the sheath and tail tube is the tail tip that is surrounded by a baseplate (magenta), with tail tip structures (orange) attached. The baseplate and tail sheath is capable of contracting, pushing the tail tip and tail tube down through the outer walls of a bacteria during infection (70)*. Second image is of unpublished 186 phage particle taken using cryoTEM by Fiona Whelan at the University of Adelaide. C) Phage belonging to the Siphoviridae family for example SPP1 and λ, consists of a capsid (purple) containing the phage DNA genome (dark blue and green circles) connected to head-to-tail connector (wheat colour) in turn is connected to a long tail tube (green) structure. Some Siphoviridae phage will also include additional structures like the tail fiber (orange) structures (λ)* (72)*. Second image is an electron microscopy image of λ* (73)*. Figure is adapted from similar figures present in literature including Figure 1 from* (74)*, Figure 2 from* (70) *and Figure 1 from* (75)*. Created with BioRender.com*

Another type of classification of phage is based on their life style, whether they are lytic, temperate or filamentous phage. Lytic phage after infection will enter the lytic life cycle, result in phage replication and bacterial host death by lysis. Temperate phage make a developmental decision after infection as they are able to either adopt a lytic lifecycle or enter what is called a lysogenic lifecycle where the phage genome is harmlessly integrated into the host genome (Figure 4). Filamentous phage, which will not be further examined in this study, are phage that develop a cooperative relationship with their bacterial host. Filamentous phage continuously replicate in their bacterial host, but unlike lytic phage will release the phage progeny using a secretion method rather than causing the host cell to lyse (76).

In the lysogenic life cycle, after attachment and injection of the phage genome into the bacterial host, the phage genome can integrate into the host's genome. Depending on the phage it can either integrate at random or at a specific chromosomal region (*attB*) which shares homology with the phage attachment sequences (*attP*). Integrases proteins mediate phage genome integration into the bacterial host, belonging to one of two major protein families: tyrosine and serine recombinases (77). Integration enables the phage genome to reside within the host genome and replicate when the bacterial host replicates. A phage will only choose a lysogenic life cycle if certain cellular conditions are met. In *E. coli,* bacteriophages 186 and lambda use a protein (called CII in each case) as a type of sensor for

multiple environmental factors (78,79), where CII is able to bias the decision towards lysogeny.

If the host cell environment becomes unfavourable due to events such as significant DNA damage, a temperate phage has the ability to switch from the lysogenic state to the lytic life cycle, where this process is referred to as prophage induction. The phage genome, referred to as prophage when integrated, begins by excising its self from the host genome usually with the help of an integrase and a recombination directionality factor (RDF). The phage will hijack the host's cellular machinery and begin replication of viral DNA and production of structural proteins. After the new phage virions have been assembled, host cell lysis will occur, to release the progeny phage into the environment. It is this lytic life cycle process that makes phage an attractive alternative to antibiotics. Two different strategies have been broadly pursued. The most widespread approach is using bacteriophage directly as the treatment. The second approach is to use the phage-derived proteins that exhibit an antimicrobial function, which will be the main focus of this thesis.

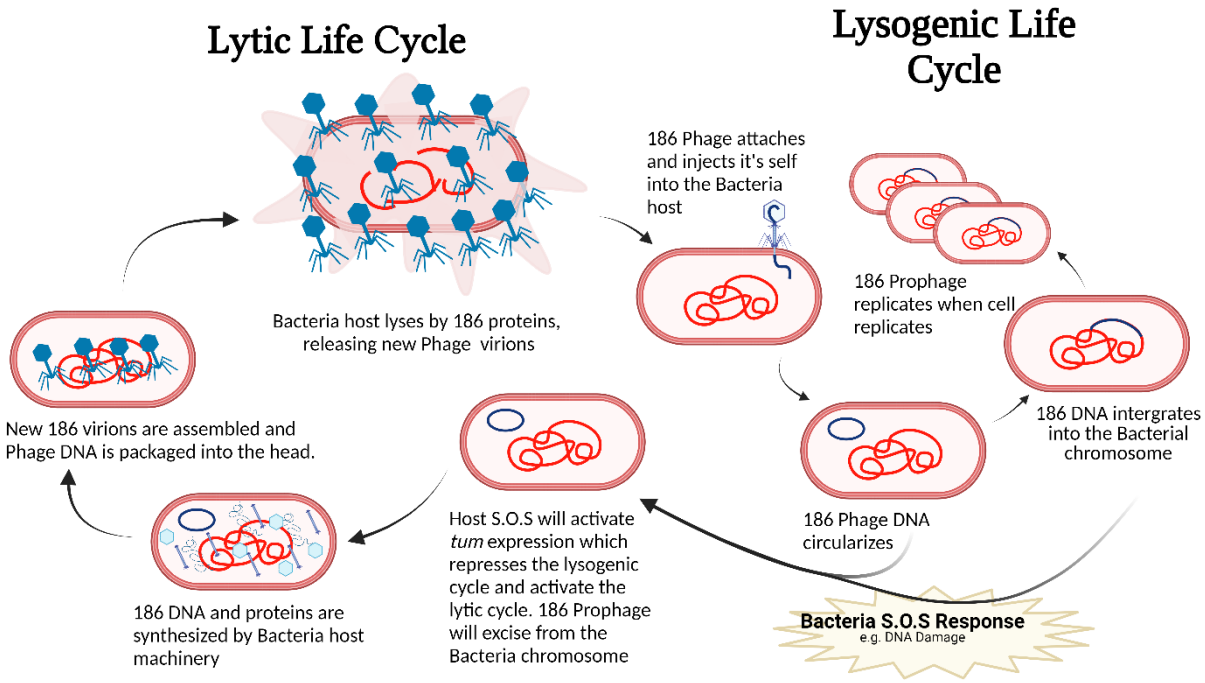


Figure 4: The General Outline of the 186 Bacteriophage Lytic and Lysogenic Life Cycle. *Infection starts when a 186 phage encounters and attaches to a bacterial host cell (E. coli) and injects its DNA. Depending upon the environment inside of the host cell, the phage DNA will follow either the lytic or lysogenic cycle. In a favourable environment the phage will enter the lysogenic cycle and the phage DNA will be integrated into the host genome. When integrated into the host genome it enters a*

dormant state and is replicates along with the host. In a less favourable environment, for example if there is expression of bacterial SOS response genes due to DNA damage, the phage is induced and enters the lytic cycle. The phage will hijack the host's cellular machinery to mass-produce phage protein and DNA to assemble and package new phage particles. Death of the host cell occurs due to the release of new phage particles into the extracellular environment via host cell lysis.

Phage produce a number of different types of proteins that have antimicrobial activity. Virion-associated peptidoglycan hydrolases (VAPGH) and lytic enzymes assist in host infection by degrading a bacterial cell wall components (80,81). VAPGH have often been found as subdomains of larger structural proteins located at the tail tip but can also be found as internal capsid proteins. VAPGH are active upon infection to assist in breaking down the cell wall (82). VAPGH represent a rich resource for therapeutic development due to their diversity and in addition often exhibit highly desirable structural properties, such as thermostability as a result of their location on the exterior of the phage particle (82).

Another class of protein produced by phage with potential therapeutic applications are the depolymerases, able to hydrolyse and disperse polysaccharides. Bacterial polysaccharides are an extremely large and diverse class of carbohydrate structures (83) which have a number of different functions in bacteria, including structural integrity, adhesion to surfaces, migration and formation of biofilms (84). Depolymerase enzymes are capable of degrading macromolecular carbohydrates that make up different components of the bacterial cell wall, including lipopolysaccharides and extracellular polysaccharides (85). Phage depolymerases are predicted to become a major therapeutic treatment for clearance of biofilms. There may be opportunities for development of combination treatments of antibiotics and phage depolymerases. Enabling the antibiotics to access exposed bacterial cells, may increase the effectiveness of the antibiotics and allow a reduction in the doses administered (85,86).

Perhaps the most noteworthy class of antimicrobial proteins are the endolysins that are expressed late in the lytic cycle. Endolysin proteins have some similarities with the VAPGH proteins. Both are capable of mediating enzymatic cleavage of the peptidoglycan (PG) cell wall, but one is active during early infection (VAPGH) and one at the end of lytic infection (endolysins). The last action during lytic infection is the release of the new viral phage particles from the host bacteria. In the majority of cases of phage that use endolysins, the endolysins degrade the PG layer leading to

cell lysis due to osmotic pressure (87,88). These proteins are noteworthy because they can clear bacteria when purified and applied externally (89–91). Research has already begun to be published looking at the use of endolysins not only in therapeutics (92–95) but also in other sectors, including food production and contamination mitigation (90).

The PG cell wall is comprised of N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG) units connected via β-1,4 glycan chain linkers (87,96) (Figure 5. C). The NAM units between layers are connected via cross-linked peptide side chains, attached via amide bonds (87,96). PG cell walls are present in both Gram-positive and Gram-negative bacteria, being a thicker layer in Gram-positive bacteria (96). In this study I will be focusing on endolysins that target the PG layer of Gram-positive type bacteria.

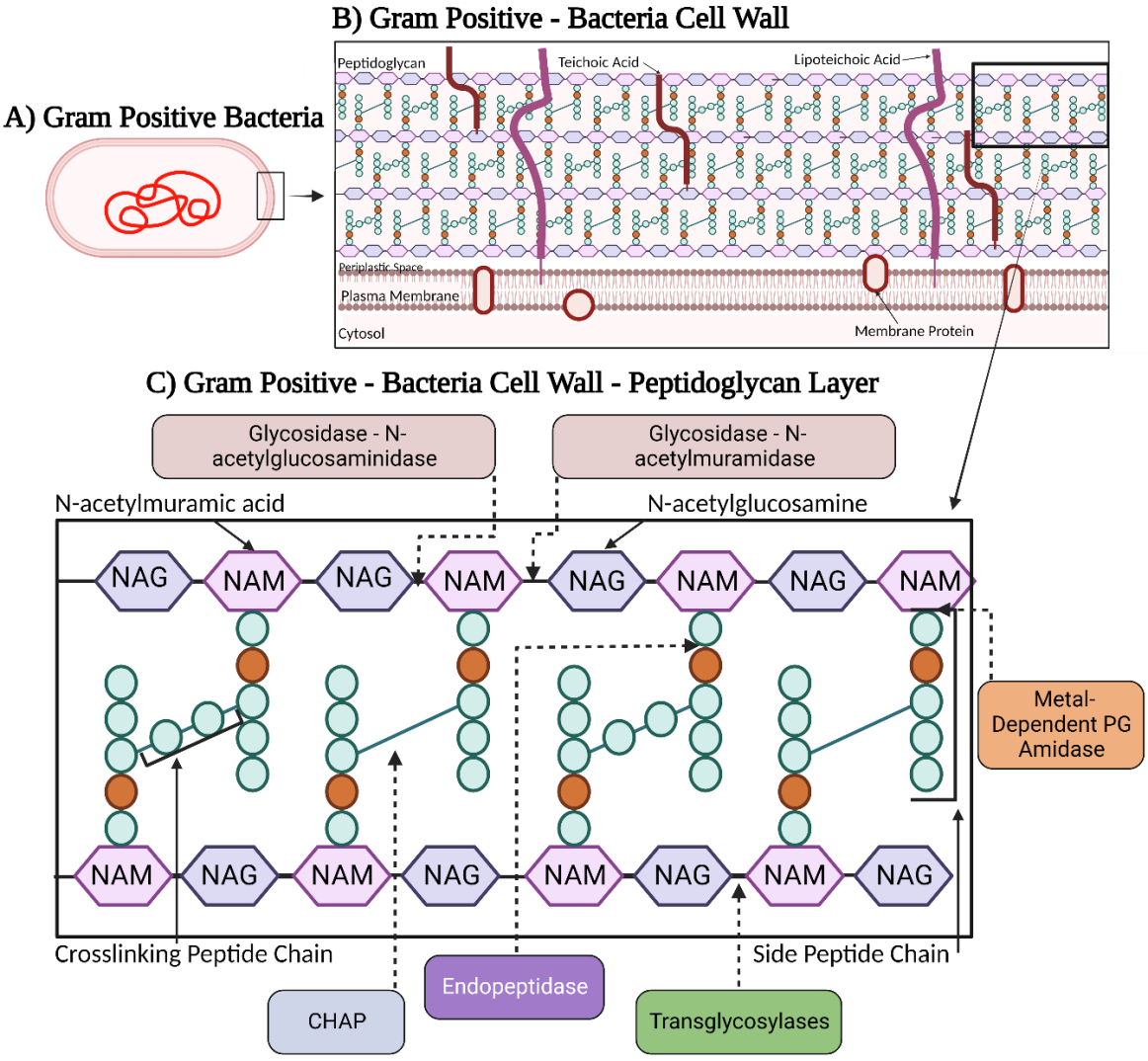


Figure 5: Gram-Positive Bacterial Cell Wall and Peptidoglycan Layer With Locations of Catalytic Domain Targets. *A) Gram-positive bacteria, with B) detail of cell wall structure, containing the*

cytosol, plasma membrane, periplasmic space, peptidoglycan, teichoic acid, lipoteichoic acid and membrane protein. C) A section of the cell wall contains an enlarged view of the peptidoglycan layer which shows the N-acetylmuramic acid (NAM), N-acetylglucosamine (NAG), crosslinking peptide chain and a unique side peptide chain (aqua and orange circles). C) also contains the target locations for five of the common endolysin catalytic domains. Glycosidase-N-acetylglucosaminidase, Glycosidase-N-acetylmuramidase and the transglycosylases domains target the bonds located between the NAM and NAG units. The Metal-Dependent PG Amidase targets the bond located between the Nam and the first peptide on the Side Peptide Chain. CHAP domain targets the bonds between the Crosslinking Peptide Chain. The endopeptidase domain targets the bonds in the Side Peptide Chain. Created with BioRender.com

1.c.ii Endolysin Proteins

Endolysins are capable of hydrolysing the PG layer from within but are kept physically separated from their substrate until required (88). Thus during late lytic development, endolysins accumulate in the cytosol of the bacterial host, followed by a process which allows endolysin access to the PG layer. The two main mechanisms that have been identified for allowing the endolysin proteins to gain access to the PG layer are (i) a secretion system and (ii) a holin-endolysin system (87,88).

Some phage take advantage of host machinery such as the Sec secretion system (97), via secretion signals located at the termini of the endolysin proteins or through expression of chaperone-like proteins (87). On the other hand, the holin-endolysin system is a stand-alone phage system that allows for the transfer of endolysin proteins through pores formed in the cell membrane by holin proteins, a process critical for cell lysis timing (98).

Endolysin genes are present in phage that target Gram-positive or Gram-negative bacteria, with some differences in structure and domain architectures, reflecting their target substrates. Gram-positive bacteria have a single thick outer wall PG layer (Figure 5. B). Endolysins from phage which target Gram-positive bacteria commonly contain two types domains: one or two N terminal catalytic domains which are responsible for breaking down the PG layer, and a single C-terminal cell wall binding (CWB) domain (99,100).

Gram-negative bacteria also have a PG layer, which is much thinner compared to Gram-positive bacteria and is surrounded by an outer membrane made up of

lipopolysaccharides. This outer membrane presents an extra challenge to the approach of using endolysins as therapeutic agents against Gram-negative pathogens, requiring additional proteins or methods to degrade the outer membrane so the endolysins can accesses the PG layer (101). Perhaps because the PG layer is much thinner compared to the Gram-positive bacteria, endolysins from phage targeting Gram negative bacteria are often only a single catalytic domain (Figure 6). However in some cases they may also contain an N-terminal CWB domain, such as in the endolysins found in phage ϕKMV and φKZ (100). Because of the difficulties associated with the outer wall of Gram-negative bacteria and because endolysins are the primary type of enzyme being investigated in this study, Gram-positive bacteria were chosen as the target bacteria in the assay system due to their easily accessible PG layer.

There are five different families of catalytic domains, each targeting a different site in the PG layer. These classes of enzymes include the most common; transglycosylase, glycosidase, N-acetylmuramoyl-L-alanine amidase (amidase), endopeptidase and CHAP domains (Figure 5. C).

Both transglycosylase and glycosidase catalytic domains target the NAG and NAM linker in the PG layer (Figure 5. C). Transglycosylases (N-acetylmuramidases, endo-b-N-acetylglucosaminidases) are a class of GH enzymes that cleaves the glycosidic bonds between the NAM and NAG via glycoside hydrolases, producing a cyclic 1,6-anhydro-N-acetylmuramic acid (87,102–104). Glycosidase enzymes target the glycosidic linkages (O-, N- and S-linked) of glycosides in the NAG and NAM linkers via hydrolysis (105).

Amidase and endopeptidases domains both target the peptide sidechains attached to the NAM subunit (Figure 5. C). The amidase domain targets the first peptide (L-alanine) in the linker that is connected to the NAM subunit (glycan-peptide linker) via hydrolases (106,107). Many of the amidase domains found in bacteriophage are metal ion dependent, such as the Amidase_2 and _3 families that are zinc dependent (108).

Endopeptidases enzymes have recently been found to be able to target and cleave multiple locations located in either of the side peptide chains following the first peptide linked to the NAM unit or in the crosslinking peptide chain (108) (Figure 5. C).

The types of residues and bonds found in these locations can differ greatly depending on the bacteria, therefore, this proteins family is large.

The final class of catalytic enzyme is the cysteine histidine-dependent amidohydrolase/peptidase (CHAP) (Figure 5. C). The CHAP domain containing proteins are a large superfamily of proteins which are recognised by the presence of highly conserved cysteine and histidine residues critical for activity (109). In most cases, the mode of action is cleavage via peptidoglycan hydrolysis. While it is classified as an endopeptidase, the CHAP domain can also display amidase activity (63) or may have both activities in a single protein (110).

A second, non-enzymatic domain family common in Gram-positive phage endolysins are the C terminal CWB domains. The CWB domain functions by binding to a cellular component on the envelope of bacterial cell walls. A theory with some supporting evidence regarding the function of the CWB domain in Gram-positive phage endolysin proteins is that it assists the endolysin in PG layer degradation by bringing the catalytic domains close to the PG layer (63,111). Another theory is that the CWB domain prevents the endolysin from diffusing away from the host bacteria during lysis. Since Gram-positive bacteria have a thicker PG layer, increased concentrations of the endolysin proteins may be required for phage release. If the endolysin diffuses away from the host, it may affect cell lysis timing and/or effect neighbouring, uninfected cells that the released progeny phage could later infect (87). Interestingly, the CWB domain may also be critical for enzyme function, with its removal leading to a partial or complete loss of activity in some cases (100,112,113). The key structural features required for CWB domain function are not fully understood. CWB domains largely confer the specificity of the endolysin protein for its host. CWB domains can be highly specific to a particular bacterial strain or broad enough to recognize more than one bacterial genus (99). Currently there are four different families of commonly recognized as CWB domains, the three-helix bundle, α/β multimers, choline binding and the bacterial SH3 (SH3b) CWB domain (99). This number will almost certainly increase in the future as new domains are discovered and classified.

A CWB domain that is present in both Gram-positive and Gram-negative phage endolysins are the three–helix type domain, although Gram-positive derived CWBs tend to have three repeats of three helix bundle (Cpl-7), while Gram-negative derived CWBs usually containing a single three-helix bundle ($gp144_{CBD}$) (114,115)..

α/β multimer domains are unique because the CWB domains are expressed separately from the catalytic domain and oligomerizes with the catalytic subunits to form the active enzyme (99). PlyC, from Streptococcal phage C1, one of the most active endolysins recorded to date (63) contains an α/β multimeric CWB domain (Figure 6). The CWB domain of PlyC is expressed separately, forming an octameric structure that contains up to eight binding sites for PG recognition. Having multiple binding sites for the PG layer may contribute to the high specific activity of PlyC. Oligomerization control and timing could act as a functional switch, adding an additional form of regulation of host cell lysis (116).

Choline binding domains are CWB domains that contain one or more choline-binding modules that recognise teichoic acids that contain choline. Choline binding modules are formed by multiple tandem copies (4 to 18) of ~20 amino acid sequences that makes up an adaptable super-secondary structure (99,117). A choline binding domain from the Pneumococcal CpI-1 endolysin forms a super-helical moiety at the N terminal and C terminal β-sheet that interacts with the endolysin catalytic domain (118).

The last CWB domain type is the bacterial SH3 (SH3b) domain family, which is the domain type commonly found in *Staphylococcal* phage (119). SH3 domains are generally identified by the presence of five to seven β-strands, organised as antiparallel β-strands connected by linkers (120). The SH3 domain has been shown to recognise and bind to the inter-peptide bridges, which in *Staphylococcus* are glycine-rich. (99,113).

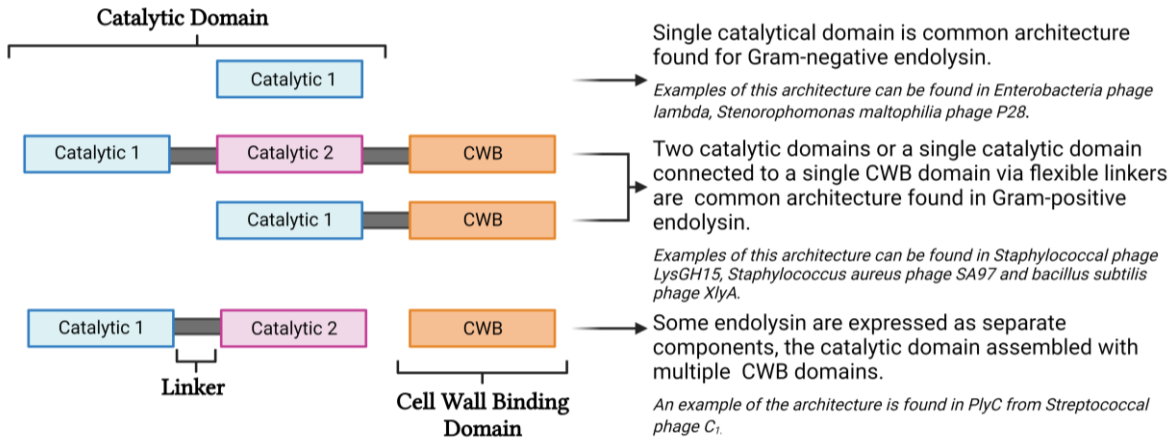


Figure 6: Bacteriophage Endolysin Architecture. *Shows the general architecture of different types of endolysin genes from different phage species. Endolysins found in Gram-negative phage are commonly found as a single catalytic domain* (121,122). *In phage which target Gram-positive*

bacteria, endolysins are more likely to have one to two catalytic domains joined to CWB domain (90,99,123,124)*. However, they can also be found expressed as separate protein with an α/β type CWB domain* (63)*. where the catalytic domains are expressed separately from the CWB domain. Figure created using BioRender.com*

These CWB domains are connected to the catalytic domains by flexible linkers, with the exception of CWB that are expressed separately. The domains are in many cases commonly perceived to act independently of one another (125). This has opened a new area of study and investigation, the generation of artificial chimeric endolysins constructed using domains from separate endolysins. Chimeric endolysins can have a number of chemical and technical advantages over their wild type counter parts (126). A common issue when investigating endolysin for lytic activity is their poor solubility which effects protein purification, leading to low yields (125,127–129). Chimeric endolysins can be constructed as a new approach to overcome solubility issues, choosing domains that individually have a high solubility (125). Constructing chimeric endolysins can also be used to target other properties such as; hydrophobicity, improved stability to pH, temperature, and longer half-life. In some cases that chimeric endolysins have been demonstrated to have a greater lethality compared to the parent endolysins (130–132). Other advantages such as increased host range and synergistic action between the chosen domains can also be pursued when constructing chimeric endolysins. Proteins are more suited for satisfying through current regulatory requirements, compared to functional intact phage and so present opportunities for intellectual property (IP) protection.

Like many recombinant proteins, poor solubility can make it difficult to test endolysins for muralytic activity (125). A common approach to testing these proteins is to generate a purified sample of the protein and test activity against a target strain by observing zones of lysis on indicator/agar plates, or by following the cell density (via OD_{600}) of liquid cultures (130,133). Many functional studies of phage antimicrobial proteins have been performed on proteins that are readily purified using protein tags and/or purification buffers (63,89,90,94,116,121,123,134–140). However, if the protein being purified is poorly soluble, it may not be possible to purify sufficient protein for these functional tests. Poor solubility will also affect a protein's capability for further drug development, with many therapies, including topical therapies, requiring proteins with high solubility (141). However, modification and site-directed mutagenesis can increase a protein's solubility (126,142).

There are other *in vivo* methods that can be used to screen endolysin proteins for muralytic activity without purification. Inducing lysis of a protein expression strain then applying the target strain is one such approach (143). One such method involves growing individual colonies of the expression strain on an agar plate and using chloroform vapours to lyse the expression strain before then adding a top layer of agar containing the target strain (143). The use of chloroform is undesirable and makes retrieval of desired clones problematic or impossible (144). Individual expression strains can be induced, lysed separately, and the lysate added to the target strain in a liquid or solid media assay. However, this dilutes the concentration of the protein of interest.

A simpler approach using ligand-induced cell lysis would be preferable. The overarching aim of this thesis is to generate a novel cell lysis system that can be adapted for library screening.

1.d The Scope of this Thesis

This PhD study aims to contribute to two areas of antimicrobial protein research. The first area is to develop new tools for testing proteins for antimicrobial activity. The second area is made up of three components; an improved bioinformatics search for putative phage antimicrobial proteins, a functional test of phage protein for muralytic activity and finally optimization of positive candidates for improved activity.

The first part of this thesis addresses the first area of study, the issue of screening antimicrobial proteins without requiring purification. This study is specifically interested in antimicrobial proteins found in phage, identified or predicted as phage endolysin proteins. In this thesis I describe the development of a novel assay system that does not require protein purification to test for antimicrobial activity. This assay system is both simple to perform and cost effective.

The approach developed here uses an *E. coli* expression strain engineered to contain three modules: a synthetic switch generated by Alejandra Isabel of the Shearwin Laboratory that uses chemical induction to induce cell lysis (module 1) (145), an engineered 186 prophage genome (module 2) and a protein of interest integration plasmid (module 3). The chemical induction is used to simultaneously activate expression of a protein of interest expression, and initiate cell lysis(145). The

chemical that is used for inducing the expression strain is cumic acid (cumate) (146), which has no other impacts on bacterial growth. This chemically inducible lysis system enables the endolysin expression strain to be co-plated with a target bacterial species, here *Staphylococcus aureus* (*S. aureus*), for visual identification of any killing of the target strain.

This study was successful in development of a novel antimicrobial functional test (which we have named the **PHEARLESS** system: **Ph**age-based **E**xpression, **A**mplification and **R**elease of **L**ytic **E**nzyme **S**pecie**s**) that does not require protein purification. It is a method that is simple to perform, not requiring any expensive reagents or equipment. Two sets of protocols were developed, the first where the aim is to simply to test if a protein has antimicrobial activity against a specific target bacterium. Potentially any protein can be investigated in this assay system as long as it can be functionally expressed in *E. coli*. The second set of protocols allows for screening a library for improved activity or expanded host range against a specific target bacterium. This assay addresses a significant issue for library screening methods, which is retaining the genotype from an active clone when the expression strain is lysed as part of the testing method. The new assay approach allows for high throughput screening of the assay. This is an advantage over current methods for screening libraries, which require a physical copy of each clone tested to be kept for retrieval of the genotype (144).

Including the time required for cloning, testing of a single protein for activity can be done in as little as two weeks. For more efficient testing, multiple proteins of interest can be tested in parallel. This research contribution will allow for faster testing of multiple proteins against any bacterial that can be cultured on solid media, enabling testing of proteins that were previously found to be to insufficiently soluble for protein purification. This will allow for a faster discover of antimicrobial proteins that would be potential drug candidates.

In addition to testing individual proteins for activity the assay can be adapted to test the effect of specific site mutations, chimeric endolysins and placement of purification tags. This assay system has also been adapted for mutant library screening, using random mutagenesis. However, testing such proteins requires the release of the protein following destruction of the host. This makes retrieval of positive clones difficult. Assay developed in this thesis greatly improves the high throughput

capabilities of screening these type of proteins with a novel way to retrieve the genetic material of the positive clones.

The second area that this study addresses is the development of a bioinformatics pipeline that searches for phage proteins that contain domains that are commonly found in proteins with muralytic activity. The pipeline started with two datasets of genomic sequences for clinical *S. aureus* strains. Each sequence was investigated for prophage genomes using an online tool called PHASTER. PHASTER produced a list of predicted putative proteins for each of the prophage genome discovered. Hidden Markov model domain searchers were performed on each protein and the predicted domains were searched against a known list of catalytic, CWB and antimicrobial domains. The final list of proteins contained a number of details about each protein; their location, size, protein sequence, number of domains discovered, type of domains found, domains E-values, domain locations in the protein, top BLAST results and if the protein sequence is present in any other strains. From both data sets of putative antimicrobial proteins that were identified, four proteins were chosen to be tested in the PHEARLESS assay for activity. In addition to the four proteins found in the bioinformatics search, eleven additional proteins were tested. Three of these were chosen from a separate bioinformatics search which was investigating the biological and evolutionary relationship between *S. aureus* phage. Five of the proteins tested were provided by the Basil Hetzel Institute (Adelaide, South Australia) as part of a collaborative project. The final four proteins were chimeric endolysins generated in an attempt to increase the muralytic compared to one of the parent endolysins, tested using the PHEARLESS system.

By developing and using the PHEARLESS system, this thesis has contributed to the number of tested phage proteins, seven of which came from two bioinformatics searches. This work also demonstrates how the PHEARLESS assay can work in a collaborative setting, being used to test proteins identified by another research group.

Chapter 2: Materials and Methods

2.a Materials

2.a.i Bacteria and Bacteriophage

Table 1 contains all bacterial strains used and created in this study. Table 1 includes details of strain ID, it's original source/genotype information, any plasmids it contains and any significant notes about it or its uses. Square brackets indicate sequences chromosomally integrated, with superscripts indicating the integration site.

Table 1: Bacteria and Bacteriophage.

ID	Bacterial Strain	Original Source/ Genotype Information/ Bacterial Species	Plasmid	Note/Use
186	NA	Wild Type	NA	186 Bacteriophage, NC_001317.1
186CI10	NA	*E. coli* Laboratory Strain	NA	186 Bacteriophage carrying a mutation in the CI (10^{th} amino acid). Mutation causes it to always enter the lytic cycle, used as control to confirm phage infection in host.
E4640	MG1655	CGSC 7925	NA	Wild type *E. coli* strain.
E4643	E4643	*E. coli* Laboratory Strain	NA	*E. coli* strain used commonly in Shearwin laboratory, made from E4640. Entire Lac operon of E4640 (*lacZYAI*) deleted via recombineering.
HB59	E4643[186]	*E. coli* Lab Strain/ E4643	NA	Strain contains a single 186 wild type phage integrated at primary *attB* site.
HB60	E4643[186]	HB59	pSIM6	pSIM6 plasmid used for integrating double stranded DNA into chromosome by recombineering. Plasmid is temperature sensitive and will activate above 30°C. Plasmid details found in Table 2: Plasmid Construct Table of Content.
E4170	POD	*E. coli* Lab Strain/E251 double 186 lysogen	NA	POD = prophage of death, contains a copy of 186 at each of the two integration sites. Grown at 30°C and used as a positive control for integration when screening new 186 lysogen strains.
E573	C600	C600 [186]	NA	Contains a single copy of 186+ at the first integration which is used as a positive control when

				screening new 186 lysogen strains.
LMS65	DSMZ 20610	*Staphylococcus gallinarum*	NA	Originally isolated from chicken skin. Used to PCR amplify "P68" Dpo7 gene
RN4220	*Staphylococcus aureus*	Non-Clinical *Staphylococcus aureus*	NA	Nonclinical strain of *Staphylococcus aureus*
NEB5α	NEB® 5-alpha Competent E. coli	*E. coli* Strain, New England Biolabs	NA	NEB 5-alpha competent *E. coli* is a derivative of the popular DH5α. It is T1 phage resistant and *end*A deficient for high-quality plasmid preparations. Used for cloning plasmids.
HER1049	*Staphylococcus aureus*	*Staphylococcus aureus P68 host*	NA	Non-clinical strain of *Staphylococcus aureus* used as host for bacteriophage P68.
E4644	*E. coli*	*E. coli* Laboratory Strain, EC100D *pir*$^{+}$	NA	For propagating plasmids with R6kγ ori ~ 15 copies, used for assembling pIT4 plasmids which contain R6Kγ origin of replication.
QC024	*E. coli* [pIT5-pBla-tdTomato]$^{186\text{-}1}$ [pIT5-pBla-tdTomato]$^{186\text{-}2}$	*E. coli* Laboratory Strain, E4643	NA	Contains fluorescent protein gene (Td-tomato) integrated into both 186 phage attachment sites. Td-tomato is constitutively expressed. Strain was constructed by Qinqin Chen, (KES lab).
HB61	E4643 [186(Δ*tum* Δ*cos, att2, Cm*)]	*E. coli* Laboratory Strain, HB60	NA	Contains a mutant 186 genome that has had the *tum/orf97* genes and *cos* sequence removed and replaced with two non-*E. coli* phage attachment sites (147) and a chloramphenicol resistance gene (*Cm*) using recombineering. Lacks a functioning *cos* sequence.
HB63	E4643 [186(Δ*tum, Δcos, att2, Cm*)]	*E. coli* Laboratory Strain, HB61	phi21 helper	Contains the helper plasmid phi21 for the pIT4-KT plasmid integration plasmid.
HB66	E4643 [pIT4-KT-cymR-pCym-empty] $^{\phi 21}$ [186(Δ*tum, Δcos, att2, Cm*)]	*E. coli* Laboratory Strain, HB63	NA	Integrated pIT4-KT-cymR-pCym-empty plasmid which contains the negative control for the Cym Tum module and a Kanamycin resistance gene. Original plasmid provided by Alejandra Isabel (KES lab).
HB67	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, Δcos, att2, Cm*)]	*E. coli* Laboratory Strain, HB63	NA	Integrated pIT4-KT-cymR-pCym-*Tum72* plasmid which contains the truncated version of tum72 and a Kanamycin resistance gene. Plasmid provided by Alejandra Isabel.
HB70	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$	*E. coli* Laboratory	NA	Integrated pIT4-KT-cymR-pCym-*Tum* plasmid which contains the wild type version of *tum* and a

	[186(Δ*tum, Δcos , att2, Cm*)]	Strain, HB63		Kanamycin resistance gene. Plasmid provided by Alejandra Isabel (KES lab).
HB74	E4643 [pIT4-KT-cymR-pCym-empty] $^{\phi 21}$ [186(Δ*Tum, Δcos , att2, Cm*)]	*E. coli* Laboratory Strain, HB66	pSBIA2-GFP	Strain HB66 carrying a green fluorescent protein (GFP) expression plasmid, pSBIA2-GFP.
HB76	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, Δcos, att2, Cm*)]	*E. coli* Laboratory Strain, HB67	pSBIA2-GFP	Strain HB67 with a GFP expression plasmid, pSBIA2-GFP.
HB78	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, Δcos, att2, Cm*)]	*E. coli* Laboratory Strain, HB70	pSBIA2-GFP	Strain HB70 with a GFP expression plasmid, pSBIA2-GFP.
HB80	E4643 [pIT4-KT-cymR-pCym-empty] $^{\phi 21}$ [186(Δ*Tum, Δcos , att2*)]	*E. coli* Laboratory Strain, HB74	pSBIA2-GFP, pZS45 186pR-ClyF	Contains GFP expression plasmid and ClyF expression plasmid.
HB82	E4643 [pIT4-KT-cymR-pCym-empty] $^{\phi 21}$ [186(Δ*Tum, Δcos , att2*)]	*E. coli* Laboratory Strain, HB74	pSBIA2-GFP, pZS45 186*pJ*-ClyF	Contains GFP expression plasmid and ClyF expression plasmid.
HB84	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, Δcos, att2*)]	*E. coli* Laboratory Strain, HB76	pSBIA2-GFP, pZS45 186*pR*-ClyF	Contains GFP expression plasmid and ClyF expression plasmid.
HB86	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, Δcos , att2*)]	*E. coli* Laboratory Strain, HB76	pSBIA2-GFP, pZS45 186*pJ*-ClyF	Contains GFP expression plasmid and ClyF expression plasmid.
HB87	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, Δcos, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP, pZS45 186*pR*-ClyF	Contains GFP expression plasmid and ClyF expression plasmid.
HB88/HB 89	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, Δcos, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP, pZS45 186*pJ*-ClyF	Contains GFP expression plasmid and ClyF expression plasmid.
HB90	E4643 [pIT4-KT-cymR-pCym-empty] $^{\phi 21}$ [186(Δ*Tum, Δcos , att2*)]	*E. coli* Laboratory Strain, HB74	pSBIA2-GFP, pZS45 empty	Contains GFP expression plasmid and an empty (no ClyF) expression plasmid. Used as a negative control.
HB91	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*Tum, Δcos, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP, pZS45 empty	Contains GFP expression plasmid and an empty (no ClyF) expression plasmid. Used as a negative control.
HB92	E4643	*E. coli* Laboratory	phi21 helper	Contains the helper plasmid phi21 for the pIT4-KT plasmid integration plasmid.

		Strain, E4643		
HB93	E4643 [186(Δ*tum, cos+, att2, Cm*)]	*E. coli* Laboratory Strain, HB60	NA	186 lysogen that has had the *tum/orf97* genes removed and replaced with two non-*E. coli* phage attachment sites (147) and chloramphenicol resistance gene, using recombineering. The lysogen carries wild type cos+ sequence in the 186 genome and can produce functional phage.
HB95	E4643 [pIT4-KT-cymR-pCym-*Tum72*] ϕ_{21} [186(Δ*tum, Δcos, att2*)]	*E. coli* Laboratory Strain, HB76	pSBIA2-GFP, pZS45 empty	Contains GFP expression plasmid and an empty expression plasmid. Used as a negative control.
HB96	E4643 [pIT4-KT-cymR-pCym-*Tum72*] ϕ_{21}	*E. coli* Laboratory Strain, HB92	NA	Integrated pIT4-KT-cymR-pCym-*Tum72* plasmid which contains the truncated version of Tum72 and a kanamycin resistance gene. Plasmid provided by Alejandra Isabel (KES lab).
HB97	E4643 [pIT4-KT-cymR-pCym-*Tum*] ϕ_{21}	*E. coli* Laboratory Strain, HB92	NA	Integrated CymR 1 pIT4-KT plasmid which contains the wild type version of *tum* and a kanamycin resistance gene. Plasmid provided by Alejandra Isabel (KES lab).
HB98	E4643 [pIT4-KT-cymR-pCym-empty] ϕ_{21}	*E. coli* Laboratory Strain, HB92	NA	Integrated pIT4-KT-cymR-pCym-empty plasmid which is the negative control for the Cym Tum switch. Carries a kanamycin resistance gene. Plasmid provided by Alejandra Isabel (KES lab).
HB102	E4643 [186(Δ*tum, cos+, att2, Cm*)]	*E. coli* lab strain, HB93	ϕ 21 helper	HB93 carrying ϕ21 helper plasmid to allow pIT4 plasmid integration into ϕ21 *attB* site.
HB103	E4643 [pIT4-KT-cymR-pCym-empty] ϕ_{21} [186(Δ*tum, cos+, att2, Cm*)]	*E. coli* lab strain, HB102	none	HB93 with the pIT4-KT-cymR-pCym-empty control plasmid integrated into the ϕ21 *attB* site. Plasmid provided by Alejandra Isabel (KES lab).
HB104	E4643 [pIT-cymR-pCym-*Tum*] ϕ_{21} [186(Δ*tum, cos+, att2, Cm*)]	*E. coli* lab strain, HB102	none	HB93 with the pIT4-KT-cymR-pCym-*Tum* plasmid integrated into the ϕ21 *attB* site. Plasmid provided by Alejandra Isabel (KES lab).
HB106	E4643 [pIT4-KT-cymR-pCym-*Tum72*] ϕ_{21} [186(Δ*tum, cos+, att2, Cm*)]	*E. coli* Laboratory Strain, HB102	NA	Strain containing 186Δ*Tum* and wild type cos sequence, with the pIT4-KT-cymR-pCym-tum72 plasmid integrated into it. Plasmid provided by Alejandra Isabel (KES lab).
HB112	E4643 [pIT4-KT-cymR-pCym-*Tum*] ϕ_{21} [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP + pZS(^)45 empty	Carries GFP expression plasmid and high copy number version of the expression plasmid with no ClyF, used as negative control.
HB114	E4643 [pIT4-KT-cymR-pCym-*Tum*] ϕ_{21} [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP + pZS(^)45 186*pJ* ClyF	Carries GFP expression plasmid and high copy number version of the ClyF expression plasmid.

HB115	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB76	pSBIA2-GFP + pZS(^)45 empty	Carries GFP expression plasmid and high copy number version of the expression plasmid with no ClyF, used as negative control.
HB117	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB76	pSBIA2-GFP + pZS(^)45 186*pJ* ClyF	Carries GFP expression plasmid and high copy number version of the ClyF expression plasmid.
HB125	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP + pZS(^)45 186*pJ* ORF11 NTD-His	Carries GFP expression plasmid and high copy number version of P68 ORF11 with a N-terminal His tag expression plasmid.
HB126	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP + pZS(^)45 186*pJ* NTD ORF11	Carries GFP expression plasmid and high copy number version of the NTD of P68 ORF11 with a N-terminal His tag expression plasmid.
HB127	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2, cos+,* Δ*Cm*)]	*E. coli* Laboratory Strain, HB104	NA	Strains containing the wild type cos sequence which contains the wild type version of *tum* which has had the chloramphenicol resistance gene in the 186 genome removed.
HB128	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, att2, cos*$^{+}$*,* Δ*Cm*)]	*E. coli* Laboratory Strain, HB106	NA	Strains containing the wild type *cos* sequence which contains the wild type version of *tum72* which has had the chloramphenicol resistance gene in the 186 genome removed.
HB130	E4643 [pIT4-KT-cymR-pCym-empty] $^{\phi 21}$ [186(Δ*tum, att2, cos*$^{+}$*,* Δ*Cm*)]	*E. coli* Laboratory Strain, HB103	NA	Strains containing the wild type *cos* sequence which contains no *tum* expression which has had the chloramphenicol resistance gene in the 186 genome removed.
HB133	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2, cos*$^{+}$*,* Δ*Cm*)]	*E. coli* Laboratory Strain, HB127	att2 helper	Strain contains the *tum* wild type inducible lysis switch and also contains the helper plasmid for pIT4 integration at the *att2* attachment site.
HB134	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, att2, cos*$^{+}$*,* Δ*Cm*)]	*E. coli* Laboratory Strain, HB128	att2 helper	Strain contains the truncated *tum*72 inducible lysis switch and also contains the helper plasmid for pIT4 integration at the *att2* attachment site.
HB135	E4643 [pIT4-KT-cymR-pCym-empty] $^{\square 21}$ [186(Δ*tum, att2,* Δ*Cm*)]	*E. coli* Laboratory Strain, HB130	att2 helper	Strain contains the empty inducible lysis switch and also contains the helper plasmid for pIT4 integration at the *att2* attachment site.
HB141	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, 186pR ClyF, Spec*)]	*E. coli* Laboratory Strain, HB133	NA	*tum* wild type inducible lysis strain with the pIT4_att2_loxP_S_pR_ClyF plasmid integrated at the att2 attachment site. pIT4 plasmid contains ClyF controlled by 186*pR* promoter.
HB143	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, 186pJ ClyF, Spec*)]	*E. coli* Laboratory Strain, HB134	NA	Truncated *tum*72 inducible lysis strain with the pIT4_att2_loxP_S_186*pJ*_ClyF plasmid integrated at the att2 attachment site. pIT4 plasmid

				contains ClyF controlled by 186*pJ* promoter.
HB145	E4643 [pIT4-KT-cymR-pCym-empty] $^{\phi 21}$ [186(Δ*tum, 186pJ ClyF, Spec*)]	*E. coli* Laboratory Strain, HB135	NA	Empty inducible strain with the pIT4_att2_loxP_S_186*pJ* _ClyF plasmid integrated at the att2 attachment site. pIT4 plasmid contains ClyF controlled by 186*pJ* promoter.
HB147	E4643 [pIT4-KT-cymR-pCym-empty] $^{\phi 21}$ [186(Δ*tum, 186pR ClyF, Spec*)]	*E. coli* Laboratory Strain, HB135	NA	Empty inducible strain with the pIT4_att2_loxP_S_pR_ClyF plasmid integrated at the att2 attachment site. pIT4 plasmid contains ClyF controlled by 186*pR* promoter.
HB149	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, 186pR ClyF, Spec*)]	*E. coli* Laboratory Strain, HB134	NA	Truncated *tum*72 inducible strain with the pIT4_att2_loxP_S_pR_ClyF plasmid integrated at the att2 attachment site. pIT4 plasmid contains ClyF controlled by 186*pR* promoter.
HB152	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, 186pJ ClyF, Spec*)]	*E. coli* Laboratory Strain, HB133	NA	*tum* wild type inducible strain with the pIT4_att2_loxP_S_186*pJ*_ClyF plasmid integrated at the att2 attachment site. pIT4 plasmid contains ClyF controlled by 186*pJ* promoter.
HB155	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, 186pJ ClyF*)]	*E. coli* Laboratory Strain, HB143	NA	Integrated 186*pJ* ClyF and truncated *tum*72 expression strain with the spectinomycin resistance gene and R6Kɣ origin of replication from the pIT4 plasmid removed.
HB156	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, 186pJ ClyF*)]	*E. coli* Laboratory Strain, HB152	NA	Integrated 186*pJ* ClyF and wild type *tum* expression strain with the spectinomycin resistance gene and R6Kɣ origin of replication from the pIT4 plasmid removed.
HB157	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*tum, 186pR ClyF*)]	*E. coli* Laboratory Strain, HB149	NA	Integrated 186*pR* ClyF and truncated *tum*72 expression strain with the spectinomycin resistance gene and R6Kɣ origin of replication from the pIT4 plasmid removed.
HB158	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, 186pR ClyF*)]	*E. coli* Laboratory Strain, HB141	NA	Integrated 186*pR* ClyF and wild type *tum* expression strain with the spectinomycin resistance gene and R6Kɣ origin of replication from the pIT4 plasmid removed.
HB163	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP, pZS(^)45 186*pJ* ORF11	Plasmid based expression strain containing the GFP expression plasmid and the ORF11 expression plasmid controlled by the 186*pJ* promoter.
HB164	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP, pZS(^)45 186*pJ* P68 Genes	Plasmid based expression strain containing the GFP expression plasmid and the P68 expression plasmid.
HB165	E4643 [pIT4-KT-cymR-pCym-	*E. coli* Laboratory	NA	Truncated tum72 inducible strain with the

	Tum72]□21 [186(Δ*tum, 186pJ, Spec*)]	Strain, HB134		pIT4_*attP*2_loxP_S_186*pJ* plasmid integrated at the att2 attachment site. pIT4 plasmid contains the *186pJ* promoter.
HB172	E4643 [pIT4-KT-cymR-pCym-*Tum*] ϕ21 [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP, pZS(^)45 186*pJ* ClyF(C36A)	Expression plasmid for ClyF mutant C36A.
HB173	E4643 [pIT4-KT-cymR-pCym-*Tum*] ϕ21 [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP, pZS(^)45 186*pJ* ClyF(H99A)	Expression plasmid for ClyF mutant H99A.
HB179	E4643 [pIT4-KT-cymR-pCym-*Tum*] ϕ21 [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB78	pSBIA2-GFP, pZS(^)45 186*pJ* ClyF(C36G)	Expression plasmid for ClyF mutant C36G.
4644	NA	*EC100D pir+*	NA	Cloning and propagation strain used for assembling and amplifying plasmids that contain the R6Kɣ origin of replication which requires a *pir* gene for replication.
HB182	E4643 [pIT4-KT-cymR-pCym-*Tum72*] ϕ21 [186(Δ*cos,* Δ*tum, att2, Cm*)]	*E. coli* Laboratory Strain, HB67	NA	Expression Strain for PHEARLESS V2.2, strain contains the *tum72* cumate lysis switch and has had the Cm gene removed. This strain also originates from the cos- strain. For phage production the Cos sequence must be re-integrated in a pIT4 plasmid.
HB186	E4643 [pIT4-KT-cymR-pCym-*Tum72*] ϕ21 [186(Δ*cos,* Δ*tum, att2, Cm*)]	*E. coli* Laboratory Strain, HB182	AH6045 Phage 2 helper	Expression Strain for PHEARLESS V2.2, contains the helper plasmid for the att2 pIT4 integration plasmid.
HB189	E4643 [pIT4-KT-cymR-pCym-*Tum72*] ϕ21 [186(Δ*cos,* Δ*tum, 186pJ ClyF, cos+)*]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the *clyF* gene controlled by the 186*pJ* promoter. Strain is capable of making functional phage particles due to the reintroduced 186 *cos* sequence.
HB191	E4643 [pIT4-KT-cymR-pCym-*Tum72*] ϕ21 [186(Δ*cos,* Δ*tum, 186pJ ClyF)*]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the *clyF* gene controlled by the 186*pJ* promoter. pIT4 plasmid integrated into the 186 prophage genome does not contain the *cos* sequence required for production of functional phage particles.
HB237	E4643 [pIT4-KT-cymR-pCym-*Tum72*] ϕ21 [186 (Δ*cos,* Δ*tum, 186pJ CFLI P7, cos+*)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the PE_ CFLI_P7 gene controlled by the 186*pJ* promoter.
HB239	E4643 [pIT4-KT-cymR-pCym-*Tum72*] ϕ21 [186 (Δ*cos,* Δ*tum,*	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the PE_GDT_P4 gene controlled by the 186*pJ* promoter.

	186pJ GDT P4, cos^{+})]			
HB241	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186 (Δ*cos, Δtum, 186pJ GDT P5, cos+*)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the PE_GDT_P5 gene controlled by the 186*pJ* promoter.
HB243	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186 (Δ*cos, Δtum, 186pJ BHI P1, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the PE_BHI_P1 gene controlled by the 186*pJ* promoter.
HB245	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186 (Δ*cos, Δtum, 186pJ BHI P2, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the PE_BHI_P2 gene controlled by the 186*pJ* promoter.
HB247	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186 (Δ*cos, Δtum, 186pJ GDT P6, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the PE_GDT_P6 gene controlled by the 186*pJ* promoter.
HB250	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186 (Δ*cos, Δtum, 186pJ BHI P3, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the PE_BHI_P3 gene controlled by the 186*pJ* promoter.
HB253	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186 (Δ*cos, Δtum, 186pJ, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Contains an empty pIT4 plasmid with a 186*pJ* promoter, used as an empty control.
HB255	E4644	*E. coli* Laboratory Strain, *4644*	pIT4_*attP*2_LoxP_S_186*pJ*_ClyF_+1kb_Cos+	Cloning strain for pIT4 plasmid that contains 1 Kb “Junk” sequence (FtsK)
HB257	E4644	*E. coli* Laboratory Strain, *4644*	pIT4_*attP*2_LoxP_S_186*pJ*_ClyF_+2kb_Cos+	Cloning strain for pIT4 plasmid that contains 2 Kb “Junk” sequence (FtsK)
HB259	E4644	*E. coli* Laboratory Strain, *4644*	pIT4_*attP*2_LoxP_S_186*pJ*_ClyF_+3kb_Cos+	Cloning strain for pIT4 plasmid that contains 3 Kb “Junk” sequence (FtsK)
HB261	E4643(pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*cos, Δtum, 186pJ ClyF, 2Kb, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the pIT4 integrated plasmid that contains 2 Kb “Junk” sequence (FtsK) (HB257). Strain used to test the genome size limitations of 186, net increase ~4.1 kB.
HB263	E4643(pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*cos, Δtum, 186pJ*	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the pIT4 integrated plasmid that contains 3 Kb “Junk” sequence (FtsK) (HB259). Strain used to

	ClyF, 3Kb, cos^{+})]			test the genome size limitations of 186, net increase ~5.1 kB.
HB265	E4643(pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ*cos, Δtum, 186pJ ClyF, 1Kb, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, contains the pIT4 integrated plasmid that contains 1 Kb "Junk" sequence (*FtsK*) (HB255). Strain used to test the genome size limitations of 186, net increase ~3.1 kB.
HB278	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186 (Δ*cos, Δtum, 186pJ C1, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, Chimeric Endolysin Gene Version 1 (CV1) expression controlled by 186*pJ* promoter
HB279	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186 (Δ*cos, Δtum, 186pJ C3, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, Chimeric Endolysin Gene Version 3 (CV3) expression controlled by 186*pJ* promoter
HB280	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186 (Δ*cos, Δtum,186pJ C4, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, Catalytic Domains Gene Version 4 (CV4) expression controlled by 186*pJ* promoter
HB281	E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186 (Δ*cos, Δtum, 186pJ C2, cos*$^{+}$)]	*E. coli* Laboratory Strain, HB186	NA	Expression Strain for PHEARLESS V2.2, Chimeric Endolysin Gene Version 1 (CV1) expression controlled by 186*pJ* promoter
HB294	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB70	pZS(^)45 186*pJ* Clinda8	Expression Strain for PHEARLESS V2.2, contains the pZS(^)45 186*pJ* expression plasmid for Clinda8 protein expression. Strain is unable to produce any functioning phage particles.
HB296	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB70	pZS(^)45 186*pJ* LysK (HB291)	Expression Strain for PHEARLESS V2.2, contains the pZS(^)45 186*pJ* expression plasmid for LysK protein expression. Strain is unable to produce any functioning phage particles.
HB298	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB70	pZS(^)45 186pJ Clinda3	Expression Strain for PHEARLESS V2.2, contains the pZS(^)45 186*pJ* expression plasmid for Clinda3 protein expression. Strain is unable to produce any functioning phage particles.
HB301	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB70	pZS(^)45 186*pJ* First Translation (FT)	Expression Strain for PHEARLESS V2.2, contains the pZS(^)45 186*pJ* expression plasmid for First Translation (FT) protein expression. Strain is unable to produce any functioning phage particles.
HB302	E4643 [pIT4-KT-cymR-pCym-*Tum*] $^{\phi 21}$ [186(Δ*tum, att2*)]	*E. coli* Laboratory Strain, HB70	pZS(^)45	Expression Strain for PHEARLESS V2.2, contains the empty pZS(^)45 expression plasmid. Functioning as a

				negative control plasmid. Strain is unable to produce any functioning phage particles.

2.a.ii Plasmids Construct

Table 2 details the plasmids used and created in this study. Table 2 includes detail about the plasmid name, features, selection marker, glycerol stock name and any specific notes about the plasmid or its applications.

Table 2: Plasmid Construct Table of Content.

Plasmid	Features	Selection Marker	Glycerol Stock	Note/Use
pET15b NTH ORF11 CTD	CTD ORF11, Amp Resistance gene	Amp100	HB53	CTD of ORF11 gene from bacteriophage P68
pSIM6	Amp resistance gene, lambda Red recombineering functions genes (*gam, exo, bet*)	Amp100, temperature sensitive origin of replication	Physical copy provided by Nan Hao, plasmid reference (148)	Plasmid used for integrating double stranded DNA into chromosome by recombineering. Plasmid is temperature sensitive (lambda CIts) and will activate above 30°C.
phi21 helper	Amp resistance gene, Integrase	Amp100, ts origin	Physical copy provided by Nan Hao plasmid reference (149)	Helper plasmid for integrating pIT4 integration plasmids into the Phi21 phage attachment site in *E. coli*. Used to integrate the cumate switch into the E4643 genome.
pIT4-KT-cymR-pCym-empty	Empty cumate switch	Kan20	Physical copy provided by Alejandra Isabel.	pIT4 integration plasmid containing no *tum* induction module and is used as a negative control.
pIT4-KT-cymR-pCym-*Tum*	*tum* wild type expression controlled by cumate switch	Kan20	Physical copy provided by Alejandra Isabel.	pIT4 integration plasmid containing the cumate induction module which represses *tum* wild type expression in the absence cumate.
pIT4-KT-cymR-pCym-*Tum72*	Truncated *tum* expression controlled by cumate switch	Kan20	Physical copy provided by Alejandra Isabel.	pIT4 integration plasmid containing the cumate induction module which represses truncated *tum* expression in the absence cumate.
pZS45	Origin of replication with low copy number	Spec50	KES 1366, plasmid reference (150)	Expression plasmid. Obtained from Keith Shearwin
pZS45 186*pR*-ClyF	ClyF expression controlled by 186*pR* promoter	Spec50	HB68	pZS45 plasmid containing a chimeric endolysin, ClyF (130), which is controlled by

				the 186*pR* promoter from 186 phage and a Spectinomycin resistance gene.
pZS45 186*pJ*-ClyF	ClyF expression controlled by 186*pJ* promoter	Spec50	HB73	pZS45 plasmid containing a chimeric endolysin, ClyF (130), which is controlled by the 186*pJ* promoter from 186 phage and a Spectinomycin (Spec) resistance gene.
pSBIA2-GFP	SF-GFP expression	Amp100	Lab stock E2942, SF-GFP in *E. coli* DH5alpha -Z1	Plasmid with IPTG inducible superfolder GFP (SF-GFP) expression. SF-GFP (151) is controlled by the lac promoter which is active in the presence of β-D-1-thiogalactopyranoside (IPTG)
pZS(^)45	High copy number origin of replication mutant version	Spec50	HB100/99	Made from pZS45, contains a high copy number version of pSC101 origin. Increases the copy number from ~3 to ~70 (152).
pZS(^)45 186*pJ* ClyF	ClyF Expression plasmid, high copy number plasmid	Spec50	HB111	High copy origin of replication pZS(^)45 plasmid containing a chimeric endolysin, ClyF (130), which is controlled by the 186*pJ* promoter and a Spectinomycin resistance gene.
pZS(^)45 186*pJ* ORF11 NTH	ORF11 from Bacteriophage P68 with a N-terminal 6xHis Tag	Spec50	HB122	High copy origin of replication pZS(^)45 plasmid containing a chimeric endolysin, ORF11 which is controlled by the 186*pJ* promoter and a Spectinomycin resistance gene. ORF11 has an N-terminal 6xHis tag.
pZS(^)45 186*pJ* NTD ORF11	N-terminal domain of ORF11 from Bacteriophage P68 with an N-terminal 6xHis tag.	Spec50	HB123	High copy of replication pZS(^)45 plasmid containing a chimeric endolysin, ORF11 which is controlled by the 186*pJ* promoter from 186 phage and a Spectinomycin resistance gene. ORF11 an N-terminal 6xHis tag.
att2 helper	att2 Helper plasmid	Amp100, origin^{ts}	AH6045, Physical copy provided by Nan Hao plasmid (149)	Helper plasmid for integrating a pIT4 containing the att2 attachment sequence into the att2 phage attachment site. Obtained from Andrew Hao (KES lab)
att3# helper	att3# helper plasmid	Amp100, origin^{ts}	AH6046, Physical copy provided by Nan	Helper plasmid for integrating a pIT4 containing the att3# attachment sequence into the att3# phage attachment

			Hao plasmid (149)	site. Obtained from Andrew Hao (KES lab)
pIT4_att2_loxP_ S_186pR_ClyF	ClyF gene controlled by *pR*	Spec20	HB132	pIT4 integration plasmid containing the ClyF gene controlled by 186*pR* promoter. pIT4 plasmid also contains two terminator sequence and a spectinomycin resistance gene.
pIT4_att2_loxP_ S_186pJ_ClyF	ClyF gene controlled by 186*pJ* promoter	Spec20	HB137	pIT4 integration plasmid containing the ClyF gene controlled by 186*pJ* promoter. pIT4 plasmid also contains two terminator sequence and a spectinomycin resistance gene.
pIT4_att2_LoxP_S_1 86pR	186*pR* promoter	Spec20	HB161	Empty pIT4 plasmid which contains the 186*pR* promoter, used as a negative control.
pIT4_att2_LoxP_S_1 86pJ	186*pJ* promoter	Spec20	HB162	Empty pIT4 plasmid which contains the 186*pJ* promoter, used as a negative control.
pZS(^)45 186pJ ORF16	ORF16 controlled by 186*pJ* promoter	Spec50	HB167	Expression plasmid for ORF16 controlled by the 186*pJ* promoter.
pZS(^)45 186pJ ClyF(C36A)	ClyF(C36A) controlled by 186*pJ* promoter, high copy plasmid	Spec50	HB170	Expression plasmid for ClyF(C36A) which contains two nucleotide mutations, changing wild type Cysteine (Cys) 36 to Alanine (Ala)
pZS(^)45 186pJ ClyF(H99A)	ClyF(H99A) controlled by 186*pJ* promoter, high copy plasmid	Spec50	HB171	Expression plasmid for ClyF(H99A) which contains two nucleotide mutations, changing wild type Histidine (His) 99 to Ala.
pZS(^)45 186pJ ClyF(C36G)	ClyF(C36G) controlled by 186*pJ* promoter, high copy plasmid	Spec50	HB176 /HB177	Expression plasmid for ClyF(C36G) which contains two nucleotide mutations, changing wild type Cysteine (Cys) 36 to Glycine (Gly).
pIT4_att2_LoxP_S_ 186pJ_ClyF_cos+	ClyF controlled by 186*pJ* promoter, att phage 2 attachment site	Spec20	HB184	pIT4 integration plasmid containing the ClyF gene controlled by 186*pJ* promoter. pIT4 plasmid also contains a spectinomycin resistance gene and the 186 phage *cos*+ sequence.
pIT4_LoxP_SL-I52002	Terminator sequences, LoxP LE/RE, Spectinomycin Resistant Gene, R6K Origin of Replication	Spec20	AH1724	Backbone used for assembling the first pIT4 plasmids assembled. Obtained from Andrew Hao (KES lab)
pIT_HF_CL_1.4kb_O R_O2_LacZ	Terminator Sequences	Cm20	AH1529	Backbone used for assembling the first pIT4 plasmids assembled.

				Obtained from Andrew Hao (KES lab)
pIT4_*attP*2_LoxP_S_186pJ_ClyF_+1kb_cos+	ClyF controlled by 186*pJ*, 186 cos sequence, "junk" fragment DNA (*FtsK*).	Spec20	HB255	A pIT4 plasmid used to increase the 186 genome size by ~3.1 Kb
pIT4_*attP*2_LoxP_S_186pJ_ClyF_+2kb_cos+	ClyF controlled by 186*pJ*, 186 cos sequence, "junk" fragment DNA (*FtsK*).	Spec20	HB257	A pIT4 plasmid used to increase the 186 genome size by ~4.1 Kb
pIT4_*attP*2_LoxP_S_186pJ_ClyF_+3kb_cos+	ClyF controlled by 186*pJ*, 186 cos sequence, "junk" fragment DNA (*FtsK*).	Spec20	HB259	A pIT4 plasmid used to increase the 186 genome size by ~5.1 Kb
pIT-HF-CL-NtrC_5.6kb_glnAp2_weakRBS	*FtsK* Fragment	Cm20	AH1606	Plasmid used as template to amplify "junk" fragment DNA (FtsK)
pZS(^)45 186pJ First Translation	LysK first Translation protein gene controlled by 186*pJ*, high copy plasmid	Spec50	HB285	High copy of replication pZS(^)45 plasmid containing the first translation product of LysK (HNH) which is controlled by the 186*pJ* promoter from 186 phage and a Spectinomycin resistance gene.
pZS(^)45 186pJ Clinda3	Clinda3 protein gene controlled by 186*pJ*, high copy number plasmid	Spec50	HB286	High copy number of replication pZS(^)45 plasmid containing the Clinda3 which is controlled by the 186*pJ* promoter and a Spectinomycin resistance gene. Putative endolysin genes provided by the Basil Hetzel Institute.
pZS(^)45 186pJ Clinda8	Clinda8 protein gene controlled by 186*pJ*, high copy number plasmid	Spec50	HB287	High copy number of replication pZS(^)45 plasmid containing the Clinda8 which is controlled by the 186*pJ* promoter from 186 phage and a Spectinomycin resistance gene. Putative endolysin genes provided by the Basil Hetzel Institute.
pZS(^)45 186pJ LysK	LysK protein gene controlled by 186*pJ*, high copy number plasmid	Spec50	HB291	High copy number of replication pZS(^)45 plasmid containing the LysK which is controlled by the 186*pJ* promoter from 186 phage and a Spectinomycin resistance gene.
pZS45up_186pJ_lysK_HNH_H77A	LysK(H77A) protein gene controlled by 186*pJ*, high copy plasmid	Spec50	Plasmid cloned by Nan Hao, transformed by Hannah Bonham, HB306	High copy number of replication pZS(^)45 plasmid containing the LysK gene with a point mutation converting the 77 amino acid from N to A. Mutant LysK(N77A) is controlled by the 186*pJ* promoter from

				186 phage and a Spectinomycin resistance gene.

2.a.iii Primers

Table 3: Primers Used During the Course of this Thesis. *List contains any of the primers that were used in this thesis. Table includes the primer number, sequence of the primer (5' to 3') and a short description of what the primer was used for.*

Primer Name (Number)	**Sequence (5' to 3')**	**Description**
597 julian_186*attP*(Nhe1) (597)	GCTCAGCTAGCTATGCACTCCT CAGGAAAGTGG	186 *attP* primer; left most region, used to screen for the integration at one of the two 186 attachment sites.
598 julian_186*attP*(Nco1) (598)	GCTCATCCATGGGCGATGGTT CTGAGTAACAGATAATAGAATG G	186 *attP* primer; right most region, used to screen for the integration at one of the two 186 attachment sites.
610 186*attB* left (610)	CTCATTCGAAACCACCCACCG	Used for screening with primer 611 for integration at the primary 186 attachment site. If no integration is present at that site PCR produces a band identifying the *attB* sequences for the primary attachment site.
611 186*attB* right (611)	GATCATCATGTTTATTGCGTGG	Used for screening with primer 610 for integration at the primary 186 attachment site. If no integration is present at that site, PCR produces a band identifying the *attB* sequences for the first attachment site.
1103 #2 186 P1 (1103)	CCCTGGAGCCAAAATATCC	Used for screening with 1104 for integration at the second attachment site, if no integration is present at that site PCR produces a band identifying the *attB* sequences for the secondary attachment site.
1104 #2 186 P4 (1104)	TCCGGAATGCCTGCATTG	Used for screening with 1103 for integration at the second 186 attachment site, if no integration is present at that site PCR produces a band identifying the *attB* sequences for the secondary attachment site.
3053 Spec F	ACAGCGCAGTAACCGGC	Used to amplify the back bone of any pIT4 integration plasmid, located in the Spectinomycin resistance gene.
3054 Spec R	CAGTCGGCAGCGACATCC	Used to amplify the back bone of any pIT4 integration plasmid, located in the Spectinomycin resistance gene.
3149 PIT4_R	GGTATATCTCCTTCTTAAAGTTA AGAGTGTTATTG	Used to amplify one section of the backbone of any pIT4 integration plasmid. The primer is located on the pET RBS sequence, amplifying the fragment located before the start codon of the protein of interest sequence.

3150 PIT4 F	ATTTCCGCTAAATGTCTAGAGC ATG	Used to amplify one section of the backbone of any pIT4 integration plasmid. Primer is located immediately after the stop codon of the protein of interest sequence.
3179 ClyF(C36A) F	GCGTAGACGTTGATGGATACTA CGGTCGCCAGGCCTGGGACTT GCCCAATTACATCTTC	Used to mutate two nucleotides at 106 and 107 position in ClyF from T and G to G and C. Mutating the 36 amino acid from a Cysteine (C, Cys) to an Alanine (A, Ala).
3180 ClyF(C36A) R	CTGGCGACCGTAGTATCCATCA AC	Used to mutate two nucleotides at 106 and 107 position in ClyF from T and G to G and C. Mutating the 36 amino acid from a Cysteine (C, Cys) to an Alanine (A, Ala).
3181 ClyF(H99A) F	TTTGGACAGGAGGTAATTACAA TTGGAATACATGGGGAGCTACT GGCATTGTCGTGGGC	Used to mutate two nucleotides at 295 and 296 position in ClyF from C and A to G and C. Mutating the 99 amino acid from a Histidine (H, His) to an Alanine (A, Ala).
3182 ClyF(H99A) R	TCCCCATGTATTCCAATTGTAAT TACCTCC	Used to mutate two nucleotides at 295 and 296 position in ClyF from C and A to G and C. Mutating the 99 amino acid from a Histidine (H, His) to an Alanine (A, Ala).
3199 ClyF(C36G) F	GCGTAGACGTTGATGGATACTA CGGTCGCCAGGGCTGGGACTT GCCCAATTACATCTTC	Used to mutate one nucleotide at the 106 position in ClyF from T to G. Mutating the 36 amino acid from a Cysteine (C, Cys) to a Glycine (G, Gly). The reverse primer used with 3199 was 3180 ClyF(C36A) R.
3219 186pJ Insert F	TCTTAACTTTAAGAAGGAGATA TACCC	Used to amplify the protein gene fragment for assembly into the pIT4 integration plasmid. The protein gene fragment template must contain the pET RBS binding sequence, which the primer sequence binds to.
3220 186pJ Insert R	CCATGCTCTAGACATTTAGCGG AAATTTA	Used to amplify the protein gene fragment for assembly into the pIT4 integration plasmid. The protein gene fragment template must contain the spacer sequence from the end of 186 located before the terminator sequence on the pIT4 integration plasmid.

2.a.iv Solutions and Buffers

Buffer and solutions were either prepared in the laboratory or purchased from the technical services unit TSU (Level 4, Molecular Life Science Building, The University of Adelaide) and were stored at room temperature unless otherwise stated.

LB Lennox Media (LB Media) Solution

Media contains 5 g/L Sodium Chloride, 5 g/L Bacto Yeast Extract and 10 g/L Bacto Tryptone.

LB Lennox 1.5% Agar (LB Agar) Solution

Agar contains 5 g/L Sodium Chloride, 5 g/L Bacto Yeast Extract, 10 g/L Bacto Tryptone and 10 g/L Agar no.1.

Agar Plates

Agar plates were prepared from 250 mL of molten LB-Agar (LB Media + 10 g/L Agar Powder No. 1). Stock concentrations of reagents required for selection and/or induction were added as required. Plates were left at room temperature to set and then stored at 4°C wrapped in heavy duty glad wrap.

TSS (Transformation & Storage Solution)

To prepare 100 mL, add 80 mL LB + 5 mL DMSO + $MgCl_2$ (to 25 mM from 1 M stock, 1/40 = 2.5 mL) to a beaker. 10 g of solid PEG was added and dissolved using a magnetic stirrer (10-15 min). Keep beaker covered with foil. Check pH and adjust to 6.5 with either HCl or NaOH. Make up volume to 100 mL with LB. Filter sterilize (0.45 µm Minisart single use filter unit (Sartorius) using 50 mL sterile syringe). Aliquot and store at -20°C until use. Ensure it is homogeneous upon thawing.

5x IOS buffer for Gibson assembly

1 mL of 1 M Tris-HCL pH 7.5 (From TSU), 50 µL of 2M $MgCl_2$, 200 µL 10 mM Deoxynucleotide Solution Mix (NEB: N0447L), 100 µL of 1 M DTT, 0.5g PEG-8000 (Sigma-Aldrich), 200 µL of 50 mM NAD (NEB: B9007S) add MQ water to 2 mL, then aliquot (80 µL) and store at -20°C.

Gibson assembly mix:

80 µL 5X IOS buffer, 0.16 µL of 10 U/ µL T5 exonuclease (Epicentre: T5E4111K), 5 µL of 2 U/µL Phusion High-Fidelity DNA Polymerase (Finnzymes: F-530S), 40 µL of 40 U/ µL Taq DNA ligase (NEB:M0208L) add MQ water to 0.3 ml, aliquot (15 µL) and store at -20°C.

Phage Storage Solution (TM Buffer):

0.25mL of 1M stock solution of Tris Solution pH 7.5 and 0.5mL of 1M stock solution of Magnesium chloride. Solution made up to 5 mL with MQ water.

2.b Basic Methods

2.b.i Plasmid Mini-Preps

Plasmid minipreps were performed on 5 mL of overnight culture, grown in the appropriate selective media, using a Monarch plasmid mini-prep kit (NEB#T1010). Purified plasmid was eluted with 30 - 50 µL of the provided elution buffer and stored at -20°C.

2.b.ii Polymerase Chain Reaction (PCR)

PCRs were performed for the purposes of cloning, screening and preparing templates for sequencing. PCRs for cloning were performed using two different PCR protocols, either KAPA2G Robust PCR Kit (KK5023) (KAPA Robust) or NEB PCR Protocol for Phusion® High-Fidelity DNA Polymerase (M0530) (Phusion). PCRs were performed using the buffers and polymerase provided by the respective companies, following their protocols. Annealing temperature and extension time were altered depending on the product length and primer characteristics. PCR products were visualised following agarose gel electrophoresis, followed where required by clean-up and measurement of concentration by Nano-Drop spectrophotometer (Thermo Scientific Nano Drop 2000).

2.b.iii Restriction Digestion

Restriction enzyme digestion reactions were performed using ~1000ng of DNA on either mini-prep DNA or PCR product. 0.8 µL was used for each restriction enzyme added to the digestion. Units differ depending on restriction enzyme, stock concentrations were 5 to 100 U/µL. 2.5 µL of the appropriate buffer for the single or multiple restriction enzyme digests was added to the reaction. Digestion samples in a final volume of 25 µL and incubated at the recommended temperature, either for three hours or overnight. Digestion was determined by visualization using agarose gel electrophoresis.

2.b.iv DNA Clean up

DNA clean-up was performed using protocols and reagents provided by either Monarch PCR and DNA Clean-up Kit (NEB #T1030) for DNA in solution, or Monarch DNA Gel Extraction Kit

(NEB #T1020) for DNA excised from agarose gels. Purified DNA was eluted with 10 µL of the provided elution buffer, then stored at -20°C.

2.b.v Gel Electrophoresis

Either 1% or 1.5% agarose gel suspensions were prepared from molecular biology grade agarose powder (VWR Chemicals), dissolved in 1x TAE prepared from a 20x stock (TSU). The agarose suspension was boiled in a microwave oven until all of the agarose powder had dissolved. Unused gel stocks were kept at room temperature until needed then microwaved when needed for pouring a new gel.

2.b.vi Making Glycerol Stocks

Glycerol stocks for long term storage of bacterial strains were made by adding 500 µL of sterile 80% glycerol (TSU) to 500µL of overnight bacterial culture, mixing and storage at -80°C.

2.b.vii Optical Density Reading of 186 Tum Variants in Liquid Cultures

In addition to the agar plate based method, lysis assays were also performed in liquid cultures, where optical density (OD_{600}) of the cultures were monitored following addition cumate.

Bacterial strains were streaked out from glycerol stocks on to appropriate antibiotic plates and incubated overnight at 37°C. Single colonies were used to inoculate 3 mL of LB media with 20 µg/mL Kan, and the culture grown overnight at 37°C.

Overnight cultures were sub-cultured into 50 mL of LB media, 1/125 dilution, containing (20 µg/µL Kan) and grown to OD_{600} 0.4. Aliquots of 5 mL for each strain were distributed to five 10 mL tubes. Each tube was supplemented with cumate from a 100mM stock solution (in ethanol) to give final concentrations; 0 µM, 40 µM, 120 µM and 240 µM. Dilutions were made such that the ethanol concentration was constant. A no ethanol control was also included.

Two 100 µL replicate aliquots from each culture were added to a flat bottomed 96 well plate (Corning) and OD_{600} readings taken every 15 minutes for three hours, incubating at 37°C between readings. Three biological replicates were performed for

each strain with the exception of one strain that had six biological replicates due to large day to day variation.

2.c Cloning Strategy

2.c.i Plasmid Fragment Cloning

Many different plasmids were created in the project. These are recorded in Table 1 and Table 2. Plasmids were created using PCR amplification and assembled using Gibson isothermal assembly. Backbone plasmid fragments and gene fragments were amplified using the Phusion PCR protocols found in section 2.b.ii Polymerase Chain Reaction (PCR) and 2.c.ii Error Prone PCR. PCR amplification products and expected product sizes were confirmed using agarose gel visualization, section 2.b.v Gel Electrophoresis. PCR products were purified using protocols found in section 2.b.iv DNA Clean up. In cases where the template for the PCR was a plasmid, *Dpn*I restriction enzyme, which digests only methylated DNA, was added directly to the PCR prior to clean-up, and incubated at 37°C for at least 15 minutes (2.b.iii Restriction Digestion).

2.c.ii Error Prone PCR

Error Prone PCR was performed to generate a library of mutants, aiming for amplified PCR fragments that contained three to four mutations per gene. The error prone protocols have been optimised to generate a minimum of one mutation per cycle (153,154). Using the lower fidelity of Taq DNA polymerase, and depending on $[Mg^{2+}]$ and optimisation of unequal dNTP concentrations, error prone PCR can produce mutation rates of 0.6 to 2.0% (155).

Phusion PCR amplification was first performed to generate a template to be used in the error prone PCR. Error prone PCR, 50 µL reactions consisted of 2.5 µL of each primer (10 µM working stocks), 5 µL of 10x Thermopol buffer, 1 µL of dNTP (10 mM), 8 µL of $MgCl_2$ (25 mM) 1.2 µL of Taq DNA polymerase (5 U/µL), 0.5 µL of Phusion template and 29.3 µL of MQ water. All reagents and DNA template were kept on ice while preparing the PCR mixture. Cycling parameters for error prone PCR were adjusted dependent on the size of the fragment being amplified and the number of mutations desired per product. The initial denaturing step was performed at 95°C for 10 seconds. For each cycle denaturing was also performed 95°C for 10 seconds,

annealing was performed at 50°C for 20 seconds, and extension for a 782 bp gene fragment was performed at 68°C, for 1 minute.

Denaturing, annealing and extension steps were repeated eight times to generate gene fragments containing two to four mutations. The number of cycles were calculated using: (0.00066 x bp Length) x number of cycles = number of mutants per fragment amplified. This calculation provided a guideline for generating the desired number of mutations. The number of cycles calculated for one to four mutations was 8 cycles. A final extension step was performed at 68°C for 5 minutes. PCRs products were visualized on an agarose gel, followed by clean-up (Monarch kit, NEB) and measurement of concentration by Nano-Drop spectrophotometer (Thermo Scientific Nano Drop 2000).

2.c.iii Plasmid Fragment Assembly - Gibson assembly

Fragments generated using PCR were assembled into plasmids using Gibson isothermal assembly (156). Approximately 175 ng of purified PCR fragment was added to 15 µL of Gibson assembly aliquot mix at a 1:1 molar ratio for each of the fragments, unless stated otherwise. Deionized water was added to bring the total reaction volume 20 µL, which was then incubated at 50°C for 1 hour. Following incubation, the Gibson assembly mixture was cleaned using the DNA clean-up methods found in Chapter 2.b and then transformed into chemically competent cells (TSS method) or electrocompetent cells (ECC).

2.c.iv Preparing and Transforming Chemical Competent Cells (TSS cells)

For routine transformations, the TSS method of preparing chemically competent cells was used (157). A 1.5mL aliquot of an overnight culture of the appropriate strain was placed in a cold Eppendorf tube, and spun at 14,000 rpm (15,777 g) using a refrigerated table top centrifuge (Eppendorf Centrifuge 5415 R) at 4°C for 5 minutes. The supernatant was carefully removed and the pellet resuspended in 150 µL of TSS (2.a.iv Solutions and Buffers). The resuspended cells were kept on ice if used the same day, or stored at -80°C. For transformation of TSS competent cells, 50 µL of TSS cells were mixed with either 5 µL of Gibson cleaned up assembly mix or 2 µL of plasmid mini-prep (2.b.i Plasmid Mini-Preps) and incubated on ice for thirty minutes. The cells were heat shocked at 42°C for one minute and thirty seconds, and placed back on ice for five minutes. LB Lennox media (250 µL) was added to the cells,

followed by recovery at 37°C, unless stated otherwise, for one to two hours. After incubation, cells were pelleted using a table top centrifuge (Eppendorf Centrifuge 5415 R) for five minutes at a force not exceeding 6000 rpm (2897 g). The majority of the supernatant was removed, leaving ~100 µL of supernatant, in which the pellet was gently resuspended and spread on an agar plate containing the appropriate antibiotic. Plates were then incubated at the appropriate temperature.

2.c.v Preparing and Transforming Electrocompetent Cells (ECC)

When high transformation efficiencies were required, electrocompetent cells were used. Overnight cultures (3-5 mL) of the required strain were prepared. The next day, 400 µL from the overnight culture was added to a fresh 50 mL culture (1/200 dilution) and grown at either 30°C or 37°C with shaking (bench top water bath,180 to 200 rpm) to an OD_{600} of 0.4.

If the strain also contained a helper plasmid (expressing recombineering functions or integrases), a pre-induction step was included to initiate expression of those functions. For the recombineering helper plasmid pSIM6, which carries the lambda Red recombineering proteins (*gam, exo, bet*), a heat induction step was performed at 42°C for 15 minutes, with shaking at 100 rpm, then chilled in an ice water slurry. phi21, Att2 (AH6045) and Att3 (AH6046) helper plasmids (149) were used to integrate plasmids at the corresponding chromosomal attachment sites. Cells carrying any of these plasmids were pre-induced at 39°C for 20 minutes (phi21), or 30 minutes (att2 or att3) with shaking at 100 rpm, followed by chilling in ice water slurry.

After heat induction of the required functions, the 50 mL cultures were chilled for 5 minutes and pelleted at 4°C 4500 g for 10 min using a high speed centrifuge (Eppendorf Centrifuge 5810 R). Supernatant was poured off and the pellet was gently resuspended in ice cold MQ water. This step of pelleting and resuspending in ice cold MQ water was repeated a second time before pelleting and resuspension in 10mL of ice cold 10% glycerol. The cells were pelleted one more time, and finally resuspended in 250 µL of ice cold 10% glycerol. The cells were aliquoted into ~40 µL lots and either transformed or stored at –80°C. ECC prepared and immediately transformed had the highest efficiently compared to ECC samples stored frozen and thawed for use.

ECC cells were transformed using a BIO-RAD MicroPulser electroporator, with between 2-5 µL of Gibson assembly/plasmid mini-prep added to a 40 µL aliquot of cells on ice. The transformation sample was transferred to a cold electroporation cuvette. Electroporation used the Ec1 setting resulting in an ideal time constant of between 5.5 and 6.0 msec. LB Lennox media (1 mL) was added to the cuvette immediately after electroporation and transferred into an Eppendorf tube, followed by incubation at 37°C for one and a half hours as a recovery step. Cells were then pelleted using a table top centrifuges for five minutes at speeds not exceeding 6000 rpm (2897 g). The cell pellet was resuspended in ~ 100 µL of LB and plated onto an agar plate containing selective media. Plates were incubated overnight at 37°C, unless stated otherwise.

2.c.vi Sanger Sequencing

Sanger sequencing was performed by the Australian Genome Research Facility (AGRF), Adelaide node. DNA samples, usually PCR fragments, were cleaned and the concentration determined using a Nano-Drop spectrophotometer (Thermo Scientific Nano Drop 2000). About 100 ng of clean DNA was added to the Big Dye sequencing reaction, along with 1 µL of sequencing primer (10 µM concentration). The final volume was made up to 13 µL using MQ water ready for collection by AGRF.

2.d Bacteriophage Integration for making 186 lysogens

A lysogen is bacterial cell which has been infected by either a temperate or filamentous phage that does not cause host cell death. Bacteriophage 186 is a temperate phage which can integrate into two alternative attachment sites in the *E. coli* chromosome (158). To create a 186 lysogen at one of integration sites, the host strain E4643 was grown to OD_{600} 0.6, 200 µL of culture was added to 3 mL of 0.7% agar and spread on an agar plate. The agar plate was left to dry at room temperature, then 10 µL of wild type bacteriophage 186 stock was spotted on the plate and incubated overnight at 37°C. The following day, cells from the centre of the spot (containing potential lysogens) were streaked out, and plates incubated overnight at 37°C.

To screen for potential lysogens, resistance to infection by a lytic variant of 186 (186CI10) was used. In this assay, 10 µL of 186CI10 phage stock was pipetted onto the plate, and one side of the plate was lifted, allowing the 10 µL of phage solution to run down the plate. After the line of phage stock had dried, several individual colonies from the re-streak plate (potential lysogens) were spread across the line of 186CI10 phage. Two additional control strains were streaked across the plate in the same manner: a positive control strain carrying a 186 prophage that will be resistant to infection by the 186CI10; and a negative control, being a non-lysogenic strain were the cells that come into contact with the CI10 phage will be killed. Plates were examined after an overnight incubation at 37°C. Streaks that produced a continuous line of growth, were expected to be lysogens, as they are resistant to reinfection (immune). These were re-streaked and individual colonies further checked by PCR.

PCR was performed both to confirm the presence of a 186 prophage and to determine which of the two possible attachment sites (159) it had integrated into. PCR screening for site 1 used four primers; 610, 598, 597 and 611, using protocols found in 2.b. iv. These primers amplify the left and right boundaries of the prophage integration sites, *attL* and *attR*. If there is a prophage present at the first integration site, PCR will produce two bands of 335 bp and 241 bp identifying the *attL* and *attR* sites. If there is no integrant at site one, a band (241 bp) corresponding to the empty attachment site (*attB*) will be produced.

Integration at the second attachment site on the E4643 chromosome is less frequent than at site 1, but was checked nonetheless (159). A different pair of primers are required; 1104, 598, 597 and 1103, which would produce two bands, 316 bp and 259 bp, for site 2 *attL* and *attR*, respectively. Should there be no integration at the second attachment site, the PCR will produce a 601 bp *attB* band. Four controls for PCR screening of phage integration were used including two positive controls, E573 which contains a single 186 prophage, and E4170 (E4170, Table 1) which contains a 186 prophage integrated at both site 1 and site 2. The negative control is a non-lysogenic strain (C600). Phage stock was also used as a control for the *attP* band, 335 bp.

2.e Bacteriophage production and Integration

Bacteriophage stocks were created using two methods: a single plaque method which gave moderate titres and longer whole plate method, that produced a higher

titre phage stock. Single plaques were picked using a Pasteur pipette, added to an Eppendorf tube containing between 100 µL and 300 µL of TM buffer and vortexed for 20 seconds. Samples are treated with a drop of chloroform to kill any bacteria, and stored at 4°C.

Whole plate phage stocks are prepared by with mixing phage with 200 µL of a host strain and 3 mL of 0.7% agar, followed by spreading on a LB Lennox agar plate and overnight incubation at 37°C. The next day, 5 mL of TM buffer was added to the plate and the top layer of agar scraped off. The buffer and top layer of agar are collected in a 10 mL sterile tube, vortexed briefly and treated with chloroform. The sample is centrifuged to pellet the agar, and the supernatant containing the phage collected and stored at 4°C.

2.f Recombineering Strategy

Recombineering (160) was used to delete a section of the 186 prophage containing the *tum* and *ORF97* genes. pSIM6 was the recombineering helper plasmid used, and carries an ampicillin resistance gene, a temperature sensitive origin of replication and the lambda Red recombineering genes (gam, exo, bet) (Table 2). Recombineering functions were pre-induced by a short incubation at 42°C while generating electrocompetent cells (2.c.v).

The DNA template used for recombineering was double stranded G-block DNA fragment (IDT), containing a *loxP* flanked chloramphenicol resistance gene with overlap sequences (25 bp minimum) at each end, matching the target sequence. Template DNA was transformed into ECC cells using the protocols found in Chapter 2.c.iv. Colonies growing on LB + Cm (20 µg/mL) plates were screened by PCR to confirm recombineering, and the final sequence confirmed by Sanger sequencing.

2.g Integration of Plasmids at Phage Attachment Sites

A number of genetic modifications were performed using chromosomal integration, both into the E4643 chromosome itself and into the bacteriophage 186 prophage. The three different cumate induction module plasmids; pIT4-KT-cymR-pCym-empty, PIT4-KT-cymR-pCym-*Tum* and PIT4-KT-cymR-pCym-*Tum72* (Table 2), were integrated into the phi21 bacteriophage attachment site in the E4643 genome. The

ampicillin resistant phi21 helper plasmid carries the phi21 *integrase* under the control of a temperature sensitive repressor, and so the integrase is only expressed at 37°C.

The helper plasmid was transformed into the desired bacterial strain using TSS-competent cells (Chapter 2.c.iv Preparing and Transforming Chemical Competent Cells (TSS cells)), the recovery step and overnight incubation being done at 30°C. ECC (Chapter 2.c.v Preparing and Transforming Electrocompetent Cells (ECC)) were prepared from the strains containing the helper plasmid, with growth of the culture to OD_{600} 0.6 at 30°C. The heat induction step for the phi21 plasmid is for 20 minutes at 39°C. After the heat induction step, the transformation follows the same protocol listed in Chapter 2.c.v, with 2 µL of DNA (~100 ng) added to the electrocompetent cells.

2.g.i Gene Removal using LoxP Sequences

An engineered 186 prophage was constructed used recombineering (methods below) with a module containing an antibiotic resistance gene. The chloramphenicol (Cm) antibiotic resistance gene was included in the recombineering module as a selection marker for successful recombineering. The Cm gene is flanked by LoxP sequences to allow removal of the antibiotic marker, using a Cre recombinase (159,161) delivered using a cosmid based delivery system (158).

Single colonies of potential recombinants were streaked out on to LB agar plates containing 20 µg/µL of Cm and incubated overnight at 37°C. The next day a single colony was used to inoculate 3 mL of LB media containing 20 µg/µL of Cm and incubated overnight. In an Eppendorf tube, 5 µL of this overnight culture was added to 5 µL of a phage λ based cosmid, carrying the pE-Cre plasmid (158).

Samples were incubated for 30 minutes at 37°C to allow binding and DNA injection by the cosmid particles, followed by addition of 1 mL of fresh LB media. Samples were then incubated at 30°C up to 7 hours to allow the Cre recombinase to act. One hundred µL of a 1/10 dilution of this culture was plated on an LB agar plate containing Amp (100 µg/µL) to select for the Cre plasmid, and incubated overnight at 30°C.

The next day single colonies were collected using a pipette tip and transferred to both an LB plate (non-selection) and an LB Cm 20 µg/µL (selection) plate. Up to fifty

single colonies were screened in this way. Both plates were incubated overnight at 37°C to all removal of the pE-Cre plasmid, which carries a temperature sensitive origin of replication. Potential recombinants that grew on the non-selection plate but not the selection plate were re-streaked and confirmed by PCR and Sanger sequencing

2.h PHEARLESS Protein Screening Protocols

Protocols for the PHEARLESS system depend on what is being tested, multiple expression strains for the PHEARLESS system were generated. Protocols for the PHEARLESS system follows one of two aims, either screening an initial protein gene for activity or screening a library of protein genes for activity. Screening an individual protein gene for lytic activity can be accomplished in the first version of the expression strain (PHEARLESS V1, Chapter 3) as well as the second version of the expression strain (PHEARLESS V2.2, Chapter 5). The protocols for screening an individual protein gene for lytic activity are as follows.

Expression and target strains are streaked out from glycerol stocks onto agar plates using the appropriate antibiotic selection. Plates are incubated overnight at 37°C, the next day 3 mL cultures were prepared from a single colony for each strain using the same selection and incubation temperatures.

These strains include three protein expression strains, and a positive control strain containing a known lytic protein. An empty control strain that does not contain a protein gene was used as the negative control. The final expression strain contains the protein (for example, endolysin) gene that is being tested for lytic activity against a specific bacterial strain (target strain).

The next day, fresh 50 mL LB + antibiotic cultures are inoculated from the overnight cultures and incubated at 37°C, shaking at 180 to 200 rpm. Antibiotic selection in the new cultures is required for PHEARLESS V1 expression strains but not for PHEARLESS V2.2 expression strains. The target strain cultures were started first, after an hour incubation, the expression strain cultures were started. After the target strain reached OD_{600} 0.6, 200 µL of culture is added to a sterile glass tube containing a specific concentration cumate. 3 mL of molten 0.7% agar (held 48°C) is poured into the glass tube containing the target strain from a separate sterile glass tube. The glass tube containing the target strain, cumate and molten 0.7% agar was briefly

vortexed and then poured onto a LB agar plate containing the same concentration of cumate. The agar plate is immediately swirled to spread the molten 0.7% agar over the surface of the agar plate. If using PHEARLESS V2.2 expression strains, a second strain labelled as the propagation strain (Chapter 5) is added at a 50:1 ratio of target strain to propagation strain. If using the PHEARLESS V1 expression strains only the target strain is added to the top layer of agar. This top layer will form a lawn of bacteria; the plates were left to dry at room temperature for a minimum of an hour. Before adding the 0.7% molten agar to the plates, plates pre-warmed at 37°C for a minimum of five minutes. This 5-minute incubation at 37°C is also recommended before adding the expression strain samples. This step will assist in preventing the formation of puddles on the agar plate that can dilute the protein samples and disrupt the lawn.

The plates were then divided into three sections, one for each of the strains; positive control, negative control and the protein being tested. The expression strains were grown at 37°C to OD_{600} 0.6 and placed on ice. After all the cultures have grown to OD_{600} 0.6, 1mL of each is taken and pelleted in an Eppendorf tube and resuspended in 1 mL of fresh LB media to remove any residual antibiotic. If no antibiotic selection was used to grow the new culture, this step can be omitted.

A series of 1 in 10 dilutions is made for each of the expression strains in fresh LB media, 50 µL of culture + 450 µL of fresh LB media. Each plate will be spotted with 10 µL of undiluted, 1/10, 1/100 and 1/1000 diluted expression strain culture. Samples can be added to the plate at room temperature. After the samples have dried, the plates were placed at 37°C to incubate overnight. Plates were observed the next day for the presence of clearing in the location were the expression strain samples were spotted.

For testing weak protein activity, an additional step can be included, after the expression strain reaches OD_{600} 0.6, 10 mL is taken and concentrated 10-fold into 1 mL. A series of 1 in 10 dilutions were created from this more concentrated sample.

2.i PHEARLESS Protein Library Screening Protocols

PHEARLESS protein library screening protocols were developed for the PHEARLESS V2.2 expression system. The aim for this set of protocols is to perform a high throughput assay, screening a library of endolysin variants.

Protocols for this assay includes previous protocols detailed above; Phusion PCR and Error Prone PCR, DNA Clean Up, Gibson Assembly and Making and Transforming Electrocompetent Cells. Phusion PCR is used to generate the fragments of the pIT4 integration plasmid (detailed in Chapter 5), error prone PCR is used to generate a library of mutant protein genes. Gibson assembly is performed with the pIT4 integration plasmid fragments and the error prone library products. The Gibson assembly is cleaned up using DNA Clean up protocols and stored at -20°C. These steps can be performed in advance and the assembled reaction stored frozen.

Chapter 3: Design and construction of the PHEARLESS system

3.a Generating the PHEARLESS Expression Strain

PHEARLESS stands for **Ph**age-based **E**xpression, **A**mplification and **R**elease of **L**ytic **E**nzyme **S**pecie**s**. This *in vivo* system has been developed to test for the anti-microbial activity of any protein of interest, such as endolysins, that can be expressed in E. coli. The core design feature of the PHEARLESS system is the ability to simultaneously (i) initiate expression of the protein-of-interest and (ii) initiate a cascade leading to cell lysis and release of the cellular contents, via induction of a modified prophage. Co-culturing of the PHEARLESS strain and the target bacteria of interest (here *S. aureus*), allows for an assessment of whether the released protein of interest has an impact on the growth of the target strain (Figure 8). The PHEARLESS strain was designed to initiate both protein-of interest expression and cell lysis via induction of a 186 prophage, through the use of the chemically induced expression of the 186 Tum protein. Tum is an anti-repressor protein which inactivates the 186 lysogenic repressor, leading to expression of the early, middle and late lytic genes. Protein of interest expression is controlled by a phage 186 promoter which becomes active following Tum expression. Phage induced cell lysis is achieved using two components that have been genetically engineered into the E4643 chromosome. (1) An integrated cumate-inducible 186*tum* expression module and (2) an engineered ∆186 prophage (Figure 7).

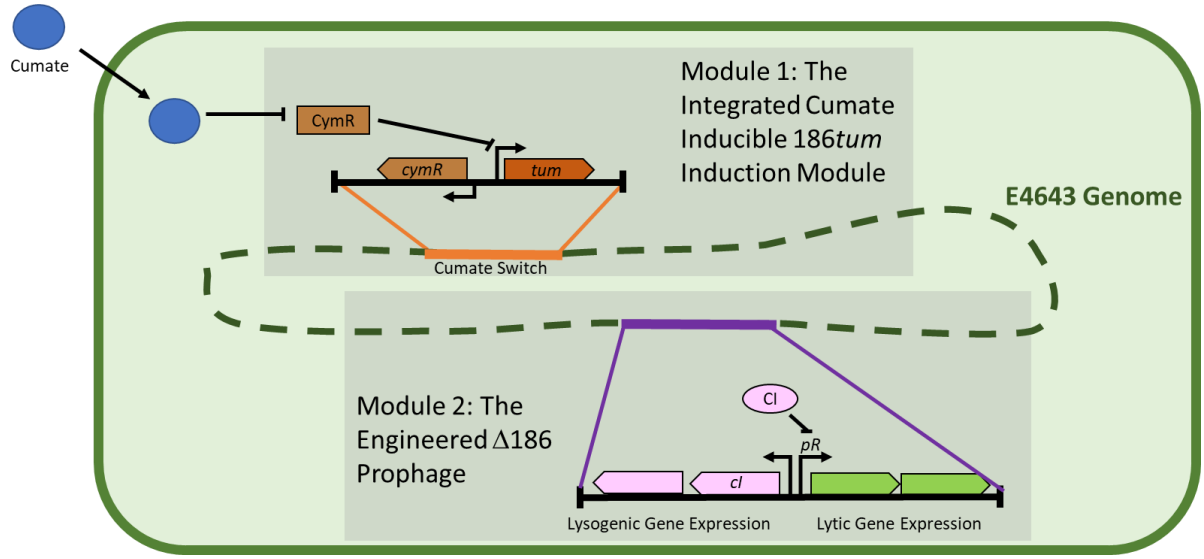


Figure 7: PHEARLESS Expression Strain Modules. *The two main modules used in the PHEARLESS system to bring about cell lysis. This diagram excludes the method used for endolysin/protein expression, which differs between the versions. Module 1 is a cumate inducible 186tum switch, integrated into the E. coli (strain E4643) genome at the phi21 integration site. Module 2 is an engineered 186 prophage, where the tum gene has been removed. The 186 lytic cycle is repressed by CI repressor (pink oval) which maintains the lysogenic cycle. In the absence cumate (blue circle), CymR which is constitutively expressed, represses the expression of Tum. When cumate is added it binds to CymR (brown square), weakening CymR binding, leading to Tum expression. Tum sequesters 186CI, inducing the prophage, leading to cell lysis.*

Tum, whose expression in a prophage is normally repressed by host LexA protein, is required for 186 prophage induction. *E. coli* 's DNA damage repair, or SOS, the system is activated in response to many different DNA damaging agents and environmental factors including UV radiation (162).

In the event of DNA damage, *E. coli*'s SOS system is turned on via activation of RecA in response to increased levels of single stranded DNA. Activated RecA accelerates auto-cleavage of LexA, leading to de-repression of many SOS related genes. The 186*p95* promoter, which controls Tum expression, is thus also de-repressed (163). When Tum is expressed it inactivates the 186 CI repressor protein which is responsible for repression of the lytic promoters *pR* and *pB.* De-repression of the *pR* promoter results in expression of the early lytic genes, ultimately leading to production and release of new phage particles.

The cumate inducible 186 *tum* induction module (cumate switch) was created by Alejandra Isabel of the Shearwin Laboratory (145). A pIT4-based chromosomal

integration plasmid containing the cumate switch was integrated into the *E. coli* E4643 genome at the phi21 attachment site, using protocols found in Chapter 2.g (Table 2). The plasmid produces the *cymR* repressor from the constitutive proA promoter. The plasmid also carries the 186 *tum* gene, controlled by a promoter, pCymR, which is repressed in the presence of CymR (164). Addition of cumate (4-isopropylbenzoic acid) removes the CymR repressor from the cumate operator (CuO), which in turn leads to the expression of the Tum gene. The pIT4 plasmid also contains a kanamycin resistance gene and a R6Kγ origin of replication which requires the *pir* gene for replication. Three versions of the cumate switch plasmid were created; Empty, *tum* wild type and *tum72*. The Empty version contains all the components with the exception of a *tum* gene, it is used as a negative control. Two versions of Tum were tested in the system the 186 wild type, full length (142 amino acids) Tum and a truncated version, consisting of the first 72 amino acids, called Tum72. Tum72 is a hyperactive variant of 186 Tum, where the truncation is thought to slow the degradation rate *in vivo* (145), resulting in a more rapid switching to the lytic cycle.

The second module of the chemically inducible lysis system is the engineered 186 prophage genome, integrated at the primary 186 attachment site. The 186 genome was engineered to remove the native *tum* and *cp97* genes using recombineering. Protocols found in Chapter 2.f Recombineering Strategy. A 186 phage lacking only the *cp97* gene has been shown to have no inhibitory effect on phage production (165). The *tum* and *cp97* genes were replaced (Figure 9.A) with a custom designed module, containing a *LoxP* flanked selection marker (chloramphenicol), as well as two well characterised phage attachment sites (*attB*), *attB2* and *attB3*, that originate from *non-E. coli* species (147). These were chosen since they are unrecognisable attachment sites for phage commonly found in *E. coli*, preventing unintended site specific integration by *E. coli* phage. The attachment sites sequences and resistance gene were incorporated into the 186 genome using recombineering of a G-Block sequence ordered from IDT (Appendix Fig. 1, Appendix Fig. 2). Successful recombineering was confirmed using PCR to amplify the left and right boundaries of the inserted recombineering fragments and also by PCR across the whole recombineered fragment (Appendix Fig. 3), Sanger sequencing was used to confirm the correct nucleotide sequence.

To summarise how these modules work together: in the un-induced system, constitutive CymR expression represses the *tum* gene (Figure 8. A). In the engineered 186 prophage, the Cl protein represses the activation of the native *pR* promoter, maintaining the lysogenic life cycle (Figure 8. A). After addition of cumate, CymR mediated repression of *tum is lost*, which results in *Tum* activating the prophage lytic cycle, by sequestration of CI. When CI is sequestered it is unable to repress the native *pR* promoter, which results in expression of the early lytic cycle genes (Figure 8. B). In addition to activating the lytic cycle, expression of a protein of interest from the protein expression plasmid occurs in parallel (Figure 8. B), activated by the same genetic elements as the phage lytic cycle. Activation of the lytic cycle eventually results in lysis of the host cell, releasing the protein of interest that has been produced throughout the lytic cycle (Figure 8. C). If the target bacterial cells are sensitive to the released antimicrobial protein, their growth will be impacted (Figure 8. C).

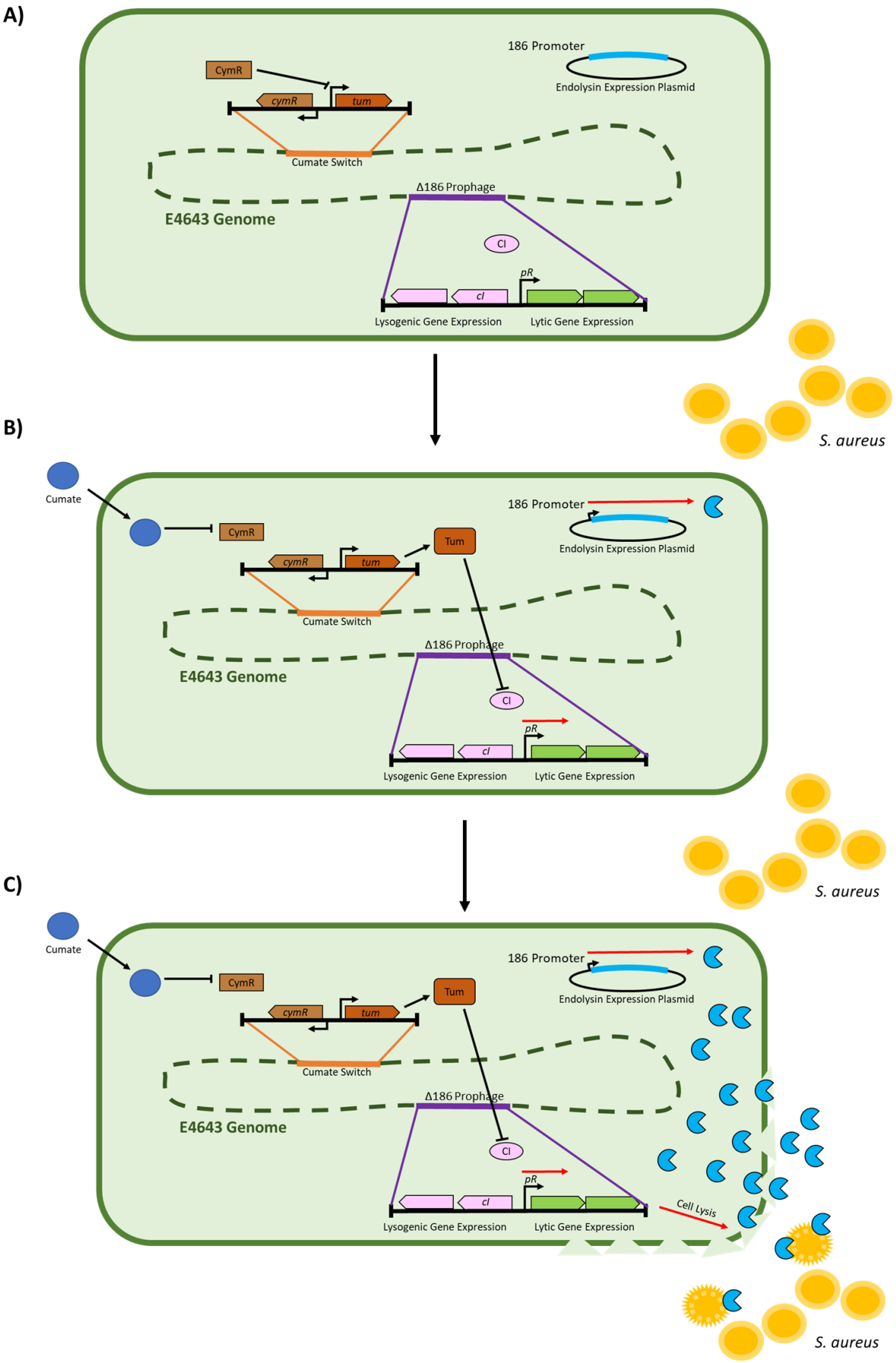


Figure 8: Activation of PHEARLESS V1 endolysin expression and cell lysis (Plasmid Version). *A) Show that without the presence cumate in the system, CymR is expressed and represses tum. Repressing Tum maintains the Engineered 186 in a lysogenic life cycle via CI (Pink Circle) repressing the early lytic*

gene expression promoter pR. The protein expression plasmids present in the cell is controlled by a 186 promoter which is also activated during the lytic cycle. The yellow circles located outside of the expression strain (Green) represent S. aureus cells. B) shows that activation of Tum expression by cumate entering the cell and sequestering CymR protein, preventing tum expression from being repressed. Tum inactivates CI which results in expression of lytic genes cassette (red arrow). When the 186 promoter is activated the protein expression plasmid is active. The protein used for the development of the expression strain contains muralytic activity against S. aureus. C) shows at the end of the of the lytic cycle the host cell undergoes cell lysis, releasing 186 phage viruses (not shown). The protein from the expression plasmid has built up in the cell and is also released apron cell lysis. When it is released into the surrounding environment, it interacts with S. aureus causing the cells to lyse.

Two variants of the engineered 186 were designed and created. The first G-Block was designed to remove enough *cos* sequence of 186 (Figure 9. B), to render it unable to package DNA and generate functional phage particles. The 186 *cos* sequence is critical for genome packaging into the phage head during the final stages of progeny phage assembly. A second version of the expression strain (Figure 9. C) was designed to leave the *cos* sequence intact and thus capable of producing functional 186 (Figure 19). Later chapters that detail PHEARLESS V1 and PHEARLESS V2.2 will state which variant is being used.

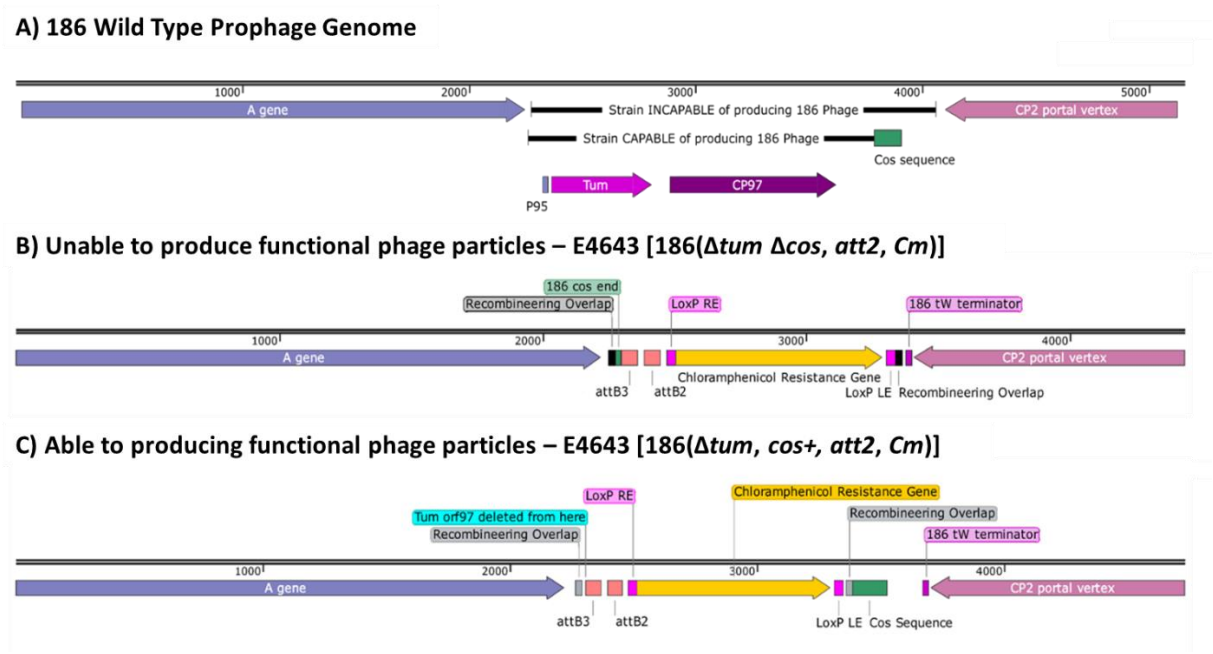


Figure 9: Map of the Wild type 186 Prophage genome and Engineered 186 Prophage Genome. *A) shows a section of the wild type 186, containing the cos sequence, A gene, tum, CP97 and CP2 portal vertex genes. The tum gene is responsible for activating the lytic life cycle, its expression is activated*

by the depression of the P95 promoter which is controlled by the host SOS system. The cos sequence located before CP2 portal vertex is essential for 186 genome packaging. Black lines labelled as "Strain INCAPABLE of producing 186 Phage" and "Strain CAPABLE of producing 186 Phage" show the fragments that were removed to generate the engineered 186 prophage genomes. B) Shows the expression strain which has had tum and cp97 removed and replaced with two S. aureus phage attachment sites (attB3, attB2) with a chloramphenicol antibiotic resistance gene flanked by LoxP sites. The expression strain E4643 [186(∆tum, ∆cos, att2, Cm)] (HB61) has had much of the cos sequence removed, and is unable to produce functional phage particles. C) Shows the second expression strain made; E4643 [186(∆tum, cos+, att2, Cm)] (HB93) which contains the same features as the previous strain (△tum, att2, Cm) but retains the entire cos sequence. Figure was generated using SnapGene DNA maps.

A liquid cell lysis assay was performed to confirm cell lysis (Chapter 2.b.vii Optical Density Reading of 186 Tum Variants in Liquid Cultures), aiming to observe the kinetics of cell lysis at different cumate concentrations. Expression strains tested included the Empty (HB66), Tum wild type (HB70), Tum72 (HB67). Three cumate concentrations were chosen: 40 µM, 120 µM and 240 µM The range of concentrations tested was based on the previous work done during the module's development (145). Strains were freshly grown from overnight cultures to OD_{600} 0.4. These cultures were then divided into five samples and cumate added to final concentrations of; 0 µM, 40 µM, 120 µM, and 240 µM. Aliquots were dispensed into a 96 well plate in duplicate and OD_{600} readings taken every 15 minutes. In between reads, samples were incubated at 37°C with shaking.

The empty strain that contains no *tum* gene was expected to grow at a normal rate in all the samples since it should not be able to induce cell lysis. The no cumate (Empty) sample was used as the baseline negative control. The lysis assay was for each strain was repeated a minimum three times using separate cultures grown from a glycerol stock. The Empty *tum* expression strain (E4643 [pIT4-KT-cymR-pCym-empty]$^{\phi 21}$ [186(∆*tum, ∆cos , att2, Cm*)]) (HB66) OD_{600} results (Figure 10) show that each condition produces a normal growth curve, confirming that the cumate itself has no effect on cell growth.

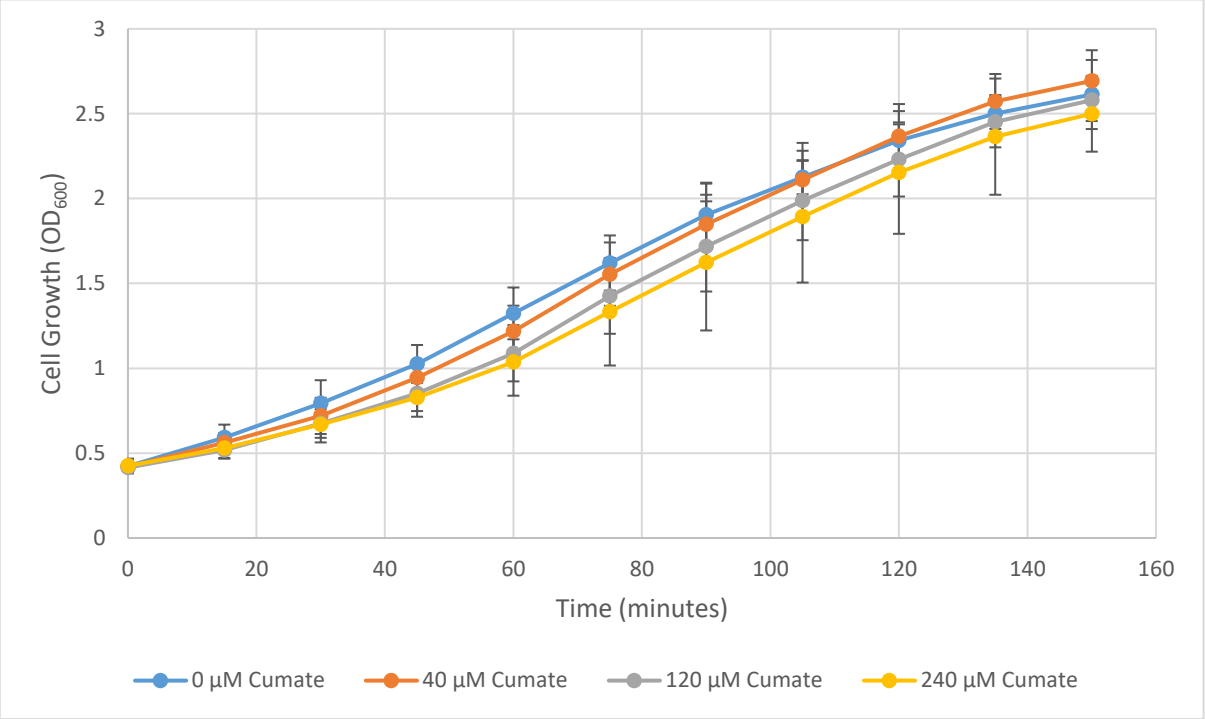


Figure 10: Empty tum Expression Stain (HB66) Liquid Culture Lysis Assay. *Graph shows cell growth over a two-and-a-half-hour period, measurements taken at 15 minute intervals. The empty tum expression strain HB66 (E4643 [pIT4-KT-cymR-pCym-empty]$^{\phi 21}$ [186(Δtum, Δcos, att2, Cm)]) was tested for cell lysis using four different cumate concentrations. 0 µM cumate (Dark Blue), 40 µM cumate (Orange), 120 µM cumate (Grey) and 240 µM cumate (Yellow). Each assay contains two technical replicates per sample and the assay was biologically replicated three times. The curves show that cumate has little effect on cell growth and that no lysis (decrease in OD) was observed.*

The Tum72 based switch (HB76) appeared to be (Figure 11) very effective at lysing the expression cells on the agar plates. The OD_{600} results show a more detailed kinetic picture of the rate of cell lysis at different cumate concentrations. In the absence of cumate the strain grows well, though slightly slower than the corresponding empty strain, perhaps due to some “leak” of Tum72.

In the presence of cumate (40 µM, 120 µM and 240 µM), all there was clear cell lysis. The cells grew like the empty controls up to OD_{600} ~0.65-0.80, for approximately 30-minutes. After 30 minutes, cell lysis begins to occur with OD_{600} plateauing at 45-minutes and then dropping to OD_{600} 0.1 after 1 hour, where it remains.

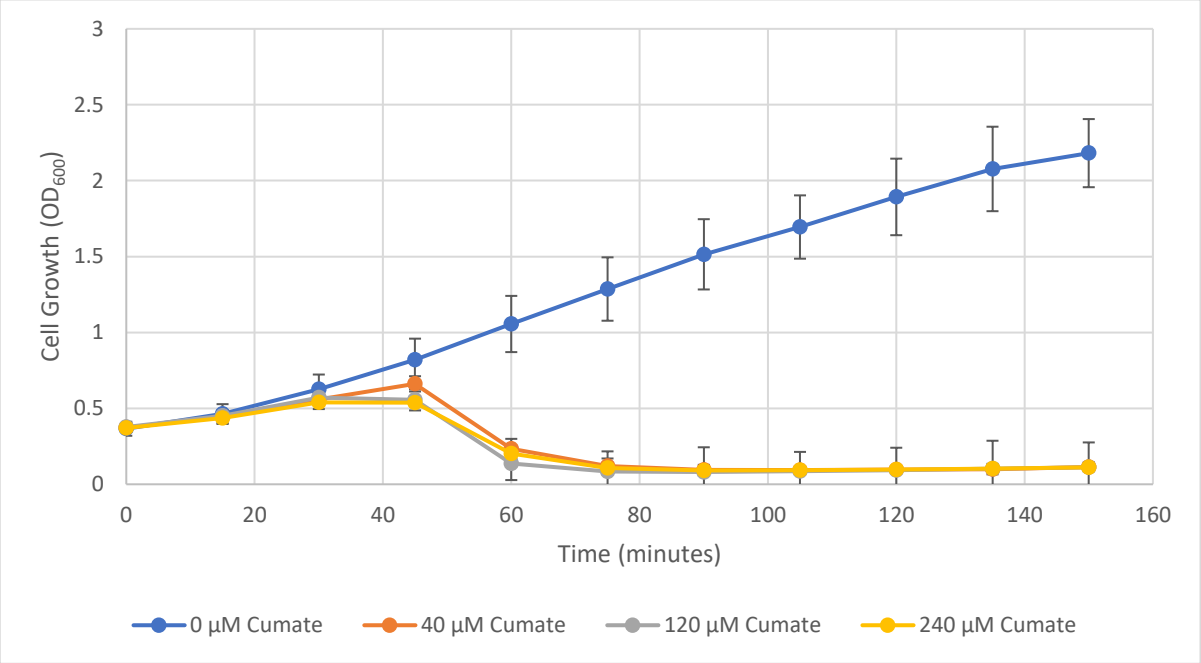


Figure 11: Tum72 Expression Stain (HB67) Liquid Cultures Lysis Assay. *Graph shows cell growth over a two-and-a-half-hour period, measurements taken at 15 minute intervals. The Tum72 expression strain HB67 (E4643 [pIT4-KT-cymR-pCym-Tum72* $^{\phi 21}$ *[186(Δtum, Δcos, att2, Cm)]) was tested for cell lysis using four different cumate concentrations. 0 µM cumate (blue), 40 µM cumate (orange), 120 µM cumate (grey) and 240 µM cumate (yellow), all samples that contain cumate also contained 12 µL of ethanol. Each assay contains two technical replicates per sample and the assay was biologically replicated three times.*

The assay of the Tum wild type switch was repeated a total of six times. The OD_{600} results (Figure 12) presented shows the averages for six repeats of lysis results for Tum wild type (E4643 [pIT4-KT-cymR-pCym-*Tum*]$^{\phi 21}$ [186(Δ*tum, Δcos , att2, Cm*)]) (HB70).

Greater variation in growth was observed at intermediate concentrations of cumate for this strain. The 120 µM and 240 µM cumate replicates produced consistent results with small error bars (Figure 12). Both samples produced a similar growth curve to Tum72, with growth continuing until the 30-minute mark, followed by the onset of lysis. Noticeable cell lysis occurs after 45 minutes, somewhat slower compared to Tum72 (Figure 11). The Tum wild type OD_{600} does not fall as far as Tum72, only falling to OD_{600} 0.4-0.3. This suggests a significant portion of cells are lysing, but that some cells stop growing (no increase in OD) but not lyse.

The 40 µM cumate sample showed significant well to well variation, evidenced by the large error bars. The 40 µM cumate concentration thus appears to be on the threshold of activating the switch to cell lysis.

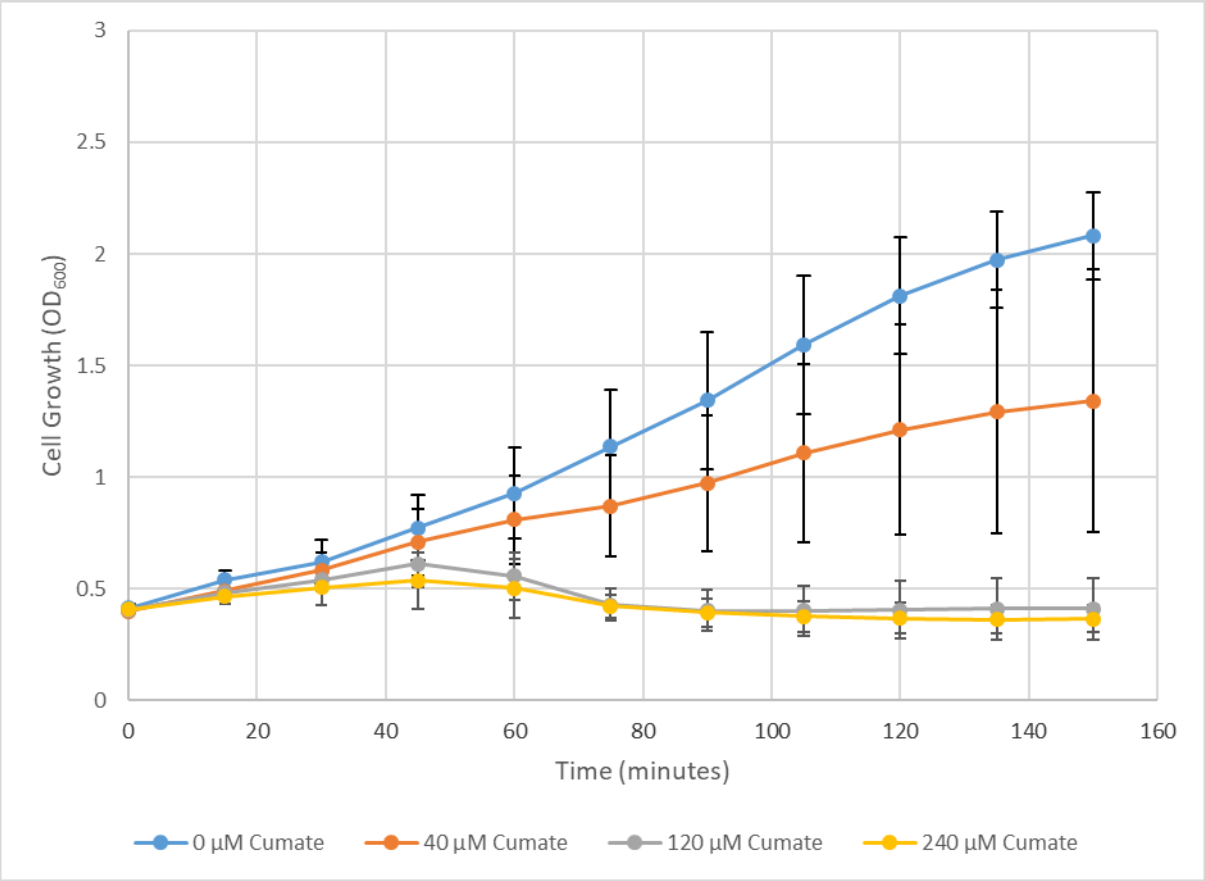


Figure 12: Tum Wild Type Expression Strain (HB70) Liquid Cultures Cell Lysis Assay. *Culture grown in the presence of 40 µM cumate (orange), 120 µM cumate (grey), 240 µM cumate (yellow) and 0 µM cumate (blue). All samples that contain cumate also contained 12 µL of ethanol. Controls confirmed that the low levels of ethanol had no effect on cell growth. Each assay contains two technical replicates per sample and the assay was biologically replicated six times.*

3.b PHEARLESS V1 Expression Strain Background and variations

Having shown that the Tum-based, chemically-inducible lysis module (a chromosomally integrated CymR-pCym-Tum module plus a 186 Δ*tum* prophage) was capable of inducing cell lysis, the second aspect of the PHEARLESS that needed to be developed was the protein expression module (Figure 8). In designing this second module, a key concept was that the lysis and expression modules would be linked, such that addition of cumate would also trigger expression of the putative antimicrobial protein of interest. This should be achievable by placing the protein of interest under the control of a bacteriophage 186 promoter which is either de-

repressed (186*pR*) or activated (186*pJ*) as a consequence of Tum expression. The native *pR* promoter is de-repressed as a result of sequestration of the CI repressor by Tum (163), while the native *pJ* promoter is activated by the 186 late activator protein B, whose expression is turned on following de-repression of the *pB* promoter by Tum (166). A second design consideration for the expression module was the level of expression of the putative antimicrobial protein required. Too little expression would be ineffective, while too much expression could be toxic to the *E. coli* host prior to induction of lysis. Expression levels can be tuned in a number of ways, including plasmid copy number, ribosome binding site strength or mRNA stability. For ease of experimental manipulation, changes to plasmid copy number via mutation of a well characterised origin of replication was the approach taken here.

As described above, the protein expression plasmid has been designed with a 186 promoter that induces expression once the lytic cycle has been activated by Tum. Two different 186 promoters were tested for protein expression; 186*pJ* (Figure 13. A) and 186*pR* (Figure 13. B), both promoters are controlled by the expression of lytic genes. Difference between 186*pJ* and 186*pR* in activation mechanism and timing may be important parameters in the context of the PHEARLESS system. The 186 CI protein is responsible for maintaining the lysogenic life cycle and repressing the lytic life cycle. When Tum prevents CI from repressing the 186*pR* promoter, lytic genes are turned, including *B*, the protein that activate expression of the 186*pJ* promoter. 186*pR* is strongly repressed by CI, when it is de-repressed it activates expression of the early lytic genes (163). In contrast, the 186*pJ* promoter activated later in the lytic cycle (167). These two promoters were chosen so that the antimicrobial protein's expression would be activated when 186 begins the lytic cycle but with different kinetics. Both 186*pR* and 186*pJ* are expected to tightly control expression, reducing the amount of leaky expression which would create issues when testing protein that are found to be toxic when over expressed. While the proteins that will be tested in this system will target *S. aureus* and not *E. coli* we will be testing high copy numbers of the expression gene. These high levels might still affect the expression strain if there is a significant amounts of leaky protein expression. Because the 186*pJ* promoter is activated later in the lytic cycle, this might be expected to reduce the amount of protein made prior to cell lysis.

In addition to two different promoters, two copy number variants of the expression plasmid were tested, pZS45 and pZS(^)45. The pZS45 plasmid contains the low copy

version of the pSC101 ori which maintains ~3 copies/cells. To increase protein production, a mutated high copy *pSC101 ori* (pSC101^) was made (152) which increases the copy number to ~70 copies/cell. The last variable that was tested was the two versions of the *tum*-based lysis module*, tum* wild type and *tum72*. Although Tum*72* was showed to be more efficient at cell lysis, both were tested again in the context of the plasmid expression system.

Another component required for the design and testing of the PHEARLESS system was an effective positive control protein with well-defined antimicrobial activity. The positive control protein chosen was a chimeric endolysin called ClyF, created by the Key Laboratory of Special Pathogens and Biosafety at the Wuhan Institute of Virology (130). ClyF was chosen due to its activity against a broad range of target strains, including thirty-one strains of *S. aureus*, three strains of *S. saprophyticus*, two strains of *S. equorum*, *S. sciuri*, three strains of *S. chromogenes*, two strains of *S. haemolyticus, S. epidermidis, S. capitis*, and *S. albus*. ClyF is a chimera consisting of an N terminal CHAP domain from the *Staphylococcus* virus 187 (Ply187) lysin and a C terminal CBD from PlySs2 lysin. PlySs2, also referred to as CF-301, was derived from a *S. sciuri* phage which shows high lytic active against both *Streptococci* and *Staphylococci* species (138). ClyF will retain activity for one month at room temperature and four months when stored at 4°C. It has also shown biofilm clearing activity, which makes it a potential positive control for future assays investigating biofilm activity.

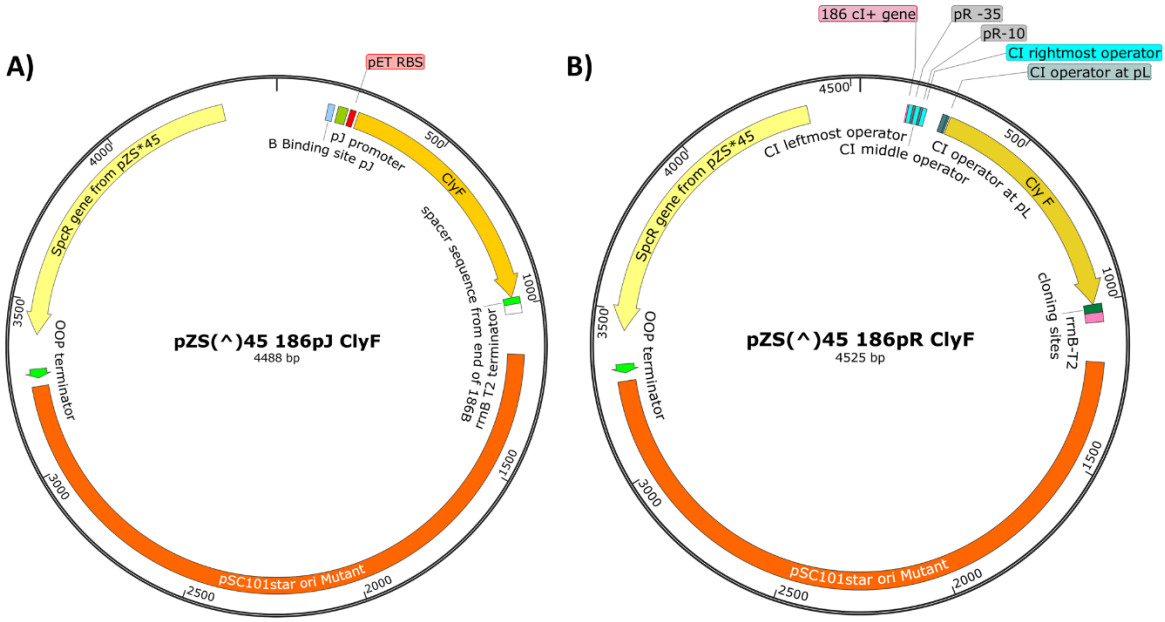


Figure 13: Plasmid Maps of Protein Expression Plasmids. *A) and B) shows the maps of the pZS(^)45 186pJ ClyF and the pZS(^)45 186pR ClyF protein expression plasmids, respectively. Both plasmids contain a Spectinomycin antibiotic resistance gene, pSC101star origin of replication (Mutant), rrnB T2 terminator sequence and the Protein of Interest Gene (here the positive control chimeric endolysin ClyF). A) The pZS(^)45 186pJ ClyF plasmid contains the 186pJ promoter 186pJ, which is activated by Protein B during the late lytic cycle. B) The pZS(^)45 186pR ClyF plasmid contains the 186 promoter 186pR, which is repressed by the CI protein. A third expression plasmid has been created but not shown here, pZS45 186pJ ClyF, this plasmid is identical to the pZS(^)45 186pJ ClyF plasmid with the exception of a single nucleotide change in the pSC101star origin of replication, N99D, which converts copy number from ~3 plasmids/cell to ~70 plasmids/cell. Figure was generated using SnapGene DNA maps.*

Thus, the design of the system was initially set up to allow testing of protein expression plasmids carrying promoters 186*pJ* or 186*pR* at both low and high plasmid copy number, and in combination with both the Tum wild type and Tum72 lysis switch modules. However, generating a high copy number plasmid carrying the 186*pR* promoter was found to be very difficult. Multiple transformations of high copy 186*pR* assembly mix would only produce a few clones per plate (~3-5), with the majority of these clones containing mutation in the 186*pR* promoter sequence and/or in the ClyF gene. When one apparently correct clone of the high copy 186*pR* ClyF plasmid was identified and transformed into the final expression strain, grew very poorly. It was possible that the additional cellular copies of the 186*pR* promoter might be causing induction of the 186Δ*tum* prophage by sequestration of CI. Thus the

186*pR* high copy plasmid was determined to be too problematic and was not pursued.

3.c Testing PHEARLESS Version 1, Using ClyF

The first assays were performed to compare the two different promoters on the endolysin expression plasmid, 186*pJ* (HB89) and 186*pR* (HB87), at low copy number. For these initial tests, the Tum wild type inducible lysis switch variants were used (Figure 14). The negative control strain contains an empty pZS45 expression plasmid, rather than ClyF. In brief, these experiments were performed by spotting aliquots (10 μL of undiluted or a series of 10 fold dilutions) of the expression/lysis strain, onto lawns of the target strain (here *S. aureus* RN4220). Cumate, at final concentrations of 40 μM and 120 μM, was incorporated into the plates, thus initiation of the lysis and expression circuit occurs only when the *E. coli* strain is spotted on the target strain. See section 2.h for full details of the protocol. Both the 186*pJ* and 186*pR* promoters produced RN4220 lysis in the undiluted sample (U) and first (1/10) dilution sample (Figure 14) (1). The negative control undiluted sample does show a RN4220 lawn disruption effect, possibly due to the large number of cells added. This physical effect might be caused by competition occurring between the *S. aureus* and *E. coli* strains when co-cultured, affecting the appearance of the lawn. Given that the Tum wild type strain has been show to not efficiently lyse at 40 μM cumate (Figure 12), competition between strains is a more likely explanation than an effect of the lysate. This effect doesn't occur when spotting the diluted *E. coli* and disappears on the 120 μM cumate plate (Figure 14) though varied somewhat from day to day. Thus a positive effect of the endolysin should be judged only with additional spots made with diluted *E. coli*.

While it was expected that a higher cumate concentration would increase ClyF expression, therefore increasing target strain RN4220 killing, any improvement was minor (Figure 14). At 40 μM cumate, the 186*pJ*-ClyF expression strain (Figure 14, 40 μM cumate) produced clearer killing up to the second dilution (2). 186pR only produced killing in the undiluted sample and light thinning in the first dilution samples. However, in the first (1) dilution and strongly in the second (2) and third (3) dilutions, cell growth that is not physically consistent with the RN4220 lawn is observed. It appears that while there may have been some expression cell lysis, strongly observed in the undiluted (U) sample, there is still a large amount of expression cells

that are surviving and continue to grow. Possibly enough ClyF is being released to inhibit RN4220 growth but the remaining expression cells overgrow in that location. This appearance is not present on the 120 µM cumate plates. Using the higher cumate concentration results in more efficient expression cell lysis of the Tum expression strain carrying the 186*pR* promoter (Figure 14, 120 µM cumate). At 120 µM cumate killing is observed again in the undiluted sample, with clear spots observed in the first (1) and second (2) dilution.

The 186*pJ* variant of the promoter shows consistent killing in the undiluted (U), first (1) and second (2) dilution samples at both 40 µM cumate and 120 µM cumate (Figure 14, 40 µM cumate, 120 µM cumate), with no appearance of the expression cell growth as seen in the 186*pR* dilutions on the 40 µM cumate plate. The problem of expression cell survival can largely be removed by increasing the cumate concentration to 120 µM, which shows more consistent cell lysis here and in the earlier liquid culture lysis assays (Figure 14, 120 µM Cumate, Figure 12).

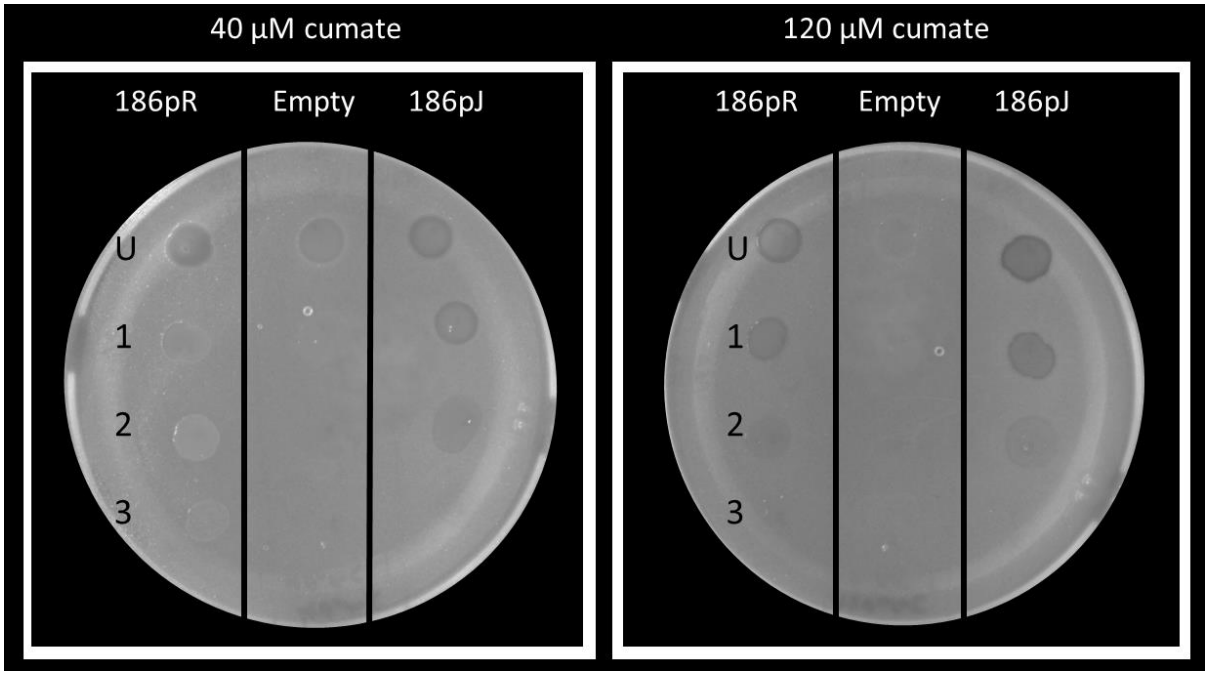


Figure 14: PHEARLESS Protein Screening Assay: Comparing Promoters in the Tum Wild Type Expression Strain on Low-Copy Number Expression Plasmids. *Comparison of low copy pZS45 plasmids for the 186pJ promoter variant(HB89) (186pJ) and the 186pR promoter variant (HB87) (186pR) in the tum wild type lysis strain. Both strains are compared against an empty expression plasmid strain (HB91) (Empty) that does not contain a promoter or the clyF gene. The empty control strain produces a lawn disruption effect in the undiluted sample on the 40 µM cumate lawn. This effect is not seen in any of the dilutions or in any of the sample on the 120 µM cumate plate. The lawn consists of the target strain, RN4220, a nonclinical S. aureus strain. Various dilutions of the*

PHEARLESS strain (undiluted (U), first (1/10) dilution (1), second (1/100) dilution (2) and third (1/1000) dilution (3)) for each expression strain were spotted onto the plates. Two plates were prepared one containing 40 µM of cumate and one containing 120 µM cumate. Photos were taken using a ChemiDoc XRS+, using the white light greyscale image settings. Three biological replicates of the assay using these conditions were performed.

The next strains tested were the low copy number plasmids with either 186*pJ* (HB89) and 186*pR* (HB87) promoters, but in the Tum72 cumate inducible lysis strains (Figure 15). Again 40 µM or 120 µM cumate was present in the plates, and the target strain was RN4220. The empty control strain (HB91) produced no lawn disruption in the undiluted sample on either of the plates, (Figure 15). Consistent with the liquid assays, the majority of the Tum72 expression strain undergoes cell lysis at both 40 µM and 120 µM cumate, with only a few surviving cells forming small colonies within the spots.

The low copy 186*pR* produced the same level of killing on both the 40 µM cumate and 120 µM cumate plates (Figure 15), clearing appearing in the undiluted (U) and first dilution (1) samples. While the 186*pJ* promoter expression strain only produces weak killing in the undiluted (U) sample on the 40 µM cumate plate (Figure 15, 40 µM cumate) it produced no killing in the 120 µM cumate plate, looking identical to the empty control strain (Figure 15, 120 µM cumate). When testing an unknown protein this would not be sufficient for confirmation of activity.

Therefore, these results could be interpreted as the Tum72 based expression strain producing less ClyF than the Tum wild type based strain. The greater stability of Tum72 could result in faster expression of 186 genes and virus particle assembly before host cell lysis, resulting in reduced build-up of ClyF. Although at this point we are not testing ClyF expression only target cell death. But by observing target cell death, speculation about the amount of ClyF of one strain compared to another could be speculated. Observing the expression strain cell lysis OD_{600} results (Figure 11, Figure 12), Tum72 strain starts to lyse at or just after the 45 minute mark, reaches the lowest OD_{600} reading between 60 and 70 minutes (Figure 11). The Tum wild type also appears to start cell lysis at the 45 minute mark, although the decline in OD_{600} is reduced the strain appears to reach lowest OD_{600} between the 75 and 80 minutes (Figure 12). While more extensive OD_{600} reading should be performed to confirm, this initial data could indicate that Tum72 faster build up in the cell results in faster progression towards host cell lysis by 5 to 10 minutes, by increasing the rate of

completely inhibiting CI and resulting in completed activity of the lytic cycle by being a more stable protein and degrading at a slower rate compared to CI. More quantitate experiments would need to be performed to confirm this increase in but currently this is outside the scope of the thesis. But from these initial results using the protein expression plasmid, the Tum72, while more efficient at activating cell lysis, in the expression is not optimal for endolysin expression using either 186*pJ* or 186*pR* promoter. Tum wild type was able to produce plaque formation up until the second (2) dilution sample for both 186*pR* and 186*pJ* at 120 µM cumate (Figure 14, 120 µM of cumate). Tum72 expression strain was unable to produces clear spots up until the second (2) dilution sample for either promoter at 40 µM or 120 µM cumate (Figure 15). 186pJ did not produce any defined spots with clear indication of activity (Figure 15, 40 µM cumate).

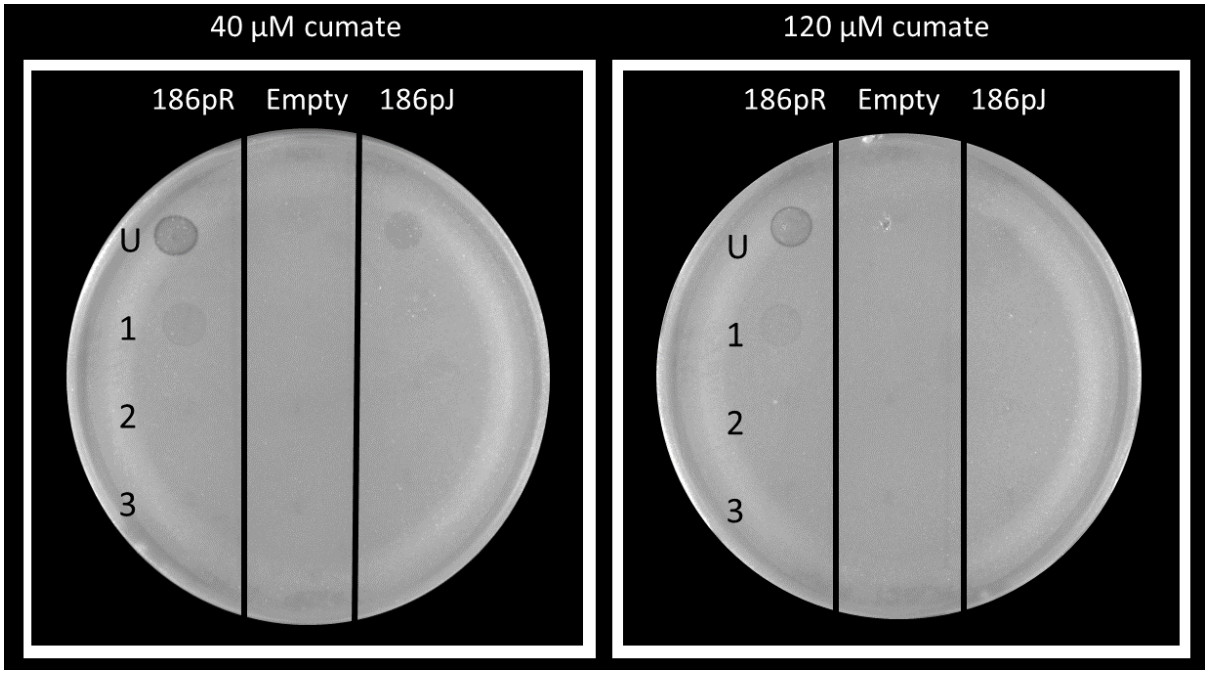


Figure 15: PHEARLESS Protein Screening Assay: Comparing Promoter Types in the Tum72 Expression Strain with Low-Copy Expression Plasmids. *Comparison of low copy plasmids for the 186pJ promoter variant (HB86) (186pJ) and 186pR promoter variant (HB84) (pR) in the tum72 expression strain in the PHEARLESS V1 system using PHEARLESS Protein Screening Protocols. Both strains are compared against an empty expression plasmid strain (HB78) (Empty) that does not contain a promoter or the ClyF gene. The lawn consists of the target strain, RN4220, a nonclinical S. aureus strain. Various dilutions undiluted (U), first (1/10) dilution (1), second (1/100) dilution (2) and third (1/1000) dilution (3) for each expression strain were spotted onto the plates. Two plates were prepared one contain 40 µM of cumate and one containing 120 µM of cumate. The Empty control strain showed no cell lawn disruption or target cell lysis on either the 40 µM cumate or 120 µM*

cumate plates. Regions of clearing are only present in the undiluted and 1/10 dilution for both 186pR promoter, on both the 40 µM cumate or 120 µM cumate plates. 186pJ promoter samples only show a slight lawn disruption in the undiluted sample which is not significant enough to show activity. Photos were taken using a ChemiDoc XRS+, using the white light grayscale image settings. Two biological replicates of the assay using these conditions were performed.

In an effort to improve the amount of ClyF being produced, a high copy number version of the 186*pJ*-ClyF expression plasmid was created, pZS(^)45 (Table 2). The high copy plasmid was compared against the low copy expression plasmids in both the Tum wild type (Figure 16) and Tum72 strains (Figure 17). Only the 186*pJ* plasmid version was tested due to the problems creating the high copy number 186*pR* plasmid stated above. The negative control plasmid used in this assay to compare the low and high was the high copy version of the empty expression plasmid, pZS(^)45 (Table 2).

Tum wild type (Figure 16) high copy (pZS(^)45) negative control (no ClyF) results (Empty) (HB112) show no RN4220 cell lysis, although there is a small lawn disruption effect in the undiluted (U) sample which is more visible on the 40 µM cumate plate (Figure 16, 40 µM cumate). Results for the low copy number 186*pJ* are similar to the previous results, in the Tum wild type expression strain (Figure 14). But with less expression strain growth on the 40 µM cumate plate (Figure 14, 40 µM cumate), which is now showing that there is killing up to the second dilution (2) sample (Figure 16, 40 µM cumate). The 120 µM cumate plate shows that low copy 186*pJ* also produced killing up to the second dilution (2) sample (Figure 16, 120 µM cumate).

The high copy number expression plasmid, which increases the copy number about 25 fold from ~3/plasmid per cell to ~70/plasmid per cell showed a large improvement in killing, increasing target cell killing to the point where the area is fully cleared (Figure 16). The high copy number 186*pJ* Tum wild type expression strain produces some killing up until the third dilution (1/1000) (3) on both the 40 µM and 120 µM cumate plates (Figure 16). Both the 40 and 120 µM cumate plate dilution samples also contain small numbers of surviving single colonies. These were re-streaked to confirm that they were the expression strain based on having the same antibiotic selection. In summary, comparison of the high copy number 186*pJ* to the low copy number 186*pJ* shows an obvious increase in target strain clearing by the high copy 186*pJ* suggests a significant increase in ClyF expression.

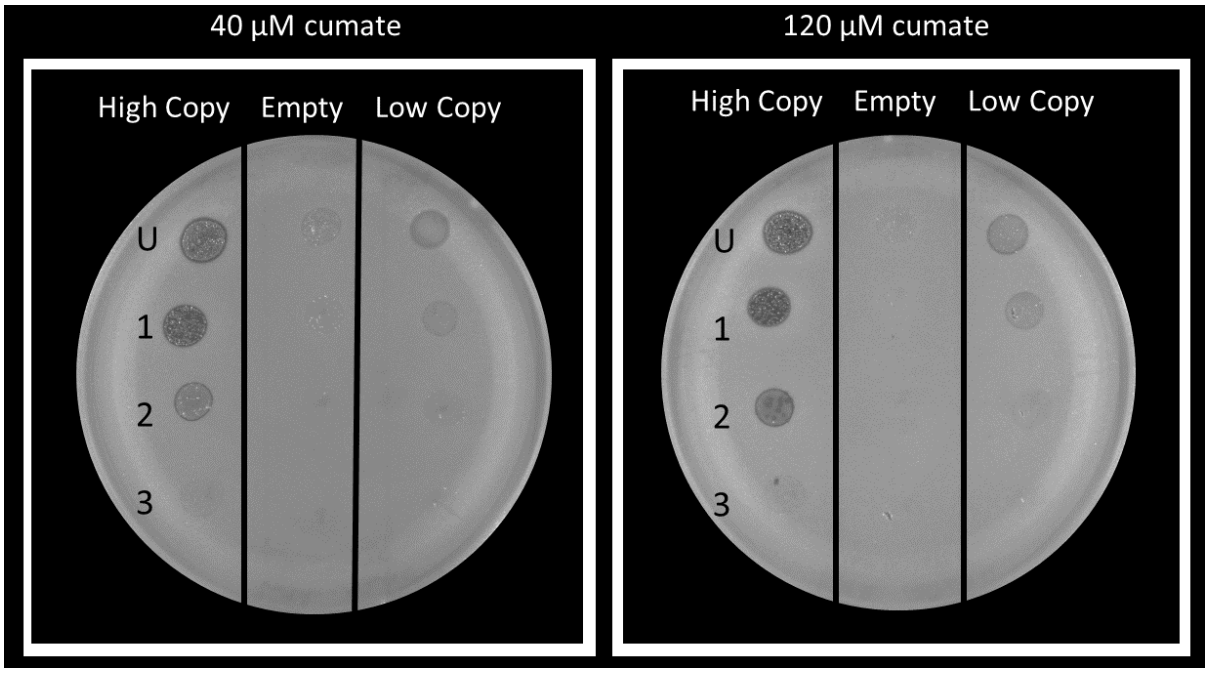


Figure 16: PHEARLESS Protein Screening Assay: Comparing Low Copy Number and High Copy Number 186pJ Expression Plasmids in the Tum wild type lysis strain*. Comparison of low copy number plasmid for the 186pJ promoter variant(HB89) (Low Copy) against the high copy number plasmid for 186pJ promoter variant (HB114) (High Copy) in the Tum wild type expression strain. Both strains are compared against a high copy number empty expression plasmid strain (HB112) (Empty) that does not contain a promoter or the ClyF gene. The lawn consists of the target strain, RN4220, a nonclinical S. aureus strain. Various dilutions undiluted (U), 1/10 dilution (1), 1/100 dilution (2) and 1/1000 dilution (3). Two plates were prepared one contain 40 µM of cumate and one containing 120 µM of cumate. Photos were taken using a ChemiDoc XRS+, using the white light grayscale image settings. Two biological replicates of the assay using these conditions were performed producing identical results.*

The final assay performed was for the Tum72 186*pJ* low copy number plasmid vs the high copy number plasmid (Figure 17). The empty control strain carries the high copy number version of the empty expression plasmid, as expected it shows no killing of the target strain and no lawn disruption effect was observed (Figure 17). The few single colonies observed within the spots for the empty plasmid are most likely the expression strain that did not lyse, similar to that seen with the low copy number plasmid. The low copy number plasmid results show identical results for the previous assay that compared the 186*pJ* promoter against the 186*pR* promoter in Tum72 (Figure 15). With the only plaque present appearing in the undiluted (U) sample on the 40 µM cumate plate (40 µM cumate, Figure 15, Figure 17). As previously stated, target cell lysis must be present in at least the first dilution for positive evidence of

antimicrobial activity. No plaques were observed for the low copy plasmid on the 120 µM cumate plates (Figure 17, 120 µM cumate), only small colonies of the expression strain.

Results for the high copy number 186*pJ* expression plasmid in the Tum72 strain (HB117) (Figure 17) showed improvement of target cell lysis compared to the low copy number plasmid (HB86) (Figure 15), but not to the same extent as the Tum wild type strain. Clearing was seen for the high copy number 186*pJ*-ClyF in the Tum72 lysis strain in the undiluted (U) sample and first dilution (1) on the 40 µM cumate plate (Figure 17, 40 µM cumate). There is also now a very clear spot in the undiluted (U) sample on the 120 µM cumate plate (Figure 17, 120 µM cumate), but nothing with the more diluted samples. With Tum72, 40 µM cumate shows better killing than 120 µM cumate, which is the reverse of the Tum wild type expression strain.

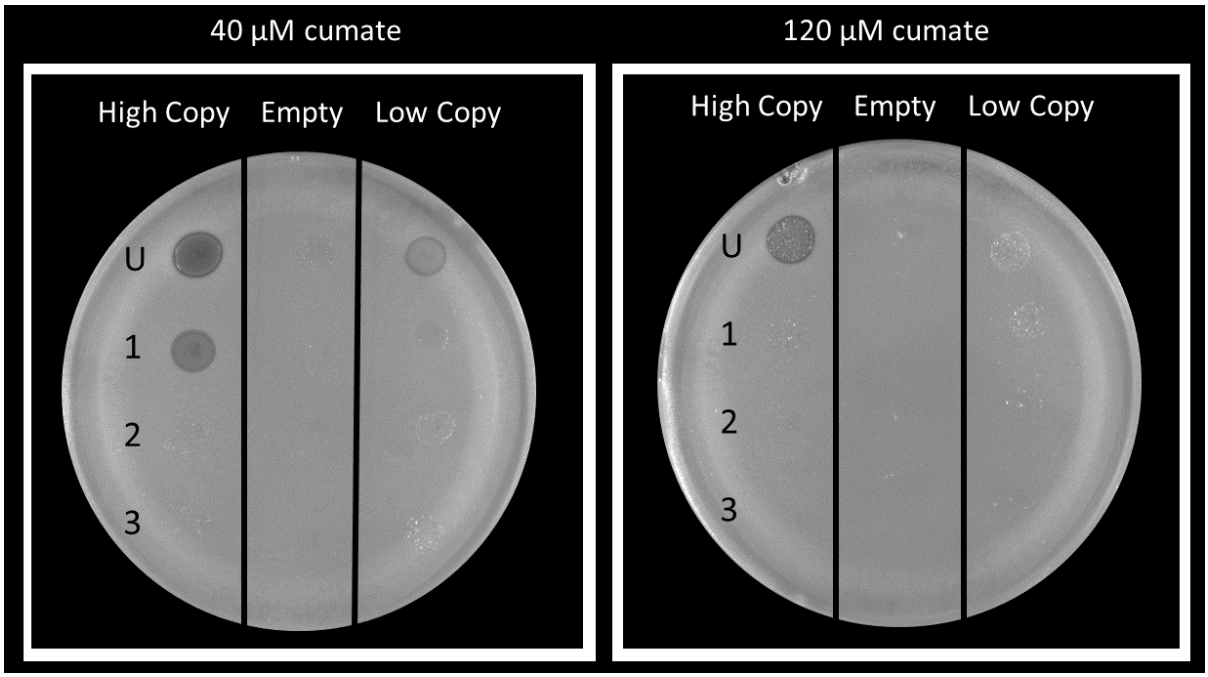


Figure 17: PHEARLESS Protein Screening Assay: Comparing the Low Copy and High Copy Number 186pJ Expression Plasmids in the Tum72 lysis strain. *Comparison of low copy number plasmid for the 186pJ promoter variant (HB86) (Low Copy) against the high copy number plasmid for 186pJ promoter variant (HB117) (High Copy) in the tum72 expression strain. Both strains are compared against a high copy number empty expression plasmid strain (HB115) (Empty) that does not contain a promoter or the ClyF gene. The lawn consists of the target strain, RN4220, a nonclinical S. aureus strain. Various dilutions undiluted (U), 1/10 dilution (1), 1/100 dilution (2) and 1/1000 dilution (3) for each expression strain were spotted onto the plates. Plates contained either 40 µM cumate or 120 µM of cumate.*

Photos were taken using a ChemiDoc XRS+, using the white light grayscale image settings. Two biological replicates of the assay using these conditions were performed, producing identical results.

Final results for testing each of the variants (Table 4) showed that for ClyF mediated killing of RN4220, the Tum wild type lysis system in combination with the high copy number 186*pJ*-ClyF plasmid was the most effective. Although Tum72 was determined to be the most efficient at activating cell lysis, (Figure 11) for the purposes for endolysin expression, this lysis seems to be too rapid, with insufficient time for protein expression prior to onset of cell lysis.

The two promoter variants produced mixed results, with 186*pR* showing more killing in both Tum wild type and Tum72 strains with the low copy plasmid. (Figure 14, Figure 15), however a stable high copy number variant of the 186*pR*-ClyF plasmid could not be generated. Low copy 186*pJ* was only capable of producing faint clearing at either cumate concentration (Figure 14, Figure 16).

The high copy number version of 186*pJ* expression plasmid however showed a clear improvement in killing of RN4220 in both Tum strains (Figure 16, Figure 17). The Tum72 strains showed improved activity with clearing present at 1/10 dilution on the 40 µM cumate plate (Figure 17, 40 µM cumate). The 120 µM cumate plate only shows a clear plaque in the undiluted sample (Figure 17, 120 µM cumate), again suggesting lysis may be happening too quickly. The Tum wild type strain containing the high copy 186*pJ* plasmid showed the best results for target cell lysis (Figure 16) with 120 µM cumate producing the best result for the Tum wild type expression strain.

In conclusion, the optimal expression strain to use for future experiments to test putative antimicrobial protein activity is the high copy number 186*pJ* expression plasmid in the Tum wild type expression strain, using 120 µM cumate (Figure 16, 120 µM cumate). It is important to note that the optimisation experiments done here used ClyF, a relatively active endolysin. Other putative antimicrobial proteins will have different levels of activity and/or solubility, likely requiring some adjustment of conditions. However, the basic system of simultaneous induction of protein expression and phage mediated lysis has been established, and simple adjustment of the cumate concentration would be the first parameter to vary in study of other antimicrobial proteins.

Table 4: Summary of the Different PHEARLESS V1 variants tested. *Table contains a summary of the results for the different tum variant, protein expression plasmid promoter and protein expression plasmid origin or replication.*

Expression Plasmid	**Tum Wild Type**	Results	**Tum72**	Results
Low Copy 186*pJ* promoter	✓	Clearing present in the undiluted (U), first dilution (1) and second dilution (2) on the 40 µM and 120 µM cumate plate.	✓	• Only slight lawn disruption is present on the undiluted sample (U) on the 40 µM cumate plate. Cannot confirm a positive or false positive result
Low Copy 186*pR* promoter	✓	• Clearing present in undiluted (U) and first dilution (1) on the 40 µM and 120 µM cumate plate.	✓	• Clearing present in the undiluted (U) on the 40 µM and 120 µM cumate plates. • Faint Clearing present in the first dilution on the 40 µM and 120uM cumate plates.
High Copy 186*pJ* promoter	✓	• Clearing present in all the samples except for the except for the last dilution (3). • Showed identical results for both cumate concentration plates .	✓	• Clearing present for both undiluted samples of the 40 µM cumate and 120 µM cumate plates. • 40 µM cumate plate also contains RN4220 cell lysis in the first dilution (1).
High Copy 186*pR* promoter	**X**	• Variant could not be tested due to issue constructing High Copy 186*pR* plasmid and instability of high copy 186*pR* in the *tum wild type* expression strain.	**X**	• Variant could not be tested due to issue constructing High Copy 186*pR* plasmid and instability of high copy 186*pR* in the *tum72* expression strain.

3.d Engineered Bacteriophage 186 Functional test

The PHEARLESS system consists of two modules – the engineered bacteriophage 186 module used to bring about cell lysis, and the expression plasmid, where the promoter controlling expression of the protein of interest, is responsive to the phage late activator protein, thus coordinating the function of the two modules. The first version of the PHEARLESS system was designed such that the bacteriophage 186 module had been engineered to remove the Tum gene and non-essential *orf97* gene,

and placing the tum gene as a separately controlled unit in the chromosome (Figure 8, Figure 9). One question in the design was whether to allow 186 after initiation of lysis, to go on to produce functional phage particles. This can be easily controlled by modification of the cos sequence, which lies adjacent to the *orf97* gene, and is essential for packaging of the phage DNA into the capsid (167). Without the cos sequence, only empty heads, and no infectious phage particles are produced.

There are arguments both for and against production of active phage particles in the PHEARLESS system. An argument against production of active phage is that without an active *cos* sequence, and hence no phage DNA packaging, lysis timing is likely to be delayed, thus allowing more time for expression of the protein of interest. Without phage production the system is a simpler, single step lysis event. On the other hand, active phage production allows the additional possibilities for more sophisticated designs, including using the phage to 'preserve and amplify' a copy of the gene of interest – a design which is explored in later chapters.

The first version of PHEARLESS used a design aimed at preventing phage packaging. Here we verify that this modification was successful, and also show that the phage module of PHEARLESS can also be modified in order to reinstate efficient phage packaging. The phage propagation strain used as a lawn for these tests was HB96, an E4643 *E. coli* bacterial strain (HB96) carrying the cumate Tum72 switch (E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$). This strain, which does not carry any plasmids or prophage, allows any *tum* minus phage variants to propagate and give rise to plaques by complimenting the missing *tum* genes, but only in the event that phage DNA packaging is functional. Phage stocks were prepared from either un-induced or induced liquid cultures of the PHEARLESS strains, (without any endolysin expression plasmid), and 10 µL of these stocks spotted onto a lawn of the propagation strain on a plate containing 80 µM cumate. Wild type 186 phage was used as a positive control for phage production, and as shown in Figure 18 (top left quadrant), gave an obvious spot/plaque.

As a negative control, lysate from the Empty *tum* expression strain (i.e. no Tum from either the module or the prophage), did not to produce any active phage particles from either un-induced (U) or induced (I) (Figure 18). Moreover, neither the Tum72 or the Tum wild type induced samples were capable of producing functional phage

particles (Figure 18) in either un-induced (U) or induced (I) samples, confirming that the cos sequence truncation completely prevents production of functional phage.

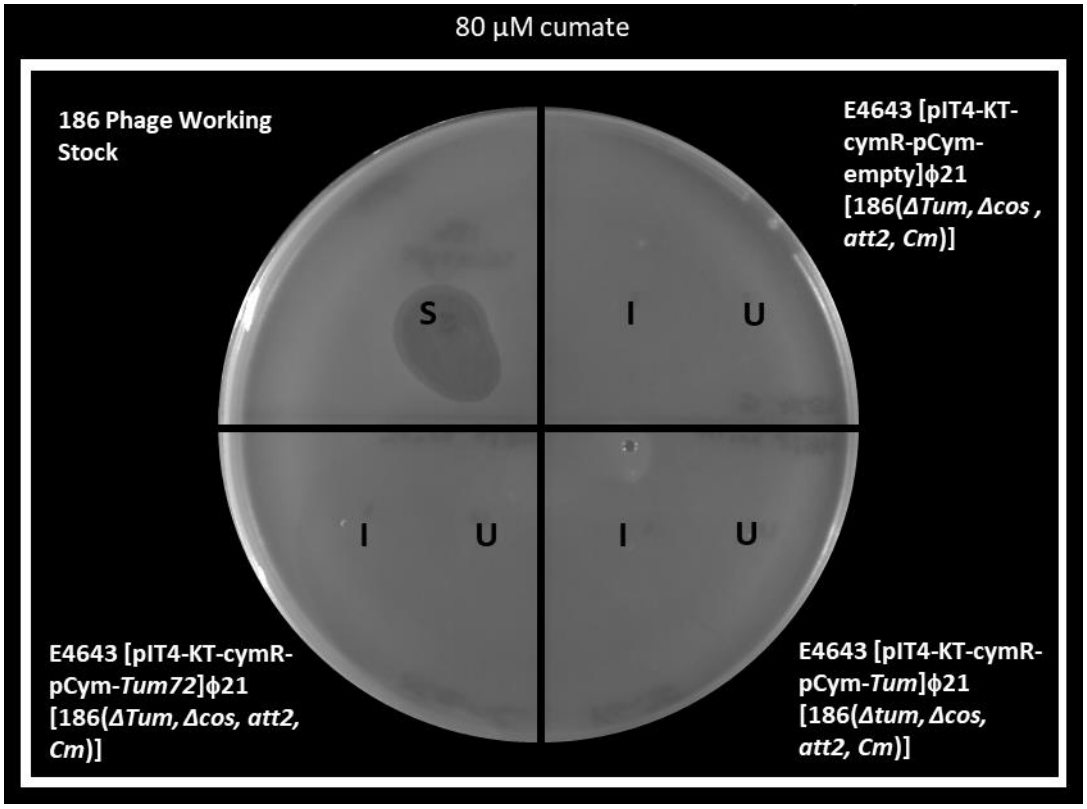


Figure 18: Testing for Production of Functional Phage Particles. *Phage stocks were prepared from the parental strains (carrying no protein expression plasmid) of the PHEARLESS strains. Wild type 186 phage stocks were used as a positive control for plaque formation. Phage stock samples were plated on a 186 sensitive E. coli indicator strain (HB96). Induced (I) and Un-induced (U) phage stocks were prepared from the parent strains of the plasmid based strain tested for ClyF activity, by omission or addition of cumate to the cultures. Phage stocks from the empty tum expression strain (HB74) (E4643 [pIT4-KT-cymR-pCym-empty* $^{\phi 21}$ *[186(Δtum, Δcos, att2, Cm)]) were prepared as a negative control to confirm that with (induced) or without (un-induced) cumate, functional phage are not produced when Tum is absent. Both the Tum wild type (HB78) (E4643 [pIT4-KT-cymR-pCym-*Tum*]*$^{\phi 21}$ *[186(Δtum, Δcos, att2, Cm)]) and Tum72 (HB76) (E4643 [pIT4-KT-cymR-pCym-*Tum72*]*$^{\phi 21}$ *[186(Δtum, Δcos, att2, Cm)]) were also shown to be unable to produce functional phage (I). Two biological replicates of the assay using these conditions were performed producing identical results.*

Release of new phage particles being critical for future variants of the PHEARLESS system, a new G-block was designed to replace the *tum* and *cp97* genes, while leaving the cos sequence intact (Figure 9, C). To recover phage production a new parental expression strain was created, E4643 [186*(Δtum, cos+, att2, Cm)]* (HB93), using recombineering to remove the 186 *tum* gene as well as gene *orf97*. The

inserted gene fragment still contained the two *attB* sites (*attB3*, *attB2*) and the Cm resistance gene flanked by *LoxP* sites (Figure 9, C). After replacing the *tum* and *orf97* genes with the new recombinant fragment, the cumate mediated Tum-based switch modules were integrated as per the ∆*cos* previous strain. The phage functional test was then repeated to confirm that the new strain was capable of producing functioning phage particles. The results (Figure 19) showed that the new strains containing the 186 *cos* sequence were capable of producing functioning phage particles in their induced samples (I). The two control strains produced the expected results, the 186 wild type stock producing a large plaque. The empty cumate switch (HB103), which lacked a *tum* gene did not produce plaques in either the un-induced or induced samples. However, both the Tum wild type (HB104) and Tum72 (HB106) strains produced plaques in the cumate-induced samples (I). In the un-induced samples (U) no large spot was produced, though there were a small number of plaques t in the un-induced sample of *Tum* wild type. These likely occurred due to spontaneous induction aided by a leak of Tum expression, which also occurs in wild type phage (145). Thus, strains that originate from HB93 are capable of producing functioning phage particles, allowing the design of new variants of the PHEARLESS system, as described in Chapter 4.

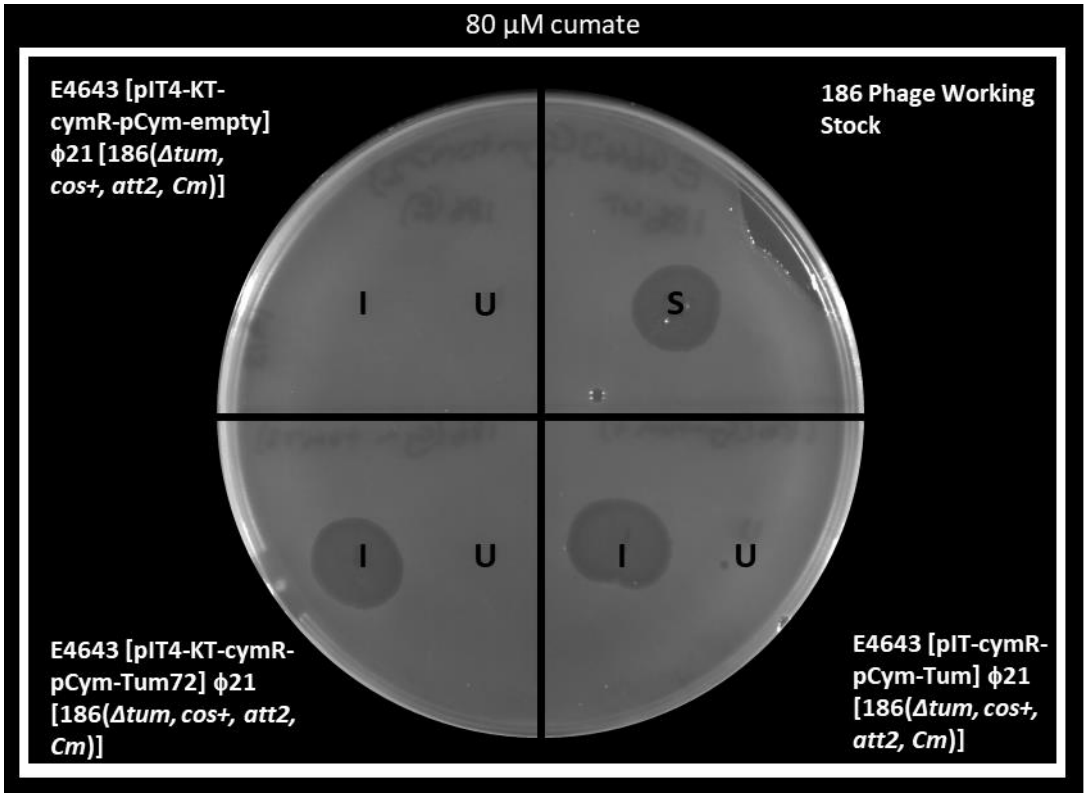


Figure 19: Testing for Production of Functional Phage Particles from Engineered 186 cos⁺ Prophage

. Phage stock samples were plated on a 186 sensitive E. coli lawn (HB96) which allows ΔTum phage

*variants to propagate.t. Induced (I) and un-induced (U) phage stocks were prepared from strains created from the new engineered 186 prophage genome that contained the full cos sequence (HB93). Wild type 186 phage stocks were used as a positive control. Phage stocks from the empty tum expression strain (HB103) (E4643 [pIT4-KT-cymR-pCym-empty]$^{\phi 21}$ [186(Δtum, cos+, att2, Cm)]) were prepared as a negative control to confirm that with (induced) or without (un-induced) cumate, the phage are unable to induce if Tum is absent. Both the Tum wild type (HB104) (E4643 [pIT4-KT-cymR-pCym-*Tum*]$^{\phi 21}$ [186(Δtum, cos+, att2, Cm)]) and Tum72 (HB106) (E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δtum, cos+, att2, Cm)]) were shown to be capable of producing functional phage particles when Tum is provided by the addition of cumate (I) and the cos sequence is intact. Two biological replicates of the assay using these conditions were performed producing identical results.*

Summarising, in this chapter, the modules of the PHEARLESS system were tested and optimised for cell lysis and protein expression. The cumate-controlled 186*tum* induction module demonstrated control over the engineered 186 prophage, causing expression cell lysis. The activation of the lytic cycle by the phage late activator B also activated the protein of interest expression located on the expression plasmid. Many parameter variations and combinations of the modules were tested: cumate concentration, Tum variant, promoter strength and the protein expression plasmid copy number. Comparing all the results, the Tum wild type expression strain carrying a high copy version of the expression plasmid with the 186*pJ* promoter plated on 120 µM and 40 µM of cumate were selected as providing the optimal behaviour of the expression strain.

.

Chapter 4: PHEARLESS V2 Assay

4.a PHEARLESS version 2 Background

In Chapter 3: Design and construction of the PHEARLESS system, the individual components of the PHEARLESS assay system were tested and optimised for expression strain lysis and protein expression. Functional phage production was also investigated and it was confirmed that the Δ*cos* strain was unable to produce functioning phage particles (Figure 9, B). A second version of the expression strain was created that was capable of produce functioning phage particles (Figure 9, C). Based on these results, a second version of the PHEARLESS system (V2) was designed, eliminating the expression plasmid and moving the expression of the gene of interest to the 186 prophage. The prophage-based expression allows gene copy number to change during the lytic cycle as the phage genome replicates, from a single copy to potentially hundreds of copies later during the lytic cycle.

The key advance in the second design is that the gene for the protein of interest will be preserved within the phage genome when functional phage particles are generated. This allows simple isolation of phage from plaques and subsequent retrieval by PCR and sequencing of the gene, potentially extending the use of the system for testing large libraries of mutants. Again, the ClyF chimeric endolysin is used as a positive control for development of the system to target *S. aureus*. This chapter will explain the changes that were made to the engineered 186 prophage that was introduced in Chapter 3.

Building upon the Shearwin laboratory's previous work on chromosomal integration (158,168) the new design was based on insertion of a plasmid carrying the gene of interest into a 186 prophage genome that had been modified to include specific sequences (*att* sites for serine integrases) for site-specific integration. The plasmid carries corresponding *att* sites, and also has a conditional origin of replication (*pir*-dependent) and so does not independently replicate once integrated into the prophage. Finally, the plasmid carries a pair of *loxP* sites, allowing removal of non-essential plasmid elements if required, leaving just the endolysin expression module embedded in the 186 prophage. Mutant versions of the Cre-*loxP* system were used (169), where these mutant *loxP* sites (*loxLE* and *loxRE*) are recognized by the Cre-

recombinase, but which upon recombination produce a double mutant scar (*loxLE/RE*) which has low affinity for the Cre protein.

Most non-filamentous phage has evolved to package a defined amount of DNA into their capsids, and this genome size can generally be increased by only about 10% (see later Chapter 4.g.iii). Hence, in designing the new system, the average endolysin gene size (1-2 kb) was considered, and recombineering of the prophage was designed to remove a similar amount of 186 sequence. In this way, an approximately wild type-sized 186 prophage genome (30.6 kb) will be regenerated after incorporation of an endolysin gene.

Thus, the 186 *tum* and *orf97* genes were removed using recombineering and replaced with two phage attachment sites (*attB2*, *attB3*) (147), and a Cm resistance gene flanked by *loxP* sites (Figure 20, 2). The two *attB* attachment sites chosen for insertion of the corresponding *attP*-containing plasmid are derived from non-*E. coli* phage (147), this was done to prevent unintended phage integration at these attachment sites, or at other sites in the *E. coli* chromosome.

In PHEARLESS version 2, the cumate induction module was still used to activate expression cell lysis by activating expression of genes involved in the early stages of the lytic cycle (Figure 20, 4). These early stages genes are responsible for phage genome replication, repeatedly replicating the 186 genome. Concurrent with genome replication, the late lytic cycle proteins are expressed, which are responsible for structural protein expression, in addition to the endolysin of interest which has again been placed under control of the late lytic activator (Figure 21, 5). While initially there is a single 186 genome to translate from, after the early stages of 186 genome replication there are many copies. This takes the number of ClyF genes being expressed for one to about one hundred copies (170). Once the new phage particles have assembled and endolysin (ClyF in our test case) concentration has built up (Figure 21, 6), host cell lysis ensues (Figure 21, 7). Lysis releases the ClyF protein into the surrounding environment, killing susceptible target bacteria in the vicinity of the expression strain (Figure 21, 8). When co-plated with a non-lysogenic, 186 sensitive *E. coli* strain (hereafter termed the phage propagation strain), the active 186 phage released following induction by cumate can then infect the propagation strain and release new phage, amplifying the production of endolysin and giving rise to zones of clearing visible to the naked eye (Figure 22).

Overall, the design of PHEARLESS version 2 should provide an advantage over the plasmid-based expression version, in that the endolysin gene is present in the genome of the new phage particles being released. When the engineered 186 prophage is chemically induced, it will create a cycle of infection, lytic cycle activation and cell lysis, allowing for continuous expression and release of the endolysin. The phage particles carry the endolysin gene, making it simple to retrieve the gene from the phage using phage DNA purification or PCR, solving the issue of recovering the genotype of the desired variant, after the loss of the expression cell, which occurs upon release of the protein. This design leads to a new application of the screening assay, instead of being used to simply test for antimicrobial activity, a mutant library screen can be done, aiming to select for improved activity.

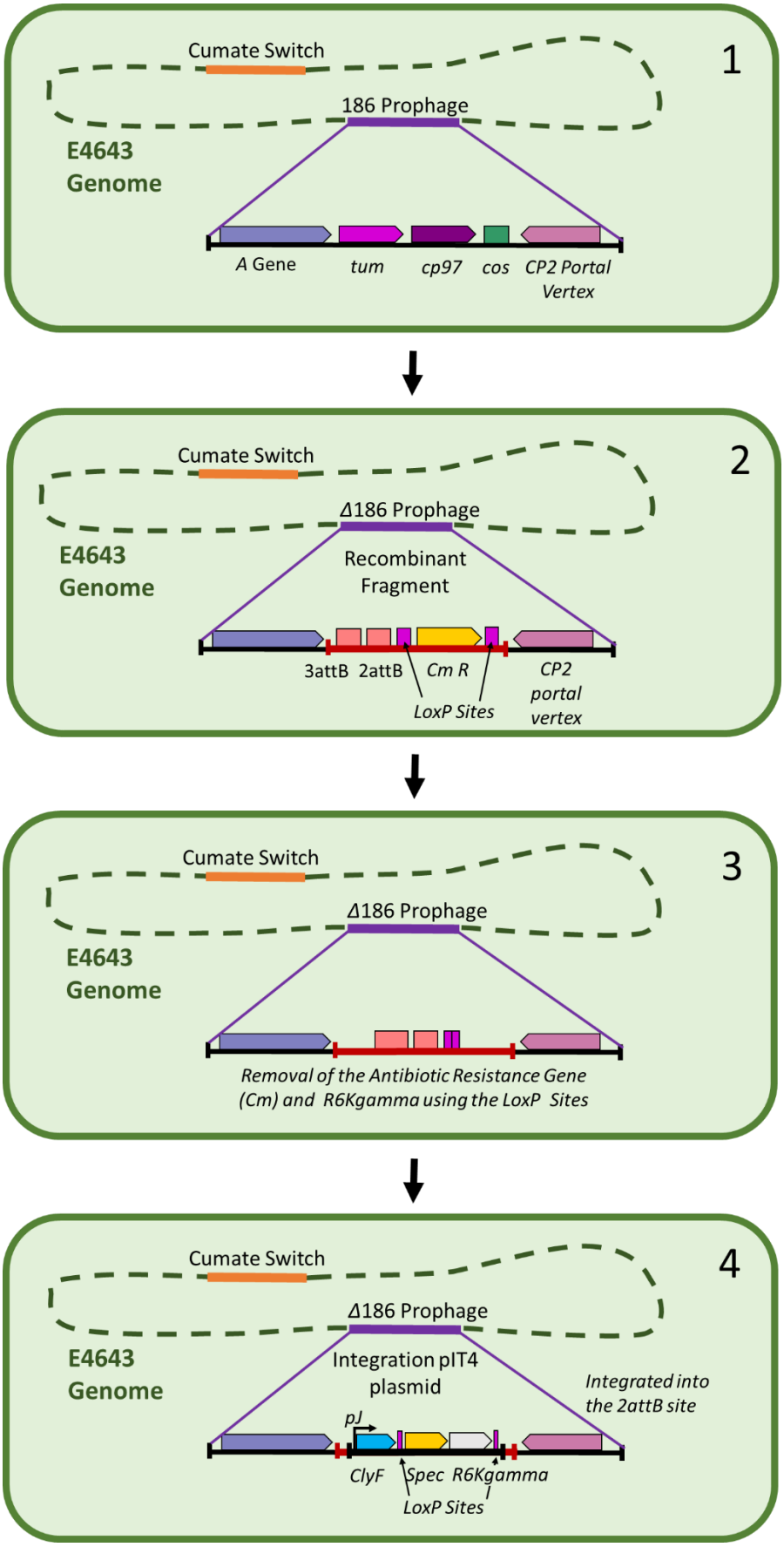


Figure 20: PHEARLESS version 2 design, part 1. *The design of the PHEARLESS V2 expression strain functions in the protein library screening experiments. Based on the cumate induced lysis strain (PHEARLESS version 1) generated in Chapter 3, E4643 [186(Δtum Δcos, att2, Cm)]. (1) shows the*

native components of the 186 prophage that have been replaced (2) with the two att sites and a Cm resistance gene (Cm R) flanked by LoxP sites are introduced by recombineering. The Cm gene allows for selection, and is then removed with a Cre recombinase (supplied from a pE-Cre plasmid, which carries a ts origin of replication) (3). A pIT4 integration plasmid containing a Spec antibiotic resistance gene (Spec) and the protein of interest (ClyF as the test example) is integrated in the att2 integration site (4). The resulting strain contains new LoxP sites flanking the Spec gene and the R6Kgamma conditional origin of replication.

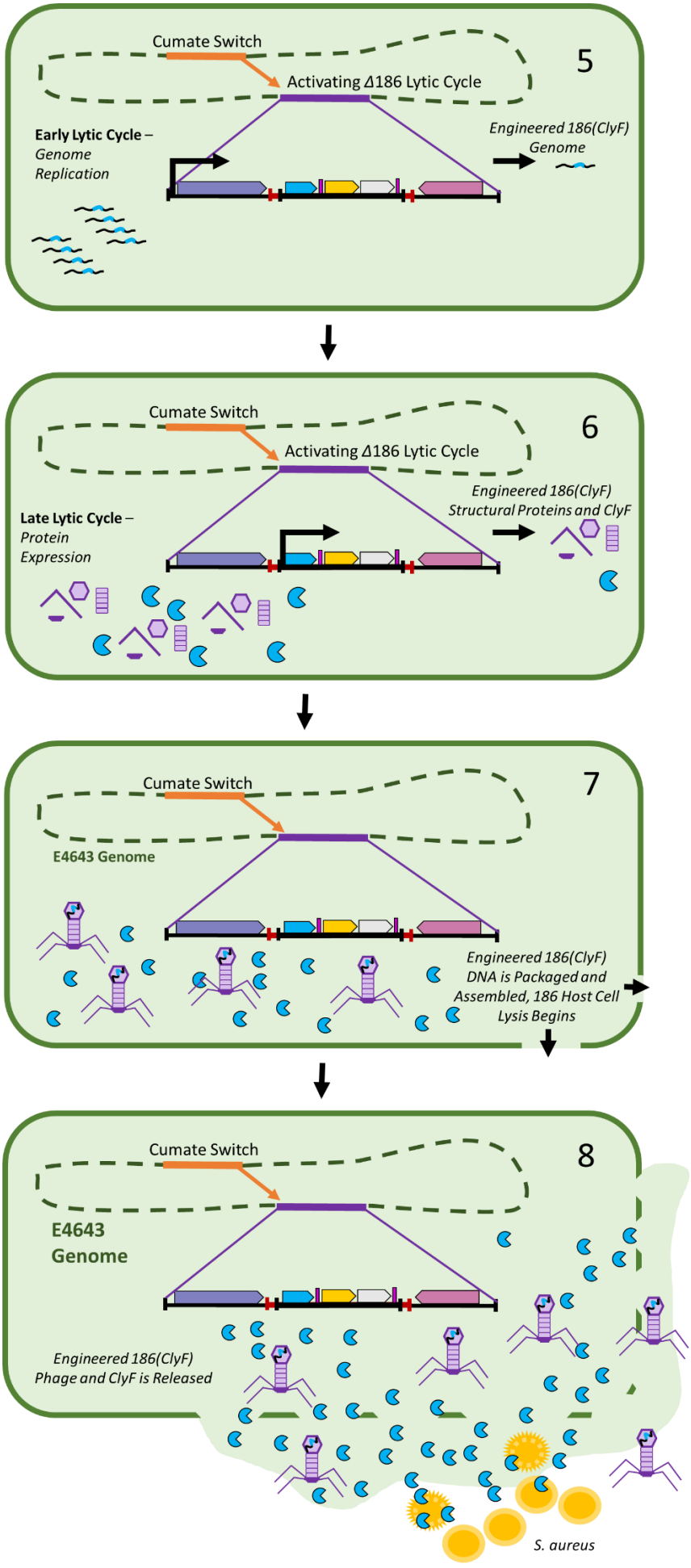

Cumate Switch
5
Activating Δ186 Lytic Cycle
Early Lytic Cycle –
Genome
Replication
Engineered 186(ClyF)
Genome
Cumate Switch
6
Activating Δ186 Lytic Cycle
Late Lytic Cycle –
Protein
Expression
Engineered 186(ClyF)
Structural Proteins and ClyF
Cumate Switch
7
E4643 Genome
Engineered 186(ClyF)
DNA is Packaged and
Assembled, 186 Host Cell
Lysis Begins
Cumate Switch
8
E4643
Genome
Engineered 186(ClyF)
Phage and ClyF is Released
S. aureus

Figure 21: PHEARLESS version 2 design part 2. *Following from Figure 20, when cumate is added to the system, activating the lytic cycle, early lytic genes are expressed, beginning to replicate the 186 genome (5). During the late lytic cycle phage protein production begins, including structural proteins as well as the protein of interest, which is controlled by a late lytic cycle promoter (186pJ) (6). Late in the lytic cycle, the phage genomes are packaged into the assembled phage head, and ClyF has also built up in the system (7). Functional phage particles are assembled after completing phage genome packaging, and the 186 begins host cell lysis (8). Upon release from the expression cell, ClyF will interact with the surrounding target strain (S. aureus in this example) resulting in lysis of the target strain by breaking down the PG layer.*

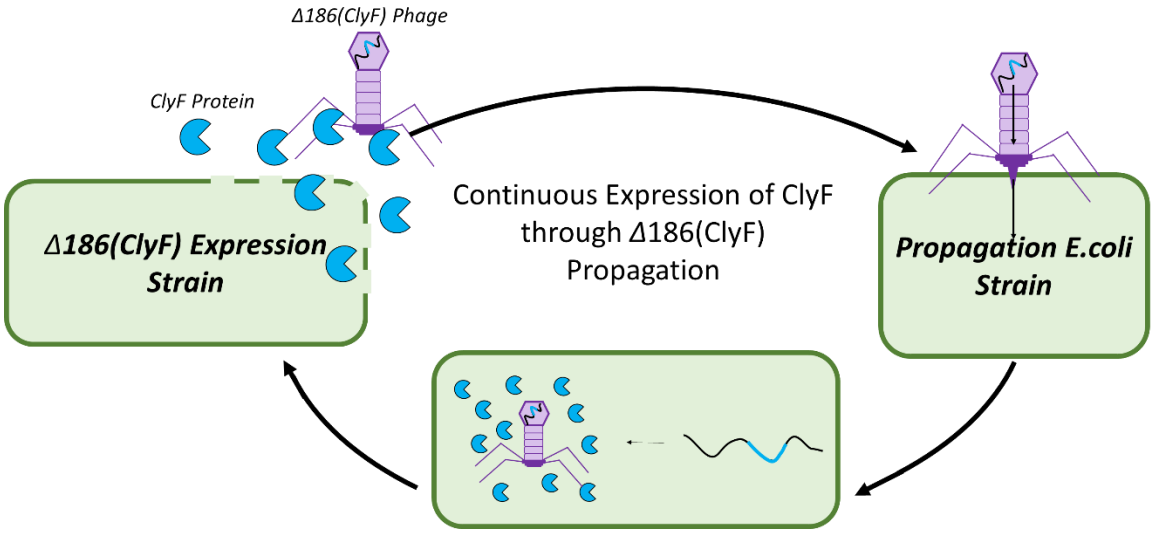


Figure 22: PHEARLESS version 2 design, continuous expression though phage propagation. *Demonstrates the advantage of the PHEARLESS 2 expression system. Because the 186 genome contains the gene for the protein of interest (ClyF), the progeny phage produced will also carry the gene, controlled by the late lytic activator. When one of these phage infects a new host (supplied as a prophage free propagation strain) activation of the lytic cycle results in production of a new round of phage particles and more protein of interest (ClyF). Overall, the gene of interest continues to be expressed via the propagation of the engineered 186(ClyF) phage. Phage samples can be used as a template for PCR amplification to retrieve the gene of interest (ClyF).*

4.a.i pIT4 Integration Plasmids

In PHEARLESS version 2, the gene of interest is placed into the 186 genome using a modified pIT4 integration plasmid, integrated at the *att2* attachment site. pIT4 plasmids carrying attachment site *att3* were generated as a back-up but were not required in this study. pIT4 plasmids (Figure 23) were designed to be small, only containing; the protein gene (ClyF), 186 promoter, R6Kgamma origin of replication

and Spectinomycin (Spec) resistance gene. A negative control pIT4 plasmid was also created with identical components including the 186*pJ* promoter but excluding the protein gene (ClyF) (Figure 23, C). The pIT4 integration plasmid contains a R6K gamma origin of replication, which requires the presence of Pir protein for initiation of replication (171). ClyF was used again as a positive control to confirm that a gene for an antimicrobial protein could be placed in the 186 genome, was well expressed and released upon cell lysis.

The first versions of the pIT4 plasmids were assembled using Gibson assembly. The promoters and ClyF gene were amplified from pZS (^)45. The backbone fragments were amplified from two pIT plasmids; AH1724 and AH1529 (Table 2), with the majority originating from the AH1724. After amplification, the backbone fragment products were re-amplified using a primer extended to include the phage attachment sites. pIT4 plasmids were generated containing either the 186*pJ* or 186*pR* promoter, along with the corresponding two empty pIT4 plasmid versions that did not contain the ClyF gene.

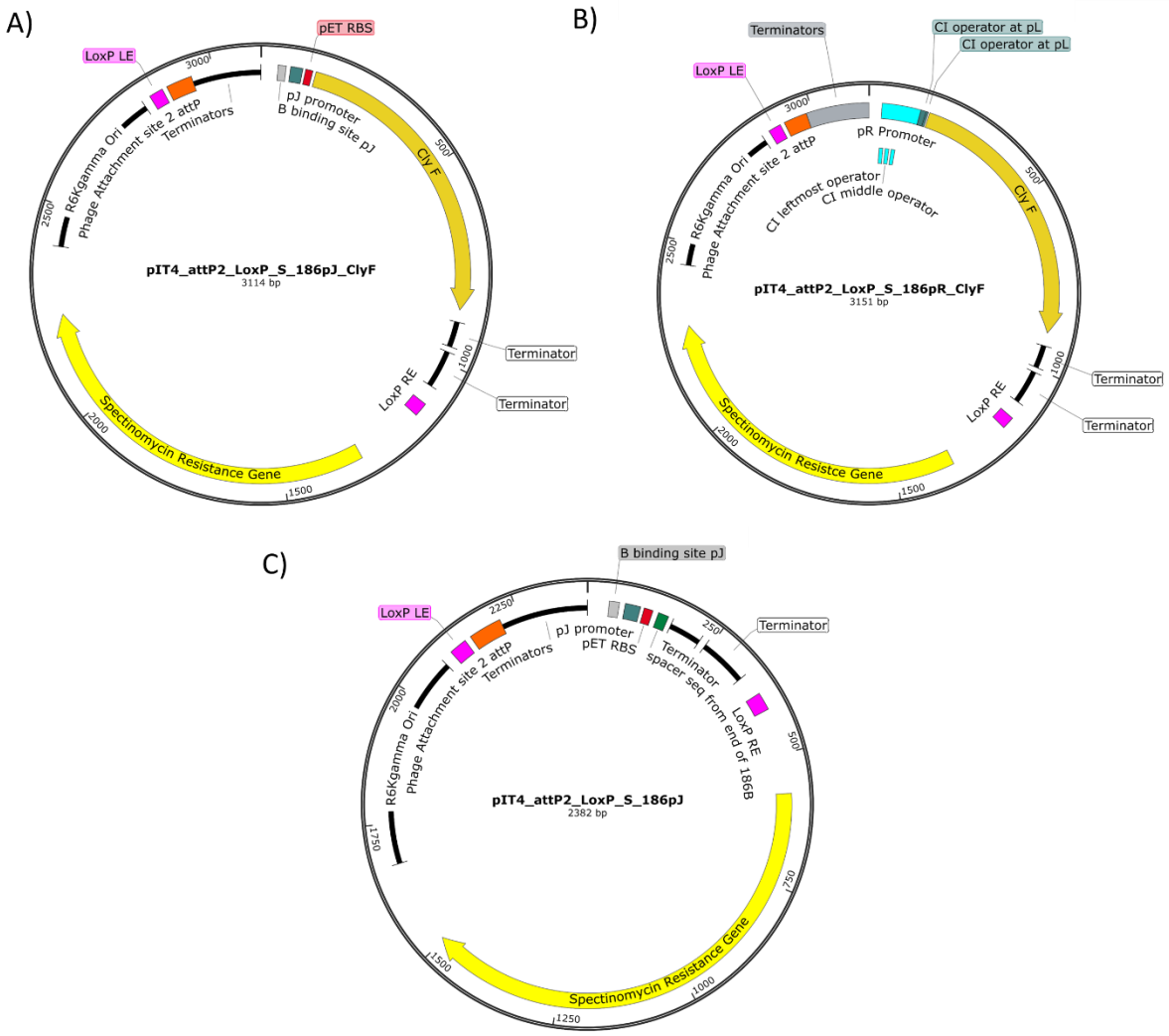


Figure 23: Plasmid Maps for the Three Variants of the Protein Integration plasmid (pIT4). *pIT4 version 1 plasmids A), B) and C) all contain a Spectinomycin Resistance Gene and R6Kgamma origin of replication flanked by LoxP sites, att2 site and terminator sequences surrounding the promoter and protein gene sequence (ClyF). A) pIT4_attP2_LoxP_S_186pJ_ClyF (HB137) contains the 186pJ promoter and the ClyF gene, B) pIT4_attP2_LoxP_S_pR_ClyF (HB132) contains the 186pR promoter and ClyF gene. C) pIT4_attP2_LoxP_S_186pJ (HB162) is used as an empty control, only contains the 186pJ promoter gene and no gene of interest (ClyF). Figure was generated using SnapGene DNA maps.*

4.b The Propagation Strain for Engineered 186 Propagation

For propagation of the engineered 186, a propagation strain is added to the plates, along with the endolysin-producing lysis strain and the target bacterial species. Three *E. coli* strains were investigated for use as the propagation strain. All three versions were based on E4643. Two strains contained one of the cumate switches; Tum wild type (HB97, E4643 [pIT4-KT-cymR-pCym-*Tum*] ϕ^{21}) or Tum72 (HB96, E4643 [pIT4-

KT-cymR-pCym-*Tum72*]$^{\phi 21}$), and the third did not carry a Tum switch, but carried two copies of fluorescent protein gene called Td-tomato (QC024) (172). The E4643 strains containing the cumate-controlled Tum switches were trialled initially, since it was unknown how efficiently the engineered 186 phage particles would lyse the propagation strain, due to lack of the *tum* gene on the phage. The QC024 strain was tested as an alternative propagation strain because both 186 phage attachment sites are already occupied - by a constitutively expressed Td-tomato gene (158). In a wild type 186 prophage, the prophage expresses the CI lysogenic repressor, preventing reinfection by a 186 phage (145,173,174). In the QC204 propagation strain, both attachment sites are unavailable and lysogeny is prevented. Thus, there is no 186 CI expression to prevent infection, and the engineered 186 phage would be forced into entering the lytic lifecycle. As an added advantage, the QC024 strain also allowed for visualization of the propagation strain, enabling better interpretation of plating results.

The Δ*tum* 186 phage was able to produce plaques on all three potential propagation strains. The Tum72 (HB96) strain had been used successfully in functional tests to confirm that functional phage are produced when Δ*tum* 186 was plated in the presence of cumate (3.d Engineered Bacteriophage 186 Functional test). Strains Tum72 (HB26) and QC024 produced the clearest plaque formation, but QC024 had the advantage of its fluorescence, distinguishing it from the target lawn. Hence, QC024 was chosen for the final propagation strain because occupation of the phage attachment sites eliminates the possibility of lysogeny and consequent development of immunity to 186 infections. After choosing QC024 as the propagation strain, the ratio of propagation strain to target strain was investigated. The propagation strain density needs to be high enough for successful propagation of the 186 phage, while allowing for visualisation of effects on the target strain. Five different ratios were tested in determining the optimal propagation to target strain ratio; 1:1, 1:5, 1:10, 1:50 and 1:100, respectively. The target *S. aureus* strain used was RN4220. Overnight cultures of each strain were grown separately in LB, sub-cultured the next day and grown to OD_{600} 0.6. Both strains were added to 3mL of 0.7% agar, vortexed and spread over LB agar plates with no selection. Plates were left to grow overnight at 37°C and observed the next day for fluorescence, using a BioRad ChemiDoc XRS+ with a 0.5 second exposure time (Figure 24). Images are overlaid with a bronze filter that turns the fluorescence into a bright white colour (propagation strain), and everything that is not fluorescent is a brown/bronze colour (target). The more fluorescence there is, the brighter the white becomes. The key ratio is where the

propagation is abundant enough for 186 propagation, but the target strain remains the most abundant cell on the plates. Ratios of 1:50 and 1:10 were chosen for the proof of principle screening assay (Figure 24). The propagation strain ratios at 1:10 and 1:50 show individual colonies that are still observable but not intense enough to create a solid appearance. The fluorescence of the 1:1 and 1:5 ratio plates was too intense, and the 1:100 ratio propagation colonies was too diluted (Figure 24). Being too diluted means that the 186 released from the expression strain is less likely to encounter and infect the propagation strain.

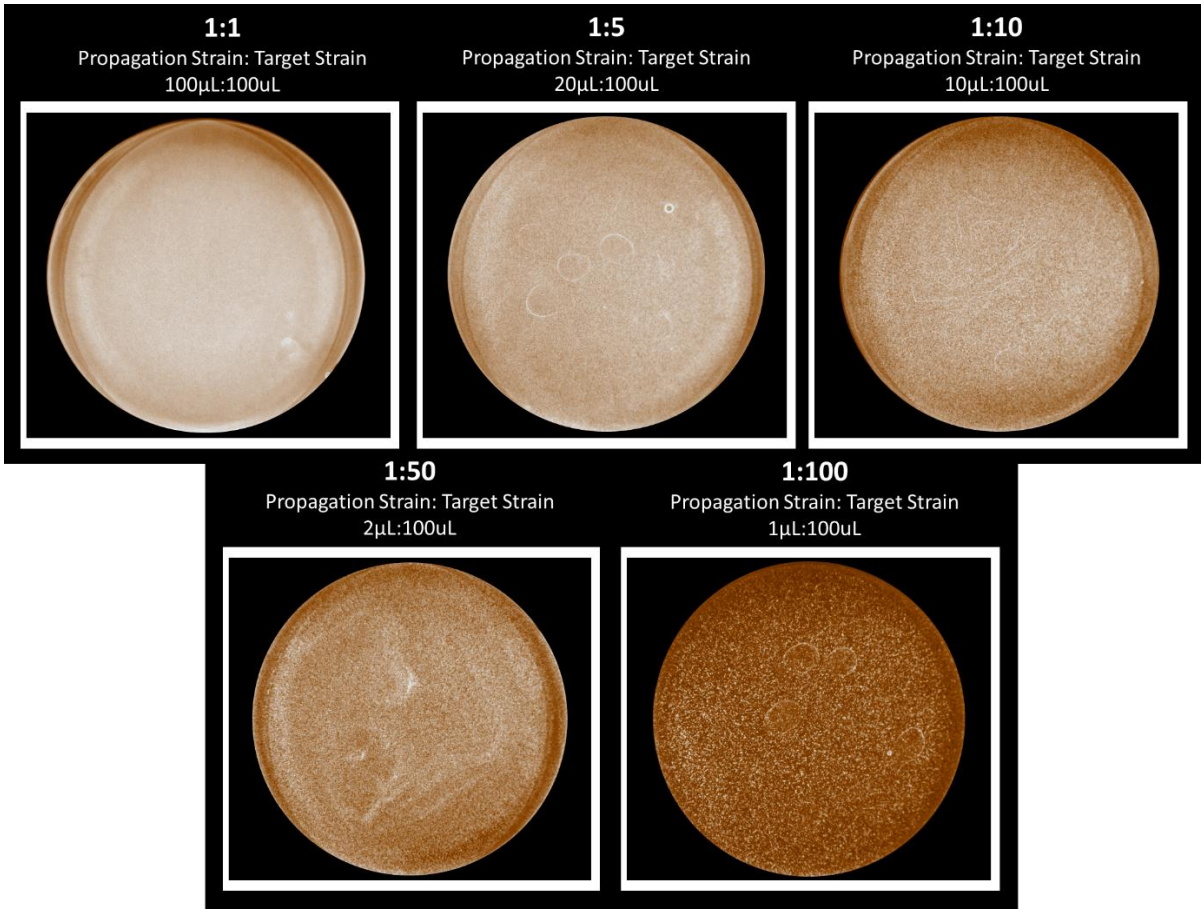


Figure 24: Determining Optimal Propagation Strain to Target Strain Ratios. *Plates contains different ratio of target strain (RN4220) to a propagation strain (QC024). The propagation strains contain two fluorescent "Td-tomato" genes, enabling observation of the two strain possible. Each photo is taken using 0.5 second of exposure time and processed using a bronze filter, with the fluorescence appearing white and non-fluorescence appearing brown. As the amount of fluorescence decreases, decreasing in propagation strain plated, the plate image becomes a darker brown/bronze colour. The target strain added to plates remained constant at 100 µL, with the volume of propagation strain decreasing in each plate. Five ratios of propagation: target strain were tested, 1:1, 1:5, 1:10, 1:50 and 1:100.*

4.c PHEARLESS V2.1 Expression Strain Development and Optimization

The parental pIT4 plasmids needed for this study (carrying the *attP2* site) were constructed using Gibson assembly. Trial experiments showed that these plasmids were able to efficiently integrate into the corresponding *att2* site that had engineered into the 186 prophage genome (methods described in Chapter 2). Next, the 186*pR* and 186*pJ* promoters driving expression of the test endolysin, ClyF, were re-tested. Plasmids were sequenced before integration and integration were confirmed using PCR. Integrating the 186*pJ* and 186*pR* pIT4 plasmid adds ~3.1Kb to the 186 prophage genome. Four new expression strains were constructed;
E4643 [pIT4-KT-cymR-pCym-*Tum*]$^{\phi 21}$ [186(Δ*tum, 186pJ ClyF, Spec*)]) (HB152),
E4643 [pIT4-KT-cymR-pCym-*Tum*]$^{\phi 21}$ [186(Δ*Tum, 186pR ClyF, Spec*)] (HB141),
E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ*tum, 186pJ ClyF, Spec*)] (HB143) and
E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ*tum, 186pR ClyF, Spec*)] (HB149).
After the integration step, PCR was performed to confirm that the pIT4 plasmids had been integrated at the correct site (Appendix Fig. 4).

With the *clyF* gene now located in the 186 prophage genome this new system was tested to confirm that; (1) it is still capable of cumate mediated self-lysis, (2) functional phage production and (3) ClyF expression. To test that new phage are still being produced after the pIT4 integration, phage samples prepared from induced expression strain cultures were compared against un-induced samples. Cultures were induced with a high concentration of cumate (400 µM) and incubated with rotation at 37°C for 2 hours. The high concentration of cumate was chosen to maximise lysis. 10 µL of induced and un-induced samples were spotted on to a plate containing a top layer of 0.7% agar containing host lawn strain E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ (HB96) and 80 µM of cumate.

Results from the phage function tests after integrating each of the pIT4 integration plasmids, pIT4_*attP*2_LoxP_S_pR_ClyF and pIT4_*attP*2_LoxP_S_186*pJ*_ClyF showed that both of the new expression strains could still produce functioning phage particles when induced with cumate (Figure 25). This was tested in both Tum expression strains, Tum72 and Tum wild type, which showed that each expression strain tested produced a killing only in the induced sample (Figure 25). A 186 wild type phage stock was used as a positive control. No plaques were produced in the

un-induced phage sample, showing little to no spontaneous induction from “leaky” Tum expression.

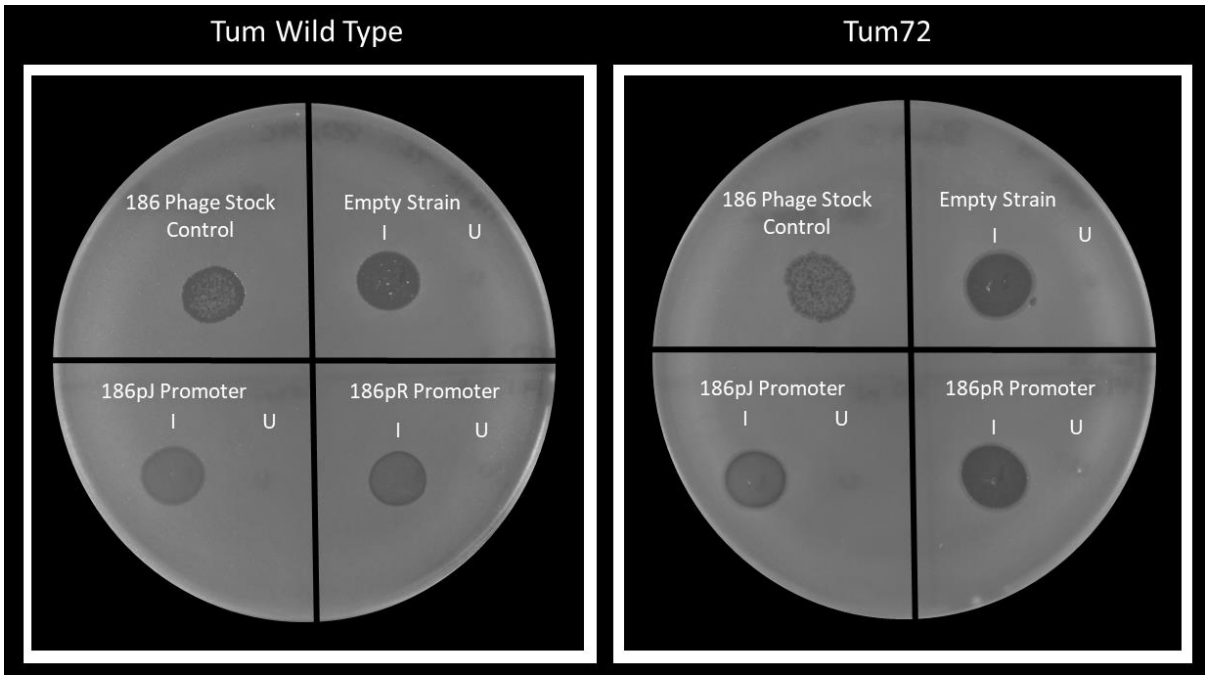


Figure 25: Phage functional tests after integration of the pIT4 integration plasmids into the Tum wild type and Tum72 lysis strains. *After confirming integrations using PCR, expression strains were induced with 400 µM of cumate to generate lysates containing phage. These phage stocks were plated on a E. coli lawn strain (HB96) and compared against a 186 wild type phage stock. Plates contained 80 µM of cumate to maintain a background expression of tum since the lawn strain contained the inducible cumate switch. Induced (I) and Un-induced (U) phage stocks were prepared from strains created from expression strains containing either the 186pJ promoter (HB152, HB143) or 186pR promoter (HB141, HB149), in both the Tum Wild Type lysis strain (Tum wild type) (HB152, HB141) and Tum72 lysis strain (Tum72) (HB143, HB149). The parental strains, HB134 (Tum72) and HB133 (Tum wild type), used to generate both ClyF expression strains (before pIT4 integration) were used as the empty control (Empty Strain). All the expression strains including the empty parental strains showed phage production in the induced sample and no induction in the un-induced sample. Two biological replicates of the assay using these conditions were performed producing identical results.*

After confirming that all of the strains are capable of (1) controlled cell lysis and (2) production of functional phage production. Each of the ClyF containing strains; HB152 (Tum wild type, 186*pJ*), HB141 (Tum wild type, 186*pR*), HB143 (Tum72, 186*pJ*), HB149 (Tum72, 186*pR*) were tested to confirm ClyF expression . This was performed using the PHEARLESS protein screening protocols optimised in Chapter 3. Cultures are grown to OD_{600} 0.6, diluted in a one in ten dilutions, with 10 µL

spotted onto a section of a plate containing the host strain and a specific cumate concentration (40 µM or 120 µM). Parental strains of the Tum wild type (HB127) and Tum72 (HB128) with the Cm resistance gene removed were used as the negative controls for the respective Tum wild type/Tum72 strains.

Assay using the control endolysin ClyF showed interesting results when comparing between the promoters, as well as comparing cumate concentrations. The new empty control strain HB127 shows no activity at 40 µM or 120 µM cumate, however, there is the “lawn thinning” effect occurring at the 40 µM cumate. As previously shown, the Tum wild type mediated cell lysis is less efficient at 40 µM leaving a proportion of the expression cells alive. This effect is only seen on the 40 µM cumate, at 120 µM cumate the lawn is not disrupted where the samples were spotted.

Comparing the promoters in the Tum wild type expression strain shows that there is little or no activity of ClyF when it is controlled by the 186*pR* promoter in the 186 genome (Figure 26). The 186*pR*-ClyF strain gave results indistinguishable from the empty control strain at 40 µM cumate. At 120 µM cumate the 186*pR* samples showed a small amount of activity, looking slightly different to empty control samples in the undiluted sample (U) and first 1/10 dilution (I). In the 186*pJ* promoter strain (Figure 26) ClyF is being effectively expressed at 120 µM cumate with killing observed into the third and fourth 10-fold dilutions.

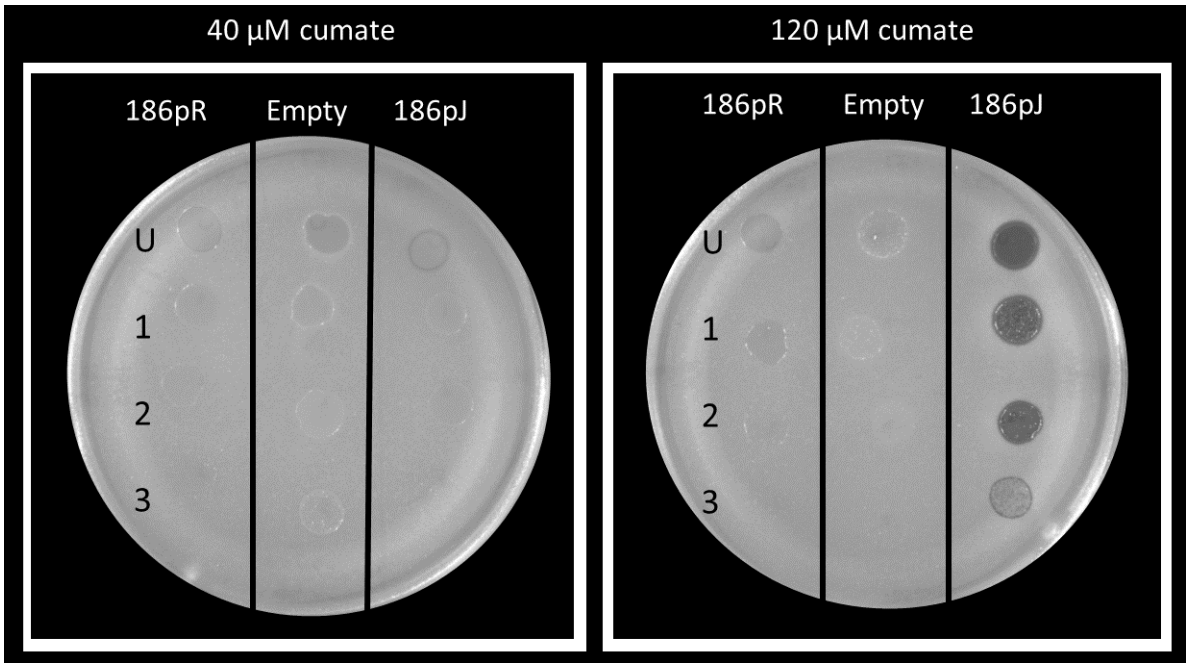


Figure 26: PHEARLESS Endolysin Screening Assay: Comparing Promoters in the Tum Wild Type Lysis Strains with integrated pIT4 ClyF Plasmid. *Comparing promoters on the phage integration plasmid,*

186pJ-ClyF (HB152) and 186pR-ClyF (HB141), in the tum wild type lysis system, using ClyF as an indicator of expression efficiency. The empty (no ClyF) expression strain (Empty) used as the negative control was the tum wild type parental lysis strain used to integrate the pIT4 plasmid (HB127). The 186pR promoter strain (pR) shows weak ClyF expression and target cell (RN4220) lysis in the undiluted (U) and first dilution 1/10 (1) sample on the 120 µM of cumate plate. No target cell lysis is observed in any of the other sample or in any samples on the 40 µM cumate plate. The 186pJ expression strain (186pJ) showed some weak target cell lysis in the undiluted (U) sample on the 40 µM cumate plate, though this is questionable when compared to the empty control. On the 120 µM cumate plate, the 186pJ promoter strain shows strong ClyF expression and target cell lysis at each dilution. Photos were taken using a ChemiDoc XRS+, using the white light grey image settings and two biological replicates of the assay using these conditions were performed producing identical results.

Results for the two promoters tested in the Tum72 lysis strain however, showed different results compared to the Tum wild type lysis strain (Figure 27). The negative control strain, HB128, as expected produced no killing of the target strain.

The 186*pR*-ClyF promoter strain was able to produce killing in the undiluted sample (U) and first (1/10) dilution sample (1) on the 40 µM cumate plate (Figure 27, 40 µM Cumate). The next dilution in the series does appear to contain a small amount of target cell lysis, shares a similarity to the “lawn thinning” effect of the control. On the 120 µM cumate plate, 186*pR* samples show identical results to the negative control strain. No killing was present in even the undiluted sample, with only the presence of individual colonies of the expression strain.

In the Tum wild type results (Figure 26, 120 µM Cumate) the 186*pJ* expression strain only showed strong expression on the 120 µM cumate plate. In contrast, for Tum72 (Figure 27, 40 µM), 186*pJ* shows the strongest lytic active on the 40 µM cumate plate, with killing present at every dilution tested, consistent with the earlier observation that Tum72 is more effective at bringing about cumate mediated lysis.

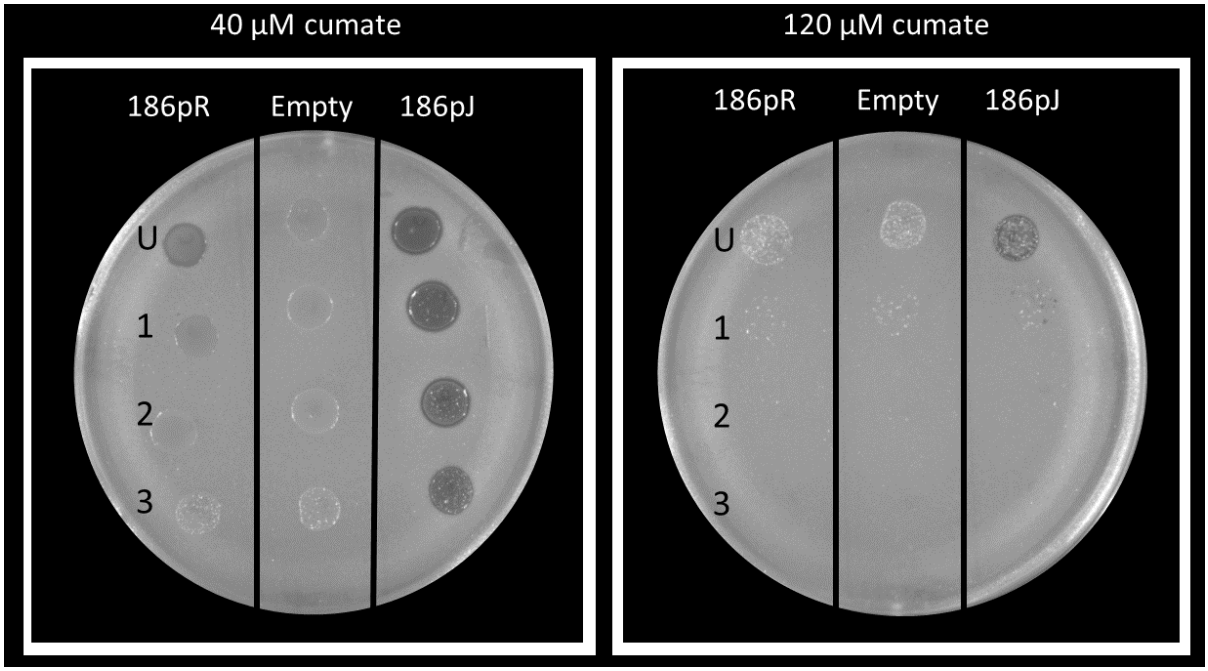


Figure 27: PHEARLESS Endolysin Screening Assay: Comparing Promoters in the Tum72 Wild Type Lysis Strains with integrated pIT4 ClyF Plasmid*. Comparing promoters on the phage integration plasmid, 186pJ-ClyF (HB143) and 186pR-ClyF (HB149), in the Tum72 lysis system, using ClyF as an indicator of expression efficiency. The empty control strain (Empty) used as the negative control was the tum wild type parent lysis strain used to integrate the pIT4 plasmid (HB128). The 186pR promoter strain (pR) shows some target cell (RN4220) lysis in the undiluted (U) and first 1/10 dilution (1) sample on the 40 µM cumate plate, with no killing present on the 120 µM cumate plate. The 186pJ expression strain (186pJ) shows efficient target cell death in every sample on the 40 µM cumate plate. Photos were taken using a ChemiDoc XRS+, using the white light grey image settings and two biological replicates of the assay using these conditions were performed producing identical results.*

With the successful design and testing of the PHEARLESS version 2 system, based on the 186 prophage carrying the endolysin gene, a second set of tests were performed where in addition to the target strain, the plating also included the propagation strain, QC024. The ratio of propagation strain to target strain used was 1:50. Plates either with or without the propagation strain, using the same expression strain dilutions were compared, to test the ability of the new progeny phage to propagate and its effect on target cell killing. If the progeny phage were successful in propagation and continuous release of ClyF, it was predicted that increased target cell lysis would be seen at the lower dilutions. As expected, for both empty strains (HB127, HB128) (Figure 28, A. Empty, B. Empty) on each of the cumate concentrations no target cell (RN4220) lysis was observed. There is a minor difference in appearance between where the empty samples have been added and the rest of the lawn. This is likely due to the difference in the phenotypes between the

target strain, *S. aureus,* and the propagation strain, *E. coli.* The *E. coli* propagation strain (QCO24) produces a speckled appearance when mixed with the *S. aureus* strain. This effect is strongest on the 40 µM cumate plates where more of the expression strain has survived. On the 120 µM cumate plate both negative control expression strains for both tum variants produced the clean results (Figure 28, 120 µM cumate), with very little indication of where the samples have been applied to the plates.

With the ClyF expression strains, the 186*pR* promoter produced some target cell lysis in both the Tum wild type and Tum72 lysis strains, with the most obvious lysis occurring on the 40 µM cumate plate in the Tum72 expression strain. 186*pR* produced the weakest lysis on the 120 µM of cumate plate with the Tum72 expression strain, with the samples appearing to look very similar to the empty control strain (Empty) with the exception of a halo around the spot where the samples have been applied.

In contrast to the results without the propagation strain, on the mixed lawn plates both Tum expression strain produced target cell killing at all dilutions on both the 40 µM and 120 µM cumate plates. For this to be occurring the phage progeny released must be infecting and lysing the propagation strain, increasing the local ClyF concentration. This provides strong evidence that the progeny phage are capable of continuing the expression of ClyF (3), and in turn should work for any potential protein of interest integrated at the attachment site in the 186 genome. The final version chosen for the PHEARLESS V2 expression strain was the Tum72 expression strain, carrying the 186*pJ* promoter. Tum72 was chosen over Tum wild type due to the requirements for efficient cell lysis in the next version of the PHEARLESS assay, adapted for Protein Library Screening.

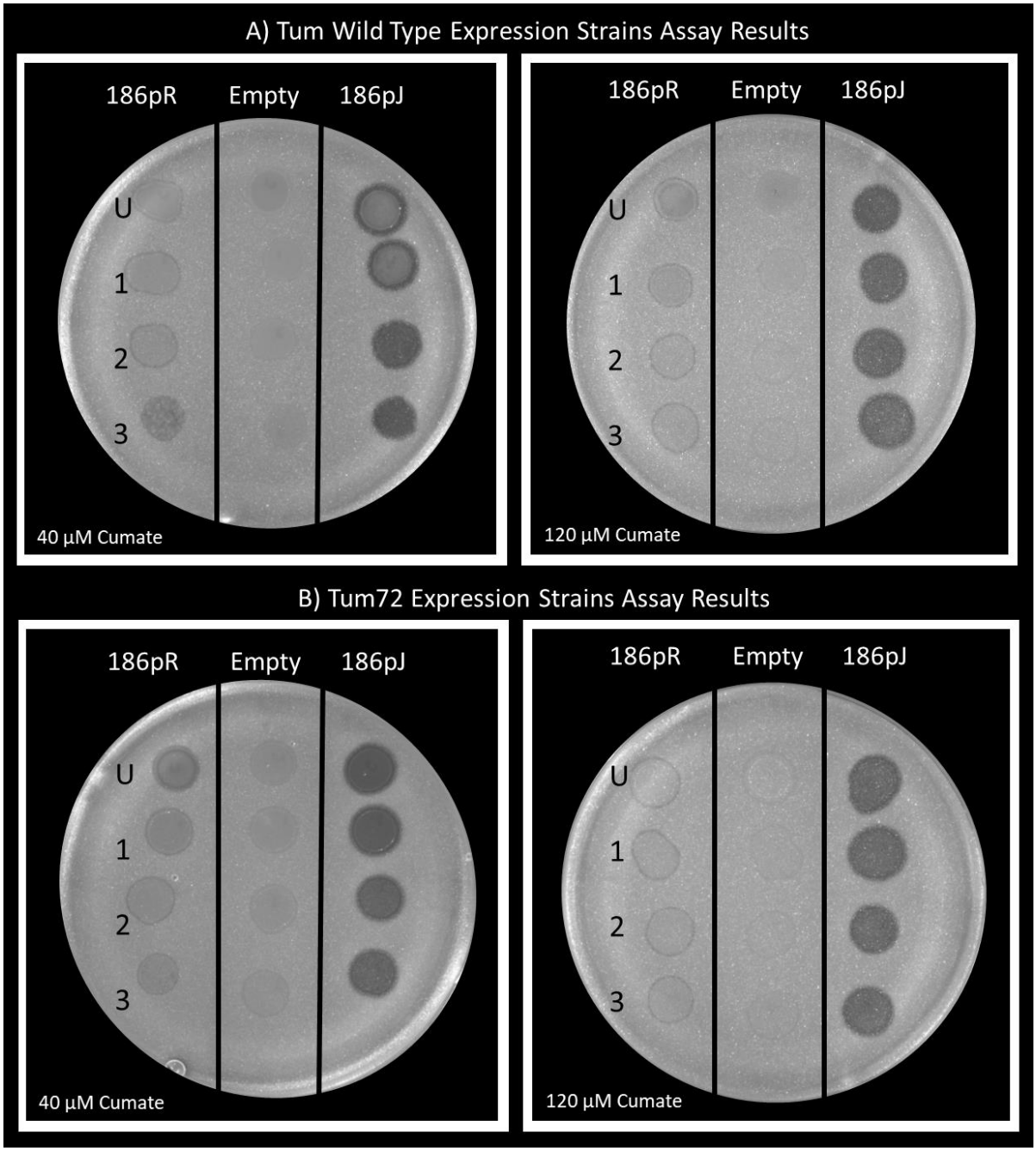


Figure 28: PHEARLESS Protein Screening Assay: Comparing the Tum lysis and Promoter-ClyF Expression Strain Variants on Plates Containing both Target and Propagation Strains. *Shows the Tum Wild Type expression strain; 186pJ (HB152), 186pR (HB141) and Empty (HB127) PHEARLESS Protein Screening Assay with the addition of the propagation strain (QC024) to target strain (RN4220) at a 1:50 ratio respectively. The 186pR promoter (HB141) strain only shows killing in the third dilution (1/1000) sample. The rest of the samples on the 40 µM Cumate and 120 µM Cumate plates contain small halos around where the sample has been added. The 186pJ promoter shows killing in all the samples on the 40 µM Cumate and 120 µM Cumate plates. B) Shows the Tum72 expression strain; 186pJ (HB143), 186pR (HB149) and Empty (HB128) PHEARLESS Protein Screening Assay with the addition of the propagation strain (QC024) to target strain (RN4220) at a 1:50 ratio respectively. The 186pR promoter shows the same haloes present in the Tum wild type samples on the 120 µM Cumate*

plate. 186pR shows clearer killing on the 40 µM Cumate plate in the undiluted (U), first dilution (1/10) with halos present in the second (1/100) and third (1/1000) dilution. 186pJ produces killing in every sample on both the 40 µM and 120 µM Cumate plates, identical to the results for the Tum wild type strain. Photos were taken using a ChemiDoc XRS+, using the white light grey image settings and two biological replicates of the assay using these conditions were performed producing identical results.

4.d PHEARLESS V2.1 Mutant Library Screening Proof of Principle

There are a number of advantages to placing the endolysin gene into the 186 genome; (1) it will exist as a single copy until expressed, (2) antibiotics are not required to maintain it, (3) the progeny phage will also contain the protein gene that can be retrieved. This third feature easy retrieval of the genotype for a successful phenotype is what enables the system to be used for library screening. Screening of a large number of variants can be used for both protein discovery and protein optimization. For screening large mutant libraries, a physical copy of each mutant clone must be retrievable. For a successful mutant screening assay, a large library size, rapid screening and simple retrieval of the best mutants are desirable properties. The main noticeable limiting factor for the system that required optimisation was the efficiency of the integration plasmid assembly using Gibson assembly and the integration of the Gibson assembly. Other potential limiting factors include whether the phage released from the initial expression cell is efficient enough to spread and infect the surrounding propagation cells. Then, following infection, if the protein expressed from the propagation cells is sufficient to produce visible killing.

To test whether the PHEARLESS Version 2 assay is usable for screening a mutant library for antimicrobial activity, a proof of principle assay was developed, based on selecting a revertant from a large library. The test was designed around creation of an inactive ClyF and screening for possible revertants in a library generated by error-prone PCR. An inactive version of ClyF carrying a missense mutation at a conserved catalytic residue was created, this inactive *clyF* gene was used as the template in error-prone PCR to create a mutant library. The mutant library was then assembled into the pIT4 plasmid using Gibson assembly, creating a library of pIT4 plasmids each containing a different ClyF mutant. The Gibson assembly reaction was integrated directly into the att2 site of the 186 prophage in the expression strain. The protocols found in Chapter 2.i PHEARLESS Protein Library Screening Protocols were then use to screen the mutant library for a revertant ClyF.

The mutant library was created using PCR and Gibson assembly, the backbone fragments were amplified from the empty pIT4 plasmid (Figure 34, B) using Phusion PCR protocols found in Chapter 2.b.ii Polymerase Chain Reaction (PCR). Fragments were digested with restriction enzyme DpnI (Chapter 2.b.iii Restriction Digestion) as well as purified using Monarch DNA Gel Extraction Kit (Chapter 2.b.iv DNA Clean up). This was done so there would be no plasmid PCR template present. In the proof of principle test the ClyF(C36G) fragment was amplified using Phusion PCR from the pZS(^)45 expression plasmid as the template. The pZS(^)45 expression plasmid was chosen since it is unable to integrate, therefore would not contaminate the phage if carried through. The ClyF(C36G) gene fragment was amplified using primers that attached to the template directly before the start codon and at the 3' end of the gene, including the stop codon. The Phusion fragment of ClyF(C36G) was digested with restriction enzyme DpnI as well as purified using Monarch DNA Gel Extraction kit. A second round of PCR was done using the Phusion fragment of ClyF(C36G) as the template. This second round of PCR used the error prone PCR protocols (Chapter 2.c.ii Error Prone PCR) to incorporate at least 2-3 mutations in the amplified gene fragments. Mutations would only occur between the primer sites, maintaining the overlap sequences. The final mutated fragments were purified using Monarch PCR and DNA Clean-up kit. This sample is a mixture of the amplified, mutant fragments and a small amount of the original wild type Phusion template.

The error prone library in assembled into the pIT4 error prone library using Gibson assembly (Chapter 2.c.iii Plasmid Fragment Assembly - Gibson assembly). The Gibson assembly reaction was purified using Monarch PCR and DNA Clean-up kit and then transformed directly into the final expression strain using electroporation (ECC) protocols (Chapter 2.c.v Preparing and Transforming Electrocompetent Cells (ECC)). The final competent expression strain carries the pre-induced helper plasmid for the *att2* attachment site and were used the same day they are prepared. Ordinarily the Gibson assembly would be assembled in a cloning strain, E4644 to allow for *pir* dependent pIT4 replication, and a DNA mini-prep made for transformation into the final expression strain. However, in the protocol developed here a single step transformation of Gibson assembly mix into the final expression strain was used. This step is expected to be the bottle neck for how many mutants can be tested. Both the Gibson assembly and integration steps need to be as efficient as possible.

4.e Generation of an Inactive ClyF Protein

Three variants of inactive ClyF (Figure 29, B, C and D) were designed, each carrying nucleotide changes in the conserved catalytic residues of the catalytic domain. The characterization study of ClyF by Yang and Zhang (130) predicted ClyF's catalytic domain to contain three predicted conserved residues, Cysteine (Cys) 36, Histidine (His) 99 and Aspartate (Asp) 116 (Figure 29, F). Cys36 and His99 were selected as the amino acids to be mutated. The aim was for the entire protein to be expressed as an intact, correctly folded protein, but lacking any catalytic activity. If mutating the individual predicted conserved residues did not result in inactivation of the protein, a premature stop codon could have be used. However, since the catalytic domain is at the N-terminal, the premature stop codon would need to be located close to the start of the protein. The amino acids were first mutated to an alanine, changing the first two nucleotides in both mutants. Alanine was chosen since it is both small and neutral (175). It was predicted that since alanine's methyl side-chain is non-reactive, it is rarely directly involved in protein function therefore would not interfere with the protein structure (176).

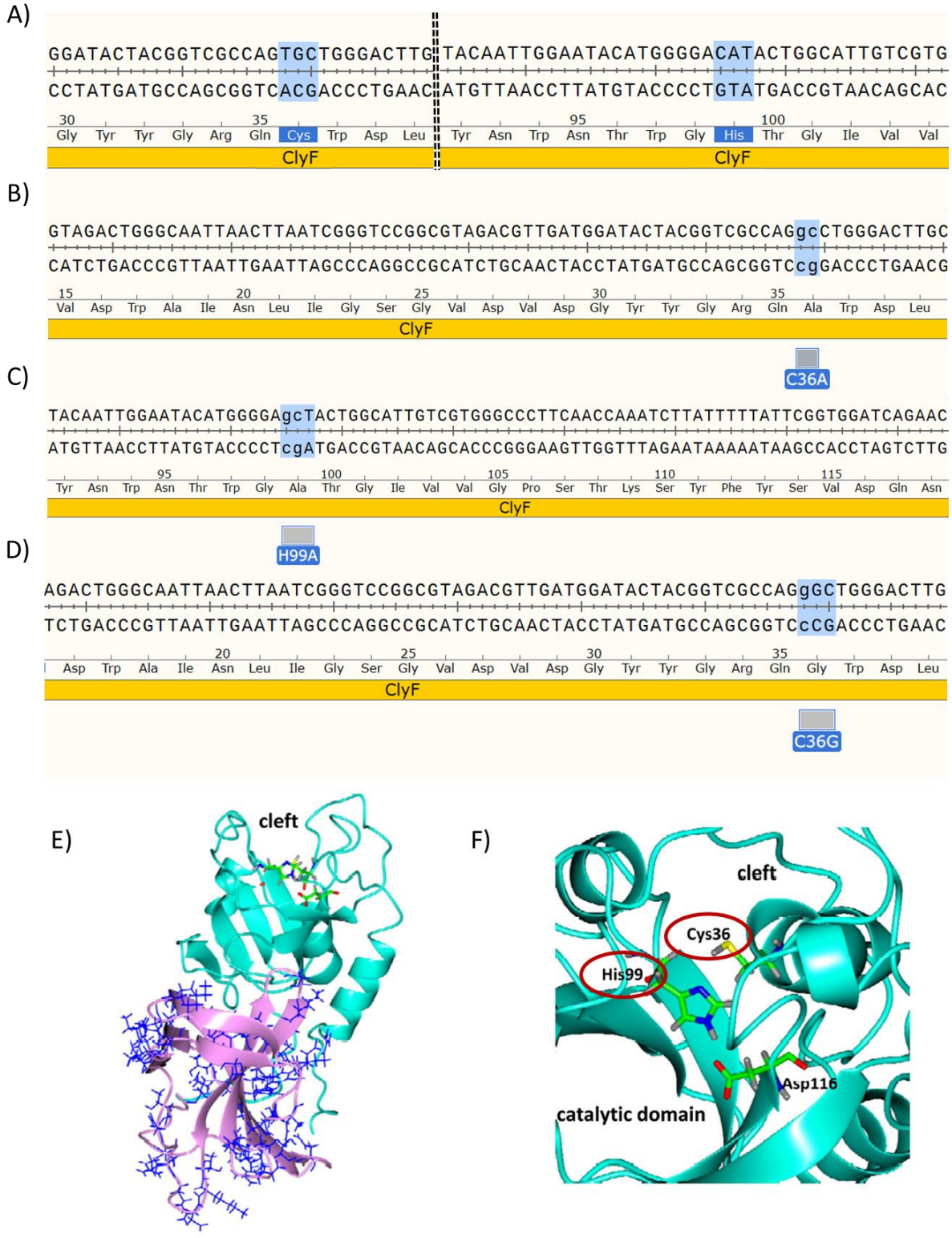


Figure 29: ClyF Inactive Mutants Constructs Sequences and 3D Structure Modelling of ClyF. *Figure A) shows the wild type sequence for ClyF, Cys at the 36 amino acid and a His at the 99 amino acid position. B) In the first mutant, amino acid 36 cysteine was changed to an alanine (C36A). C) In the second mutant, amino acid 99 histidine also changed to an alanine (H99A). D) The third mutant was also a Cys 36 amino acid mutation, but mutated using a single nucleotide to change Cys to Gly (C36G). Both amino acids are conserved catalytic residues in the catalytic domain. D) and E) are reprinted and adapted respectively from "A novel chimeric lysin with robust antibacterial activity*

against planktonic and biofilm methicillin resistant Staphylococcus aureus" by Yang et al. (130) shows the structure of ClyF, the cell wall binding domain (pink) and the catalytic domain (blue). E) Close-up view of the catalytic domain's conserved catalytic residues, the two circled in red were chosen to make inactive mutants.

Both mutants were created using PCR and Gibson assembly (Chapter 2.c Cloning Strategy) with a primer containing the two nucleotide mutation (Table 3: Primers Used During the Course of this Thesis.). The ClyF mutants, ClyF(C36A) and ClyF(H99A), were assembled into the pZS(^)45 expression plasmid for initial testing using the PHEARLESS version 1 expression strain. Both mutants were tested using the PHEARLESS Protein Screening Protocols comparing them to the wild type ClyF, using protocols found in Chapter 2.h PHEARLESS Protein Screening Protocols. Both mutants were compared to either the negative control (Empty) (HB115) containing no ClyF or the wild type ClyF control (ClyF) (HB114) on individual 120 µM cumate plates. The samples were plated this way so that the mutants, ClyF(H99A) (HB173) and ClyF(C36A) (HB172), could be directly compared to each other.

The ClyF wild type control plate (Figure 30) shows lytic activity in the undiluted sample and each of the dilution samples. Compared to the wild type ClyF, neither ClyF(H99A) or ClyF(C36A) showed any activity (Figure 30).

The Empty control plate (Figure 30) shows no lytic activity present in the empty strain samples and both mutant clones show identical results to the empty control strain on this plate. From these results (Figure 30) both mutations completely abolish lytic activity individually.

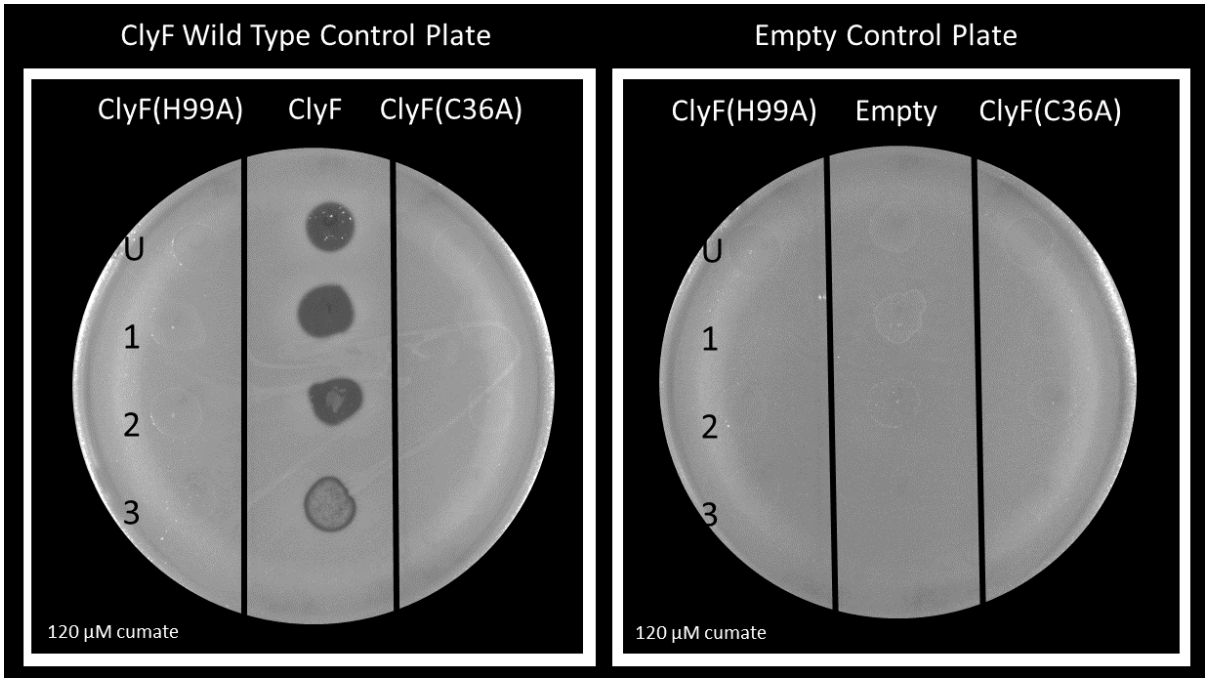


Figure 30: PHEARLESS Protein Screening Assay in the PHEARLESS V1 Expression System: ClyF Mutants Tested for Antimicrobial Activity*. Mutant ClyF proteins, ClyF(H99A) and ClyF(C26A), were tested for antimicrobial activity compared to the ClyF wild type and an empty control strain (HB115). The wild type ClyF on the ClyF Wild Type Control Plate show clear killing at each dilution. The Empty control strain on the Empty Control Plate shows no killing, indicating no target cell (RN4220) lysis. Both mutants (C36A, H99A) produce identical results on both plates, producing no killing. Photos were taken using a ChemiDoc XRS+, using the white light grey image settings and two biological replicates of the assay using these conditions were performed producing identical results.*

However, converting either catalytic residue to an alanine was performed by creating two nucleotide changes. This would not be suitable for the proof of principle assay since it would require both nucleotides to be mutated back to the wild type nucleotides, reducing the chance of generating a revertant in the error prone PCR. The Cys 36 position was mutated to a Glycine (Gly), a non-polar neutral amino acid and was generated with a single nucleotide change.

This mutant was constructed using the same method that the previous two was created. ClyF(C36G) (HB179) was tested also using the PHEARLESS Protein Screening assay to confirm changing this amino acid to a glycine would still result in a loss of function ClyF (Figure 31), showing that the Cys 36 mutation to Gly also shows no activity, like the Cys 36 to Ala mutation, producing identical results to the negative control strain (Empty).

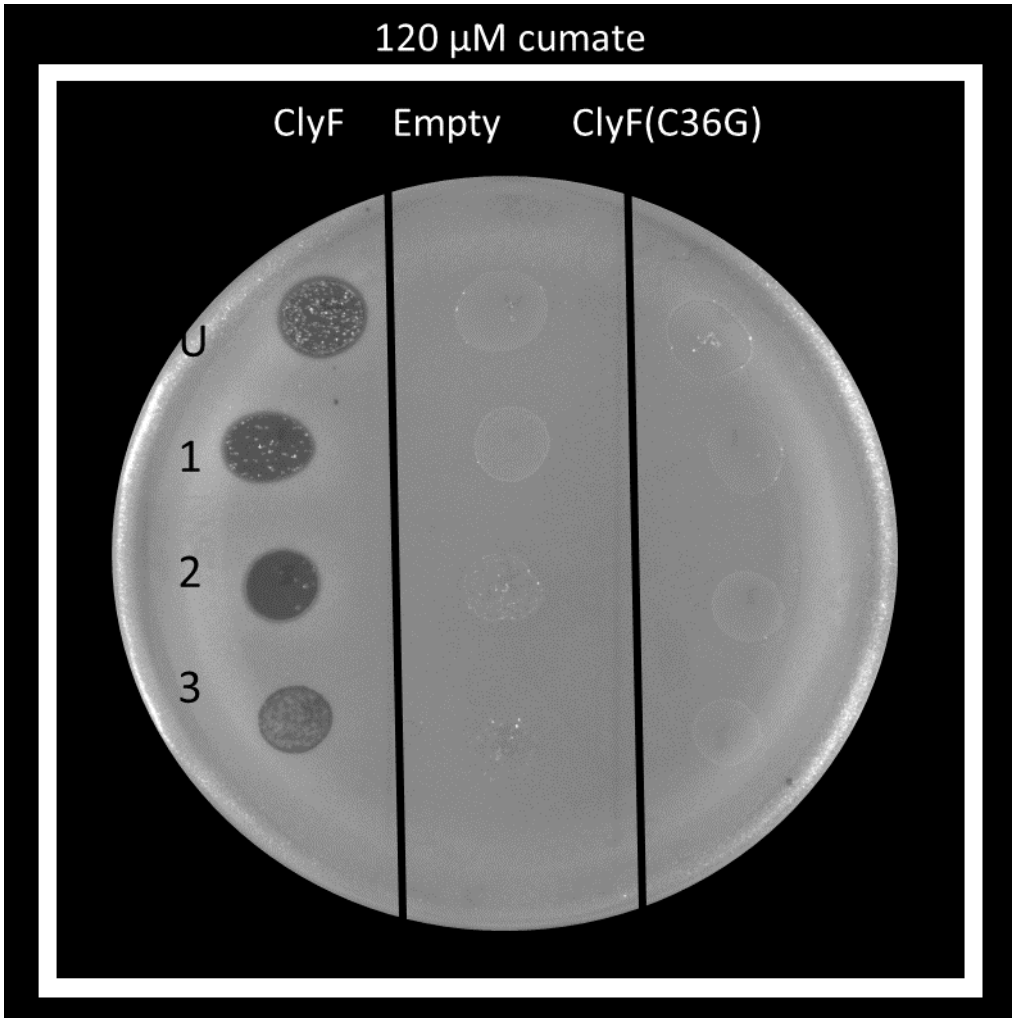


Figure 31: PHEARLESS Protein Screening Assay in the PHEARLESS V1 Expression System: ClyF(C36G) Mutant Tested for Antimicrobial Activity. *Mutant ClyF(C36G) (HB179) was tested for protein activity in the PHEARLESS V1 expression strain against RN4220 and compared against an empty control (HB115) and the ClyF wild type protein (HB114). ClyF showed killing in every sample plated on the 120 µM Cumate plate. Both the Empty control and ClyF(C36G) showed identical results with no target cell lysis observed. Photos were taken using a ChemiDoc XRS+, using the white light grey image settings and two biological replicates of the assay using these conditions were performed producing identical results.*

4.f Analysis of the Sensitivity of the PHEARLESS Protein Library Screening Assay

A critical feature for high throughput screening is being capable of identifying the positive clone in an abundance of negative clones. This sensitivity can place a limitation on the screening system, limiting how many negative clone expression cells can be present without masking the functional expression cells. The aim was to test as many negative colonies per plate as possible, compared to a fixed number of positive colonies, to determine how sensitive the system is for observing clones

containing functionally active proteins surrounded by non-functional clones at different cell counts. In the final assay, the positive clone plaques need to be visible to collect the phage in that location. This was initially performed by plating different ratios of expression strain containing an empty control strain E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ*tum, 186pJ, Spec*)] (HB165) (186(Empty)) against a matching expression strain carrying ClyF, E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ*tum, 186pJ ClyF, Spec*)] (HB143) (186(ClyF)).

Using a fixed number of 186(ClyF) expression cells on each plate (~100 cells), the amount of 186(Empty) expression cells was increased per plate. It was predicted that as the number of 186(Empty) cells increased, the number of visible ClyF clearing would decrease as competition for the propagation strain would grow. This would provide an estimation of a 'safe' number of expression cells to plate per plate for positive plaque detection. Fresh cultures of the 186(Empty) expression strain and 186(ClyF) expression strain made from sub-cultured overnight cultures were grown in 50 mL LB Lennox broths at 37°C, with shaking at 180 to 200 rpm. With the empty pIT4 plasmid or the pJ-ClyF pIT4 plasmid both integrated into the 186 prophage, no antibiotics were required during culture growth. After reaching OD_{600} 0.6 the cultures were diluted in a series of 10-fold dilutions. 50 µL of the 5th dilution of the 186(ClyF) strain was plated onto a LB agar plate with no selection. This dilution was used in every plate containing the 186(ClyF) expression strain, and gave approximately 100 colonies. A series of 10-fold dilutions was also made of the 186(Empty) expression strain, 100 µL of the 6th dilution was plated on a LB agar plate with no selection. The cell count from this plate was used to determine the cell count present on the LB agar plates testing the different ratios of 186(Empty) expression strain.

The plates that contained the ratios of 186(Empty):186(ClyF) were co-plated with the target strain RN4220 and propagation strain QC024 at a 50:1 ratio, both grown to OD_{600} 0.6 using the protocols above. All four strains were added to 3mL of 0.7% agar, then plated on a LB agar plate containing 240 µM cumate. A cumate concentration of 240 µM was chosen due to the abundance of expression strain cells being added to the plate. The amount of 186(ClyF) expression strain remained constant with 50 µL of the 5th dilution added to each of the plates. The concentration of the 186(Empty) expression strain was increased with each plate. The concentration added for the five plates were; 50 µL of the 3rd dilution (~30,000 cells), 100 µL of the 3rd dilution (~60,000 cells), 200 µL of the 3rd dilution (~120,000 cells),

50 µL of the 2nd dilution (~300,000 cells) and 100 µL of the 2nd dilution (~600,000 cells). Plates were left overnight at 37°C and observed the next day for clearing. As previously stated, only phage containing the ClyF gene should produce visible plaques. 186 will be able to propagate using the Td-tomato expressing propagation strain, enhancing visibility of the plaques.

Although each dilution above has an estimated cell count for the 186(Empty) expression strain which was estimated using predicted cell count at OD_{600} ($6x10^8$ cell/mL). The actual Empty cell count was determined for each assay, which is included in the figure (Figure 32). This assay was performed up to four times using different biological samples, providing a clear picture of the sensitivity of the assay system and the limits on the number of negative cells. Plate photos from one set of replicates (Figure 32) show that at up to 40,000 186(Empty) expression cells, the plaques created by the 186(ClyF) expression strain were clearly visible and show identical plaque sizes. Above the 100,000 186(Empty) expression cell count, the number of visible plaques created by 186(ClyF) expression decreased in both number and plaque size (Figure 32).

A graph combining the results from the four biological replicates compares the number of 186(Empty) expression cells to the percentage of 186(ClyF) cells plated (Figure 33). The percentage was determined using the 186(ClyF) strain on a non-selection control plate for each replicates. The 186(ClyF) strain was plated on a non-selection plate and the cell count was used to determine the percentage. However, it must be taken into account that batch to batch variation occurs between plates. On a replicate that was removed due to low 186(Empty) expression cell concentrations the standard deviation of the 186(Empty) expression cell strain was determined to be +/- 19 clones. Cell mixtures also add to the batch to batch variation. The propagation strain while viable is still at a concertation which is 50x less than the target strain. There is a completion between the 186(Empty) expression cell and the 186(ClyF) expression cell for the propagation strain. Depending on starting positions a 186(ClyF) expression cell may end up closer or further away from and available propagation strain cell, reducing the 186(ClyF) propagation and plaque size. From the four replicates recorded, within the first 50,000 186(Empty) expression cell per plate, the majority of 186(ClyF) propagation is observed. With the lowest percentage sitting at around 70% for Replicate 4 (Figure 33). In number of the replicate a percentage score can increase rather than decrease at the 186(Empty) expression

cell count increase. Even increasing above the 100% value, as stated above this is due to the batch to batch variation in the 186(ClyF) expression cell count. Therefore this graph should be observed as a rough estimation of the optimal plating cell count plated from the library screening system.

A graph that combines results for four biological replicates shows that the optimal cell count that could be used from plating is about 50,000 to 100,000 cells per plate for library screening (Figure 33). Plate photos for plates that contained between 10,000 and 40,000 empty control cells demonstrated visually consistent killing formations (Figure 32). However, after increasing the number of 186(Empty) expression cells (Figure 32) above 40,000, killing formation size and the number of individual killing formations are shown to decrease.

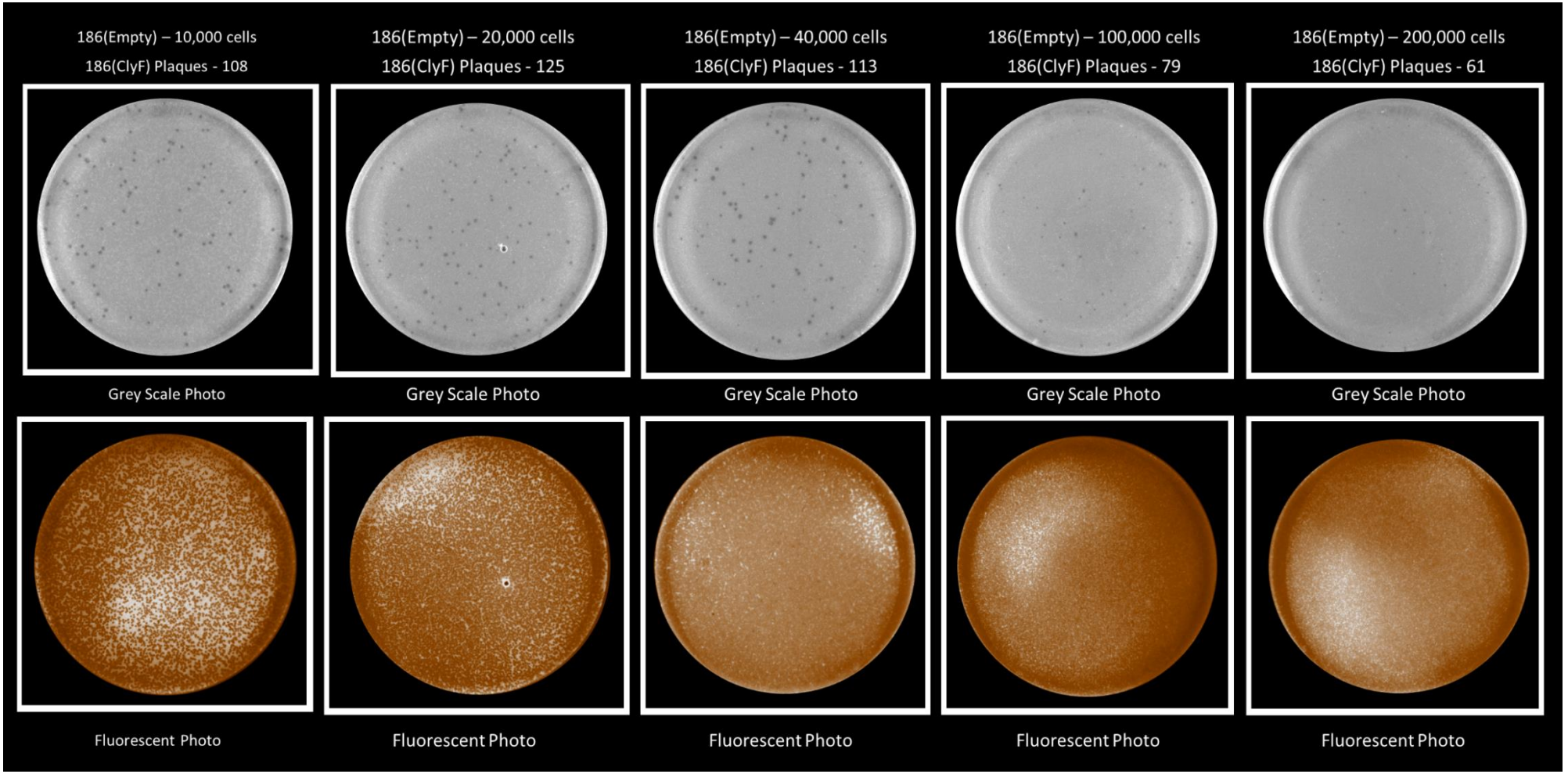

186(Empty) – 10,000 cells
186(ClyF) Plaques - 108
186(Empty) – 20,000 cells
186(ClyF) Plaques - 125
186(Empty) – 40,000 cells
186(ClyF) Plaques - 113
186(Empty) – 100,000 cells
186(ClyF) Plaques - 79
186(Empty) – 200,000 cells
186(ClyF) Plaques - 61
Grey Scale Photo
Grey Scale Photo
Grey Scale Photo
Grey Scale Photo
Grey Scale Photo
Fluorescent Photo
Fluorescent Photo
Fluorescent Photo
Fluorescent Photo
Fluorescent Photo

Figure 32: Sensitivity of the PHEARLESS Protein Screening assay. *PHEARLESS V2 expression strains carrying a ClyF protein gene (HB143) or an empty pIT4 control plasmid (HB165), integrated into a 186 prophage, were plated at different ratios. The aim is to determine how observable the 186(ClyF) plaques are (E. coli propagation cells infected with the ClyF-producing phage kill the S. aureus target strain), when surrounded by large numbers of plaques produced by the 186(Empty) phage, (which infect the E. coli propagation strain, but do not kill the S. aureus target strain). This set of experiments included five ratios and was repeated four times. The lawn was composed of two strains, the target strain (S. aureus RN4220) and the phage propagation strain (QC024) at a 50:1 ratio, respectively. Each plate contained a dilution of the 186(ClyF) expression strain designed to give ~ 100 colonies/plaques per plate. A separate series (not shown) of platings were performed to estimate the number of 186(Empty) expression cells plated on each plate. The top row of images are grey scale photos which show the 186(ClyF) plaques/clearing. The bottom row of images are fluorescent images taken of the Td-tomato fluorescent gene expression produced by the propagation strain (white light= Td-tomato). Using the fluorescent photos, at 10,000 186(Empty) expression cells, individual ClyF$^+$ plaques can be observed. The white pigment observed in the lower panel of photos is the Td-tomato protein, which decreases in amount as the amount of 186(Empty) expression cells are increased, and the propagation strain is consumed. Photos were taken using a ChemiDoc XRS+.*

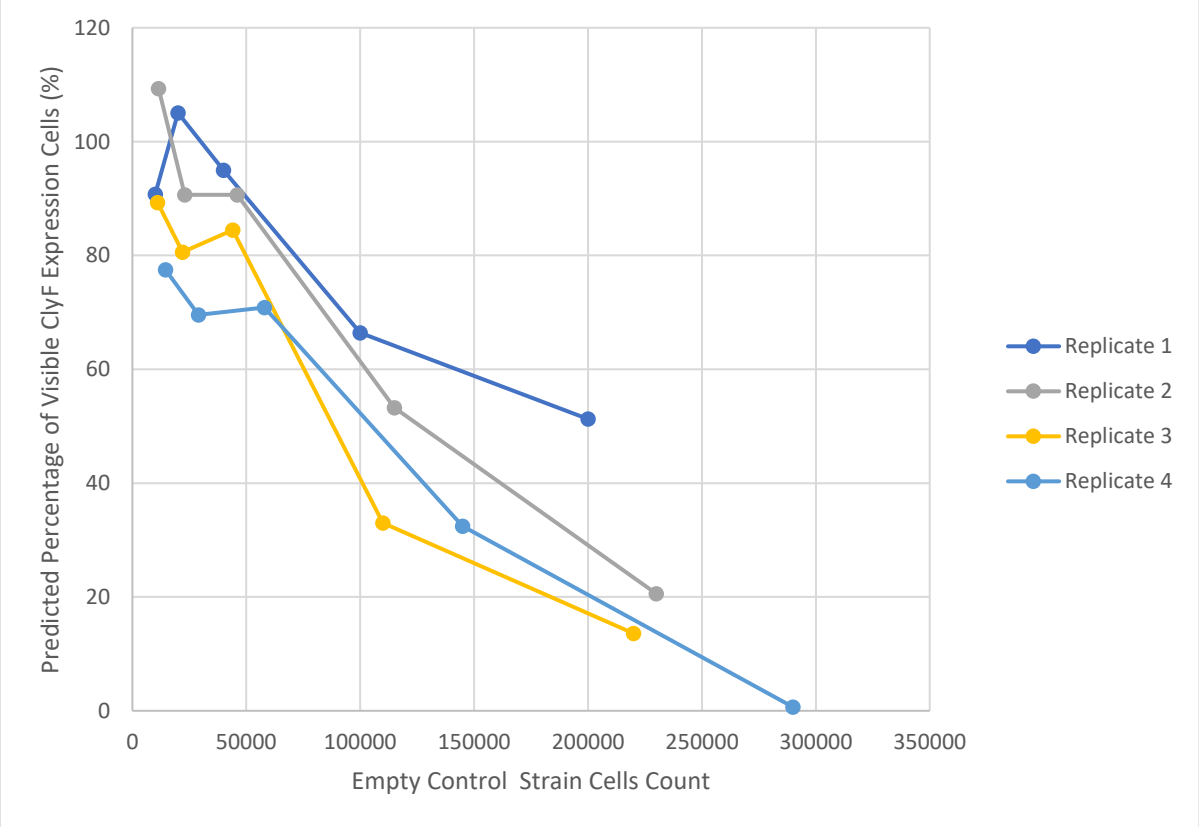


Figure 33: Assay Sensitivity Estimations for the PHEARLESS Protein Library Screening Assay using the PHEARLESS V2.1 Expression Strain. *The graph plots the percentage of ClyF expression cells visible, as the number of background empty control strain cells increase. Four biological replicates*

were performed, each with five plates of increasing number of empty expression cells. All replicates show that below 50,000 empty expression cells, at least 70% of the ClyF clones are visible. Above 100,000 empty expression cells, the percentage of visible ClyF clones decreases to between 60% to 30%. The percentage of visible ClyF clones falls to below 20% in the presence of 200,000 empty cells.

4.g PHEARLESS V2.2; Improved Version of the Expression Strain

An initial concern with the PHEARLESS assay was that during the mutant library screening, successfully integrated clones would not appear due to the expected large number of cells where the pIT4 plasmid had not integrated into the *att2* site within the 186 prophage. This is because a 186 prophage without the integrated pIT4 plasmid would still drive cell lysis, releasing functioning phage particles. When investigating the sensitivity, Chapter 4.f Analysis of the Sensitivity of the PHEARLESS Protein Library Screening Assay, as the number of Empty expression cells (negative control) increased, the number of visible plaques dropped (Figure 32). The Empty phage would dominate infection of the propagation strain, due to their large numbers. This would reduce the number of propagation cells available for infection by the phage carrying the ClyF protein, reducing plaque size. This would be an impediment in the mutant library screen, since the library would always contain a significant proportion of cells without an integrated pIT4 plasmid, effectively acting as a non-integrated (or inactive) library member.

One approach to removing these un-integrated cells from the population would be to select only cells where pIT4 had integrated, using the spectinomycin marker present on pIT4. If, after transformation of the pIT4 library, the sample was plated on a Spec20 agar plate, and incubated overnight at 37°C, the successfully integrated cells could be harvested from the plate the next day, diluted and plated together with the propagation and target strains. While this was a solution to remove the non-pIT4 containing cells (recalling that the plasmid itself is unable to replicate in the expression strain), the integrated cells collected needed to be heavily diluted for plating. It was a concern that any desired mutants would be missed when making the dilutions. Therefore, a new version of the PHEARLESS V2 expression strain was designed (PHEARLESS V2.2) that reduced the number of non-integrated (no pIT4 integration) phage being released and removed the need of a selection step.

4.g.i The Modified pIT4 plasmid

The new version of the expression strain (V2.2) contains all the same components of the V2 expression strain, but improves it by removing the *cos* sequence from the prophage, and placing it on the pIT4 plasmid. Thus, only cells that carry a pIT4(cos^+) plasmid integrated into the modified 186 prophage will be able to package the phage DNA and produce an active phage particle. This re-design would not require antibiotic selection, since the *cos* sequence acts as the selectable marker.

In this new design, it was presumed but had not been tested, that placing *cos* on the pIT4 plasmid would rescue phage production. Support for the concept is provided by the cosmid delivery system (158), including protocols used in Chapter 2.g.i Gene Removal using LoxP Sequences, where adding the λ phage *cos* sequence on to a pE-Cre plasmid, allows the plasmid to be packaged into the λ phage head, rather than the λ genome (158).

To test whether this design would work for a 186 prophage, a modified version of the pIT4 plasmid was created (Figure 34, A and B) containing the *cos* sequence for 186 phage. When the pIT4 plasmid is integrated into the 186 prophage at the *att2* site, the *cos* sequence would be placed nearest to the CP2 portal vertex gene (Figure 34, C) D)), which is close to the native 186 *cos* sequence location (Figure 9, C). The new version of the pIT4 plasmids (Figure 34, A and B) were assembled using fragments amplified from the previous pIT4 plasmid empty and ClyF plasmids (Figure 34, A and C). and with a 186 *cos* sequence fragment located adjacent to the *att2* sequence.

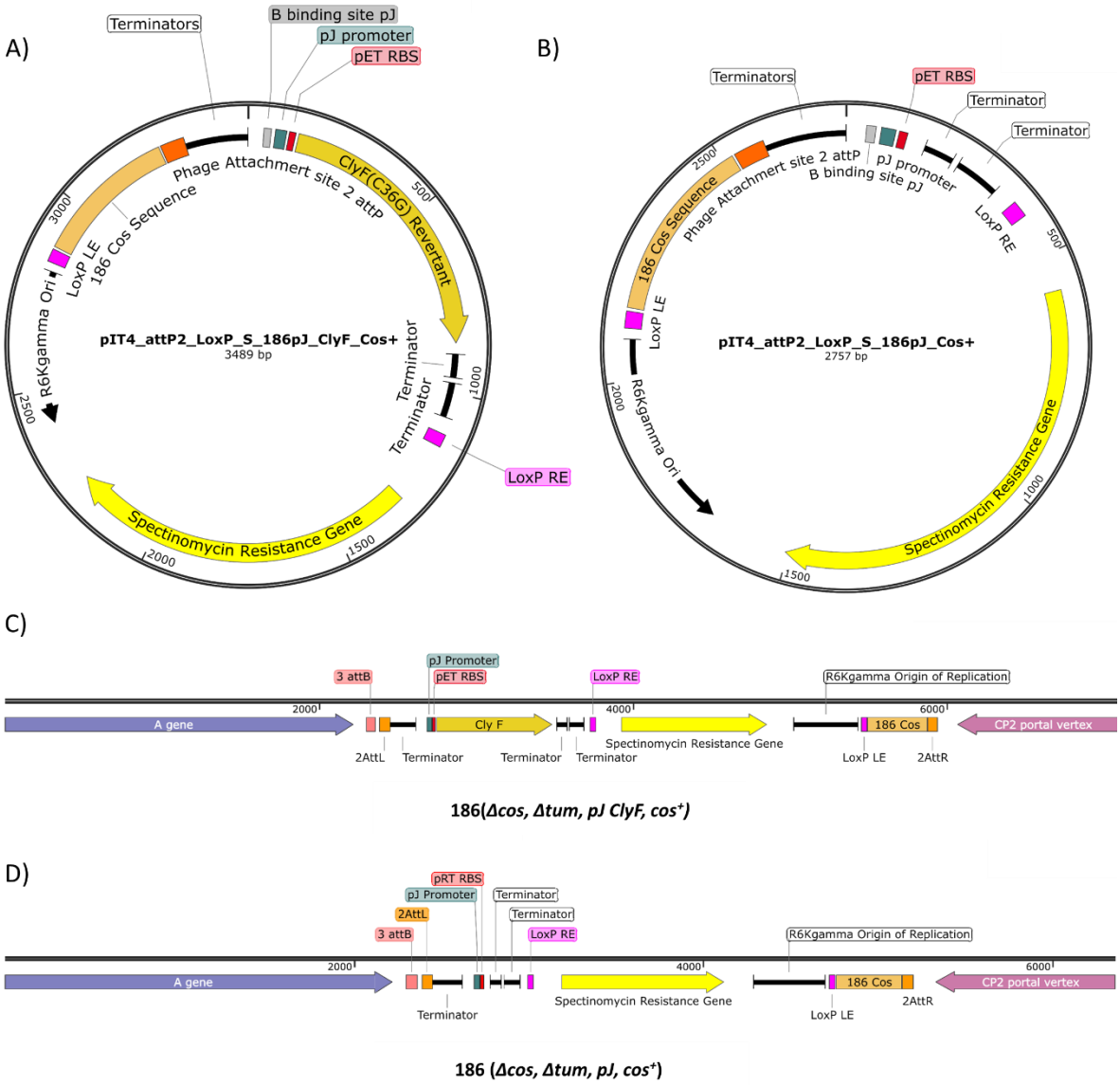


Figure 34: Map of modified pIT4 plasmids that contain the 186 cos sequence and the integration map for the 186 genome. *A) Shows the modified pIT4 integration plasmids containing the ClyF gene controlled by the 186pJ promoter, pIT4_attP2_LoxP_S_186pJ_ClyF_cos+. The plasmid contains all the features of the previous pIT4 plasmid (Figure 23), with the exception of the 186 cos sequence located in between the LoxP LE sequence and the Phage attachment site 2 attP. B) Shows the new empty pIT4 integration plasmids, pIT4_attP2_LoxP_S_186pJ_cos+, that is used as a negative control, containing all the features of the ClyF pIT4 plasmid except for the ClyF gene. C) Shows how the pIT4_attP2_LoxP_S_186pJ_ClyF_cos+ plasmid would be positioned when integrated into the 186 genome, showing the 186 cos sequence located nearest to the CP2 portal vertex gene. D) Shows how the pIT4_attP2_LoxP_S_186pJ_cos+ plasmid is integrated into the 186 genome, showing the 186 cos sequence located nearest to the CP2 portal vertex gene. Figure was generated using SnapGene.*

Next, the redesigned of the pIT4 plasmid that was tested in Chapter 3, Figure 18 showed that the expression strain HB76 which carried the *tum72* cumate switch, and

a 186 prophage without an intact *cos* sequence, was unable to produce functioning phage particles. The parent strain of HB76, HB67 (lacking the GFP plasmid), was used to test whether phage production could be rescued by integrating the 186 *cos* back into the prophage using the pIT4 plasmid. The *att2* helper plasmid (AH6045) was transformed into HB67 to give HB186 which was used to create two expression strains; a *cos*+ expression strain (HB189) and a *cos*- expression strain (HB191). Integration of the pIT4 plasmids were confirmed by PCR.

The *cos*$^{-}$ strain (HB191) was created by integration of the previous version of the pIT4 plasmid pIT4_*attP*2_loxP_S_186*pJ*_ClyF (from HB137) into the new expression strain HB186. With this strain lacking the *cos* sequence from both the 186 genome and pIT4 plasmid, it should be unable to produce functioning phage particles. The *cos*$^{+}$ strain (HB189) has the new pIT4 plasmid containing the *cos* sequence pIT4_*attP*2_loxP_S_186*pJ*_ClyF_*cos*+ integrated into the 186 prophage *att2* site. Both the *cos*$^{+}$ and *cos*$^{-}$ strains will be compared to the positive control strain, E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ*tum,* 186*pJ ClyF, Spec*)] (HB143), which is simply the expression strain carrying its wild type *cos* sequence.

This phage functional assay used four strains (Chapter 2.e Bacteriophage production and Integration) the expression strain stated above and the propagation strain, QC024, which was used to allow plaque formation. All of the strains were grown to OD_{600} 0.6, cultures were then induced with 400 µM of cumate. Before adding the cumate, 1 mL of culture was taken and prepared as the un-induced sample, demonstrating no induction occurs without the presence of cumate. Induced cultures were left for two to three hours, 1 mL of induced culture was taken and centrifuged to pellet the cell debris. The supernatant was treated with three to five drops of chloroform and vortexed. The samples were then pelleted and the top layer of liquid was removed. This step was repeated to prevent any chloroform carryover, giving the final phage sample. A top layer of 0.7% agar containing 200 µL of QC024.

The results for plaque formation showed that the positive control *cos* expression strain, E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ*tum,* 186*pJ ClyF, Spec*)] (HB143), worked as expected, producing a plaque in the induced sample and no plaque in the un-induced sample (Figure 35). The parental strain, E4643 [pIT4-KT-cymR-pCym-*Tum72*] $^{\phi 21}$ [186(Δ***cos,*** Δ***tum, att2, Cm***)] (HB182), without any 186 *cos* sequence is unable to produce any phage particles shown by no plaques present in

the induced sample (Figure 35). The expression strain E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ*cos, ⊿tum,* 186*pJ ClyF, cos*$^{+}$*)*] (HB189) was created from HB182, having the 186 *cos* sequence reintroduced into the 186 genome via integration of the pIT4 cos^{+} plasmid (Figure 34, A). This strain is now capable of producing functional plaques as shown in the induced sample (bottom right, Figure 35).

To confirm that it is the *cos* located on the pIT4 plasmid rescuing phage production, the last expression strain tested was E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ*cos, ⊿tum,* 186*pJ ClyF)*] expression strain (HB191) that had the previous version of the pIT4 plasmid, lacking a cos sequence (Figure 23, A) integrated into the 186 genome. This expression strain produced no plaques with the induced or un-induced sample (Figure 35), demonstrating that the *cos* sequence on the new pIT4 integration plasmid is solely required for producing of functioning phage particles.

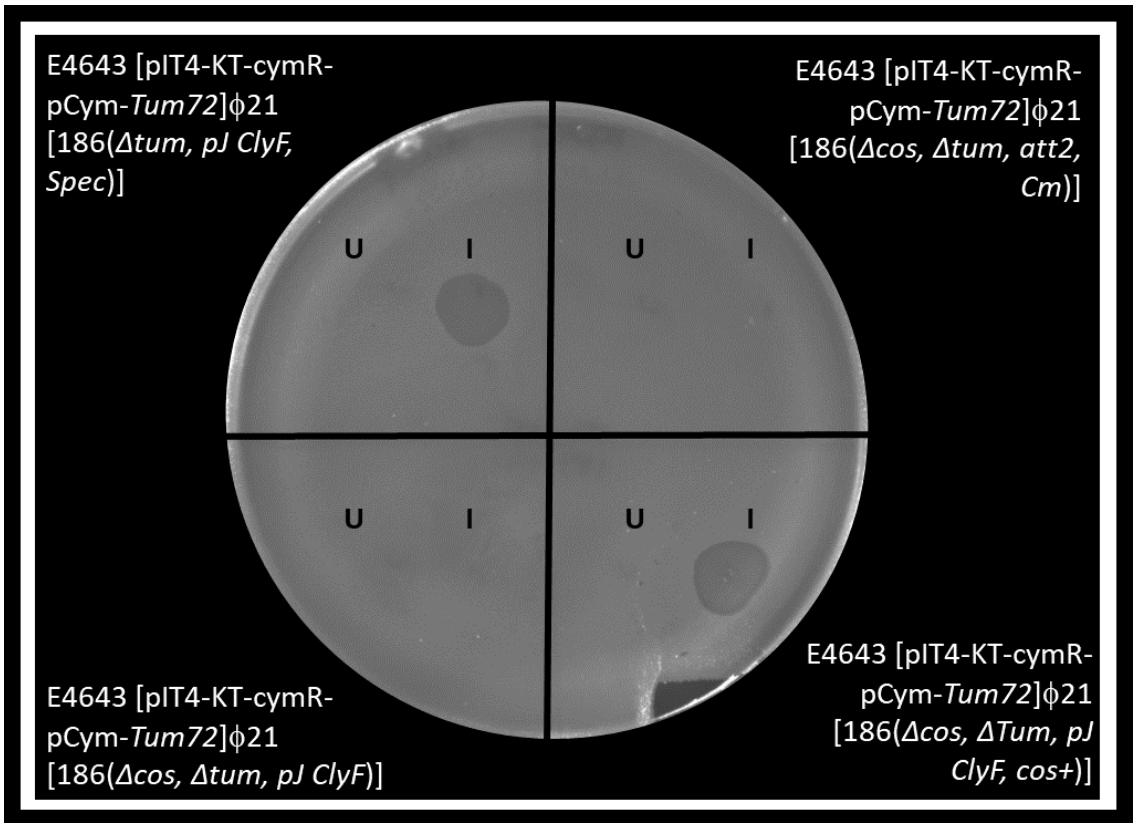


Figure 35: Recovery of Phage Production by Re-introducing the 186 cos Sequence into the PHEARLESS Expression Strain. *Plate contains a QC024 lawn and is divided into four sections. Each section was spotted with an induced (I) and an un-induced (U) sample. Top Left section contains the positive control strain E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δtum, 186pJ ClyF, Spec)] (HB143), producing a clear plaque in the induced sample and no induction in the un-induced sample. Top Right section contains the negative parental control strain that does not contain a native 186 cos sequence and has not had a pIT4 plasmid integrated into it. It shows no phage production present in either*

*sample. The Bottom Right is the induced and un-induced samples for the E4643 [pIT4-KT-cymR-pCym-*Tum72*] ϕ^{21} [186(Δcos, ⊿Tum, 186pJ ClyF, cos$^+$] (HB189) expression strain which has had the cos sequence re-introduced into the 186 genome via the new pIT4 plasmid. This strain produces plaque in the induced sample and none in the un-induced sample. The Bottom Left expression strain belongs to expression strain, E4643 [pIT4-KT-cymR-pCym-Tum72]ϕ^{21} [186(Δcos, ⊿tum, 186pJ ClyF)] (HB191) that is unable to produce functioning phage particle and has had the previous (no cos) pIT4 plasmid integrated into it. This strain is unable to support phage production, shown by the lack of plaque in the induced sample.*

4.g.ii Assembly and Integration Optimization

Now that a successful approach had been developed for ensuring all members of the phage library will carry an integrated pIT4 plasmid, optimisation of the pIT4 assembly reaction was performed. First, it was investigated how many DNA fragments could be used for efficient isothermal assembly of an intact plasmid. The backbone of the pIT4 fragment is ~ 2.4 kb. From previous experience, fragments that are larger than 2 kb are less efficient in isothermal assembly than fragments smaller than 2 kb. A preliminary test of three vs two fragment assembly, was performed. These used either one or two pIT4 backbone fragments, plus a ClyF insert fragment. Samples were prepared as described in Chapter 2 and transformed into the final expression strain E4643 [pIT4-KT-cymR-pCym-*Tum72*]ϕ^{21} [186(Δ*tum, att2,* Δ*Cm*)] carrying the *att2* helper plasmid (HB186) (Table 1), using electroporation protocols and selecting for spectinomycin. Control transformations of Gibson assembly samples containing only the backbone fragments and no insert were also performed to confirm that no pIT4 template had carried through into the assembly reaction.

Transformation results (Figure 36) showed that the three-fragment assembly visibly produced more integration than the two-fragment assembly. The controls (not shown) also confirmed that there was no pIT4 template in any of the fragment samples, giving no colonies. Electrocompetent cell samples are extremely concentrated, and cells that have not undergone transformation and integration are killed by the antibiotic selection, appearing as a smears.

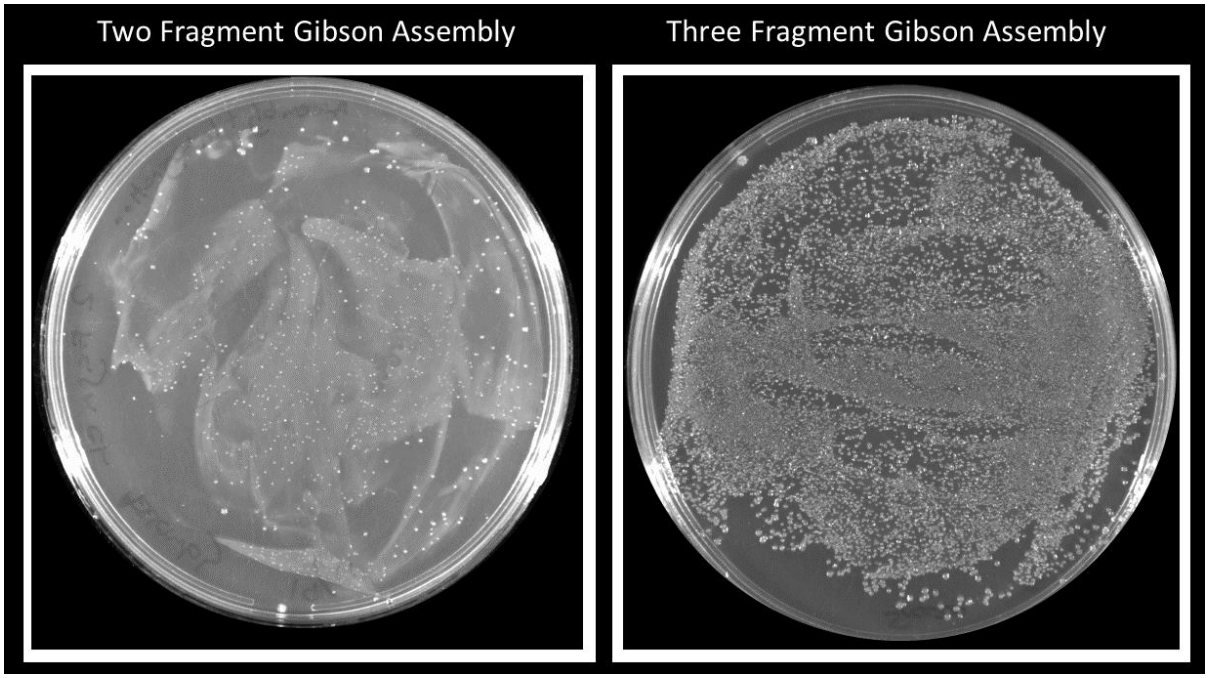


Figure 36: Preliminary Gibson Assembly Assay: Testing Number of Fragments used in Gibson assembly for Protein Library Screening Protocols. *Testing Gibson assembly for pIT4 integration plasmid backbone with a protein of interest gene (ClyF). Two fragment (left) or three (right) fragment assembly with a pIT4 backbone and a fragment carrying the 186pJ promoter and ClyF gene were used. Integrants were selected using spectinomycin, the selection marker for the pIT4 plasmid.*

To investigate the transformation and integration steps separately, three plasmids samples were transformed into the final expression strain (HB186). HB186 ECC were made using protocols found in Chapter 2.c.v , and three different plasmid samples were transformed in to determine the transformation and integration efficiency.

Cell counts are taken at different dilutions to estimate the number of cells in the transformation sample and the number of cell that were successfully transformed and/or integrated. Only a small portion of the competent cell will be successful in transformation and integration. Comparing cell count between the selection and non-selection plate provides an estimation of the efficiency of the process. Competent cells we made the day of the transformations and transformation were performed using aliquots of the same batch of competent cells. It is predicted that at each step; transformation of plasmid, transformation and integration pIT4 plasmid and finally transformation and integration of Gibson assembly reaction. The number of cells that survive selection is expected to reduce with the increasing number of steps.

To reduce the number of plates required, a “drip” assay was developed to estimate colony numbers, where 10 µL of a culture to be sampled is placed near the top of the

agar plate, the plate is then tipped at an angle, allowing the sample to run down the plate. For testing transformation and integration efficiency, a series of dilutions were made. A sample of ach dilution; undiluted, 1/10, 1/100, 1/1000, 1/10000 was used in a drip assay using two plates, one with and one without selection. The number of colonies on the selection plate would indicate the fraction of the transformation sample that had been efficiently transformed and/or integrated.

A pZS(^)45 186*pJ* ClyF(C36G) (HB177) mini-prep was used to test only the efficiency of the plasmid transformation step (Figure 37, A). While the transformation strain did also contain the helper plasmid used to integrate at the 2 *attB* phage attachment site this was not expected to affect transformation of a plasmid which did not contain the 2*attP* sequence. The antibiotic selection plate was a Spec50 plate and there are countable single cells present in the third and fourth dilutions. The third dilution (1/1000) had a cell count of 116 cfu/10 µL and the fourth dilution has a cell count of 11 cfu/10 µL (Figure 37, A). From these cell counts, the undiluted sample is predicted to have 116,000 cfu/10 µL therefore the expected total number of successful transformation from 40 µL of ECC is $1.16x10^7$.

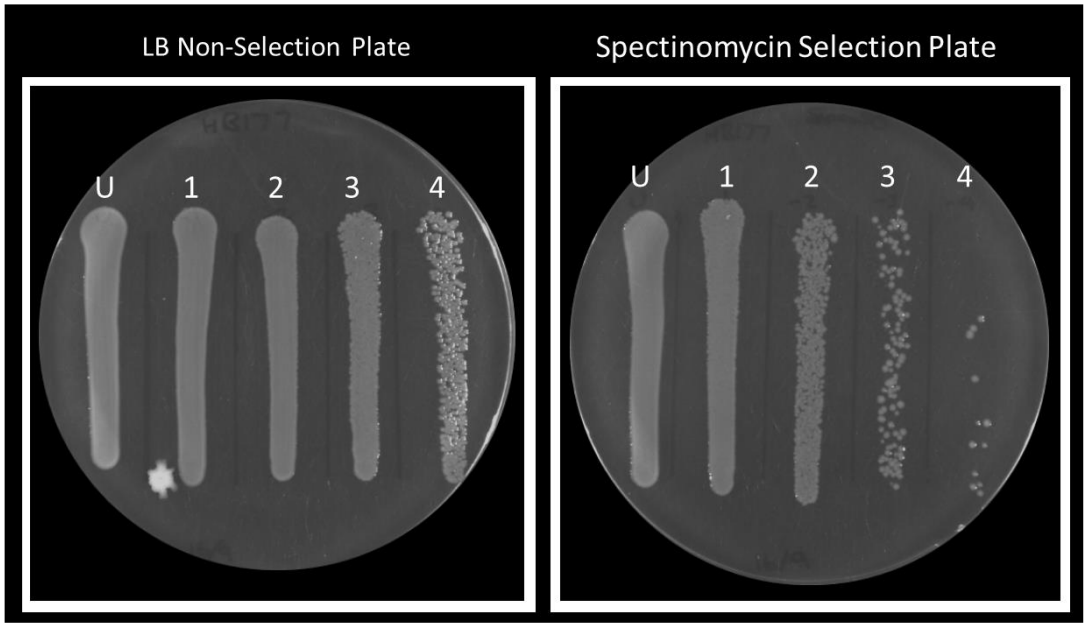
The second plasmid transformed into the transformation strain was pIT4_*attP*2_LoxP_S_186*pJ*_ClyF_*cos*+ (HB184) mini-prep (Figure 37, B). This experiment was to investigate the efficiency of the integration step in addition to the transformation step. The antibiotic selection plate used was Spec20 due to the antibiotic resistance gene being a single copy once integrated. Countable single cells were present in the third and fourth dilutions (Figure 37, B). The third dilution (1/1000) had a cell count of 117 cfu/10 µL and the fourth dilution had a cell count of ~10 cfu/10 µL, with some cell colonies being very small (Figure 37, B). The final cell count for the undiluted sample is 117,000 cfu/10 µL with the expected total number of successful transformation and integrations for 40 µL of ECC is $1.17x10^7$.

50 ng of DNA was used in the pZS(^)45 (Figure 37, A) and pIT4 plasmids (Figure 37, B) transformations. From these results, there appears to be no efficiency lost in the integration step. Every pIT4 plasmid that is successfully transformed into the strain appears to be also successful integrated. Since the pIT4 has a R6Kgamma origin of replication, it is *pir* gene dependent for replication. If it had only transformed the pIT4 plasmid into the cell, it would be lost during cell division. The transformation efficiency for pZS(^)45 (Figure 37, A) and pIT4 (Figure 37, B) was $2.32x10^8$ cfu/µg

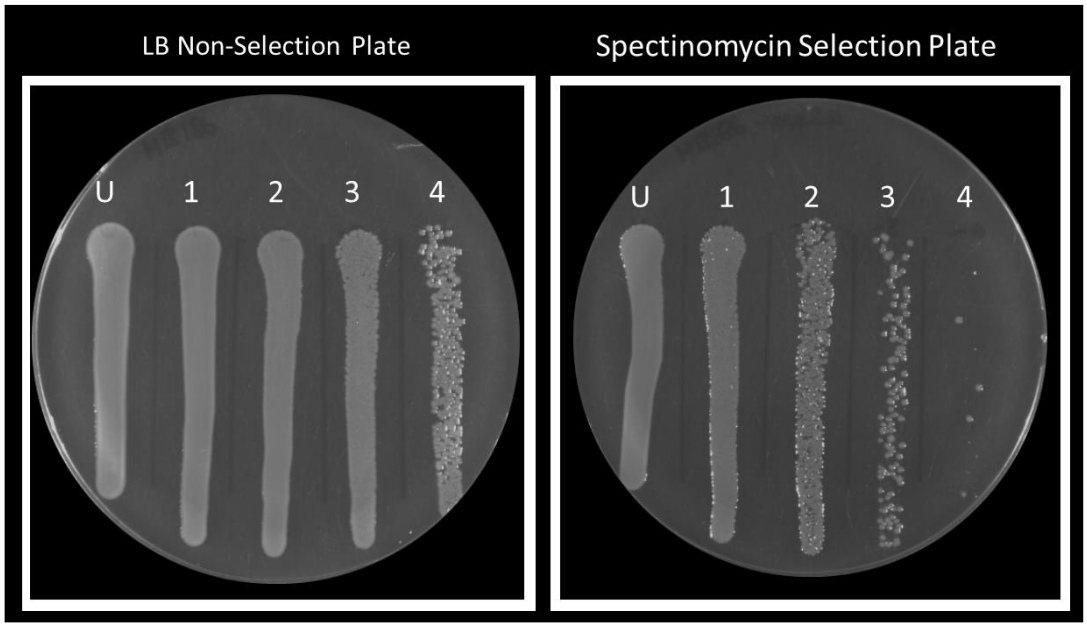
DNA and 2.34x10^8 cfu/μg DNA, respectively. A transformation efficiency of 10^8 cfu/μg DNA for both transformation and integration is at the high end of non-commercial and some commercial protein expression strains (e.g. NEB BL21 Competent *E. coli* Catalog# C2530H). No further optimization was pursued for making electrocompetent cells.

The last transformation was of a Gibson assembly sample, rather than intact plasmid, reflecting the final DNA sample used in the PHEARLESS mutant library screen assay (Figure 37, C). This test examined the transformation and integration steps of the Gibson assembly sample, which contains an unknown amount of assembled plasmid DNA. Unlike the two previous transformations, countable single cells are only present in first (1/10) and second (1/100) dilutions, indicating a loss of transformation and integration efficiency (Figure 37, C). The first dilution had a cell count of 128 cfu/10 μL, 10 μL is the amount for each sample plated on Figure 37 plates. The second dilution had a cell count of 9 cfu/10 μL. Therefore, the total number for the undiluted sample is about 1,280 cfu/10 μL, with the expected number of the total successful transformation and integrations for a 1 mL transformation sample is ~1.28x10^5 cfu/3 μL of assembly mix and 40 μL of ECC. The exact transformation efficiency in cfu/μg DNA cannot be determined, since it is unknown how much plasmid has been assembled from the fragments. In the mutant library assay, the Gibson assembly reaction is purified to remove any salts, going from 20 μL to 10 μL. 3 μL of the cleaned Gibson assembly is transformed into 40 μL of ECC. To improve the number of colonies screened, all of the cleaned Gibson assembly is transformed into three separate transformations. If on average a single transformation sample can produce 1.28x10^5 individual colonies, transforming all of the cleaned Gibson assembly will produces about 3.84x10^5 individual colonies. Therefore, the best solution to deal with the Gibson assembly bottle neck is to increase the number of transformations.

A)


LB Non-Selection Plate
Spectinomycin Selection Plate
U 1 2 3 4
U 1 2 3 4

B)


LB Non-Selection Plate
Spectinomycin Selection Plate
U 1 2 3 4
U 1 2 3 4

C)

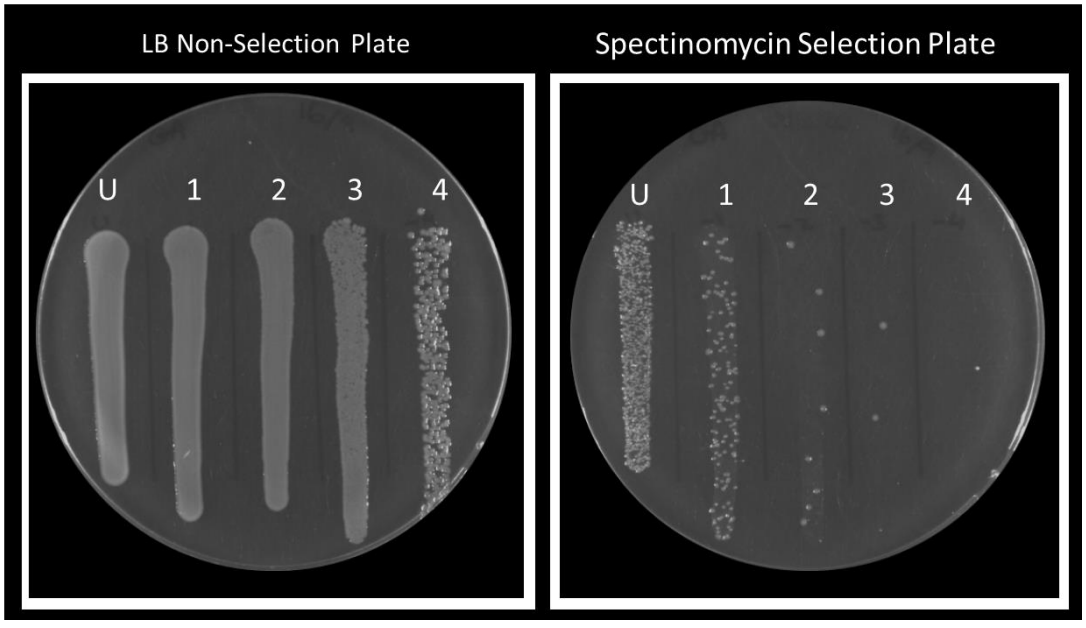

LB Non-Selection Plate
Spectinomycin Selection Plate
U 1 2 3 4
U 1 2 3 4

Figure 37: Transformation and Integration Efficiency Assay for each step. *Each transformation sample was plated onto two plates, one containing spectinomycin (20 µg/µL or 50 µg/µL) and an LB non-selection plate. Each plate contains four samples, the undiluted sample (U), first 1/10 dilution (1), second 1/100 dilution (2), third 1/1000 dilution (3) and the fourth and final dilution 1/10000 (4). A) Transformation Efficiency Assay, Transforming pZS(^)45 186pJ ClyF(C36G) mini-Prep into the Final Expression Strain for the PHEARLESS Protein Library Screening Assay. The selection plate (Spec50) produces single colonies in the third (1/1000) (3) and fourth (1/10000) (4) dilution, 116 cfu/10 µL and 11 cfu/10 µL respectively. The non-selection plate shows the presence of the un-transformed cells with no individual single colonies on the lowest dilution of the plate unlike the selection plate. B) Transformation Efficiency Assay, Transformation and Integration of pIT4 Plasmid Mini-prep into the Final Expression Strain for the PHEARLESS Protein Library Screening Assay. The selection plate (Spec20) produces single colonies in the third (1/1000) (3) and fourth (1/10000) (4) dilution, 117 cfu/10 µL and 10 cfu/10 µL respectively. The non-selection plate shows the presence of the un-transformed cells with no individual single colonies on the lowest dilution of the plate unlike the selection plate. C) Transformation Efficiency Assay, Transformation and Integration of pIT4 Plasmid Gibson Assembly Mix into the Final Expression Strain for the PHEARLESS Protein Library Screening Assay. The selection plate (Spec20) produces single colonies in the first (1/10) (1) and second (1/100) (2) dilution, 128 cfu/10 µL and 9 cfu/10 µL respectively. The non-selection plate shows the presence of the un-transformed cells with no individual single colonies on the lowest dilution of the plate unlike the selection plate.*

4.g.iii Confirming the 10% "Rule of Thumb" of the 186 Genome Size Limitations

Functional ClyF protein was successfully expressed after being integrated into the 186 prophage genome. However, a potential limitation of the system is that as a "Rule of Thumb" most phage are unable to function if the genome is increased by more than 10%. During phage reproduction, the phage genome is packaged into the head portion of the phage, before joining of the head section to the separately assembled tail section of the phage. While the DNA is being packaged, it undergoes DNA cleavage at specific or non-specific locations (177). After the head section has been filled with DNA, termination cleavage of the DNA occurs. DNA cleavage for head packaging is a tightly controlled, too much cleavage will produce non-functioning phage particles, while too little will produces unfilled phage particles (177).

The genomes size of phage 186 is 30,624 bp, therefore the "Rule of Thumb" suggests that the 186 genome could only be increased by ~3000 bp. The final

expression strain for the PHEARLESS V2.2 before adding the integration plasmid is E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ***cos,*** Δ***tum, att2, Cm***)] (HB182). The 186 engineered genome, which does not contain sufficient *cos*+ sequence for production of functioning phage particles, has a net loss of – 1,369 bp. The empty integration plasmid that contains the 186 *cos*+ sequence (pIT4_*attP*2_LoxP_S_186*pJ_cos*+) is 2,757 bp, increasing the 186 genome by a net 1,388 bp. The addition of the ClyF gene to the integration plasmid increases the plasmid size to 3,489 bp. After integration, this results in a net increase of 2,120 bp, still within the “Rule of Thumb”.

The *clyF* gene is 735 bp long, only containing two domains: a catalytic domain and a CWB domain. While this type of domain organization is not uncommon, an arrangement consisting of two catalytic domains, followed by a C terminal CWB domain is far more common in nature. With the addition of a second catalytic domain and two additional linkers the size of a common endolysin gene is around the ~1400 bp giving a net gain of genome size of approximately 2.8 kb.

While the size of the protein in PHEARLESS version 1 is not limiting, being plasmid based, protein size is a potential limit for the PHEARLESS V2.2 expression strain in screening putative proteins for antimicrobial activity.

To determine how accurate the “Rule of Thumb” is in the case of phage 186, three pIT4 integration plasmids were constructed for the purposes of increasing the genome size. Three plasmids were constructed containing ‘Junk’ DNA (*E. coli* FtsK) amplified from plasmid AH1606 to increase the pIT4 ClyF plasmid (pIT4_*attP*2_LoxP_S_186*pJ*_ClyF_*cos*+) by increments of 1 kb (Figure 38). After integrating the plasmids into the final expression strains, the net 186 genome increases were; **+1,369** bp (Empty +1.3 Kb) (E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186 (Δ***Cos,*** Δ***tum,*** 186*pJ, cos*$^{+}$)]) (HB253)), **+2,120** bp (ClyF +2.1 Kb) (E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ***cos,*** ***ΔTum,*** 186*pJ ClyF, cos*$^{+}$])) (HB189)), **+3,120** bp (+3.1 Kb) (E4643(pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ***cos,*** Δ***tum,*** 186*pJ ClyF, 1Kb, cos*$^{+}$)]) (HB265)), **+4,120** bp (+4.1 Kb) (E4643(pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ***cos,*** Δ***tum,*** 186*pJ ClyF, 2Kb, cos*$^{+}$)]) (HB261)) and **+5,120** bp (+5.1 Kb) (E4643(pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ***cos,*** Δ***tum,*** 186*pJ ClyF, 3Kb, cos*$^{+}$)]) (HB263)).

To determine if these expression strains were capable of producing functional phage particles after increasing the 186 genome size, phage functional assays were performed using the phage plating protocols detailed above. Expression strains were grown to OD_{600} 0.6, then diluted to produce ~100 cells per plate. The expression strain was plated using a top layer of 0.7 % agar containing 200 µL of propagation strain, QC024, and 240 µM of cumate to induce lysis. A plate only containing the diluted expression strain was also prepared using a LB agar to estimate the expression strain cell count. The number of single colonies on the LB plate should correspond to the number of plaques present on the QC024 lawn strains, because close to 100% of cells induce under these conditions. Plates were incubated overnight at 37°C and observed the next day for plaque formation.

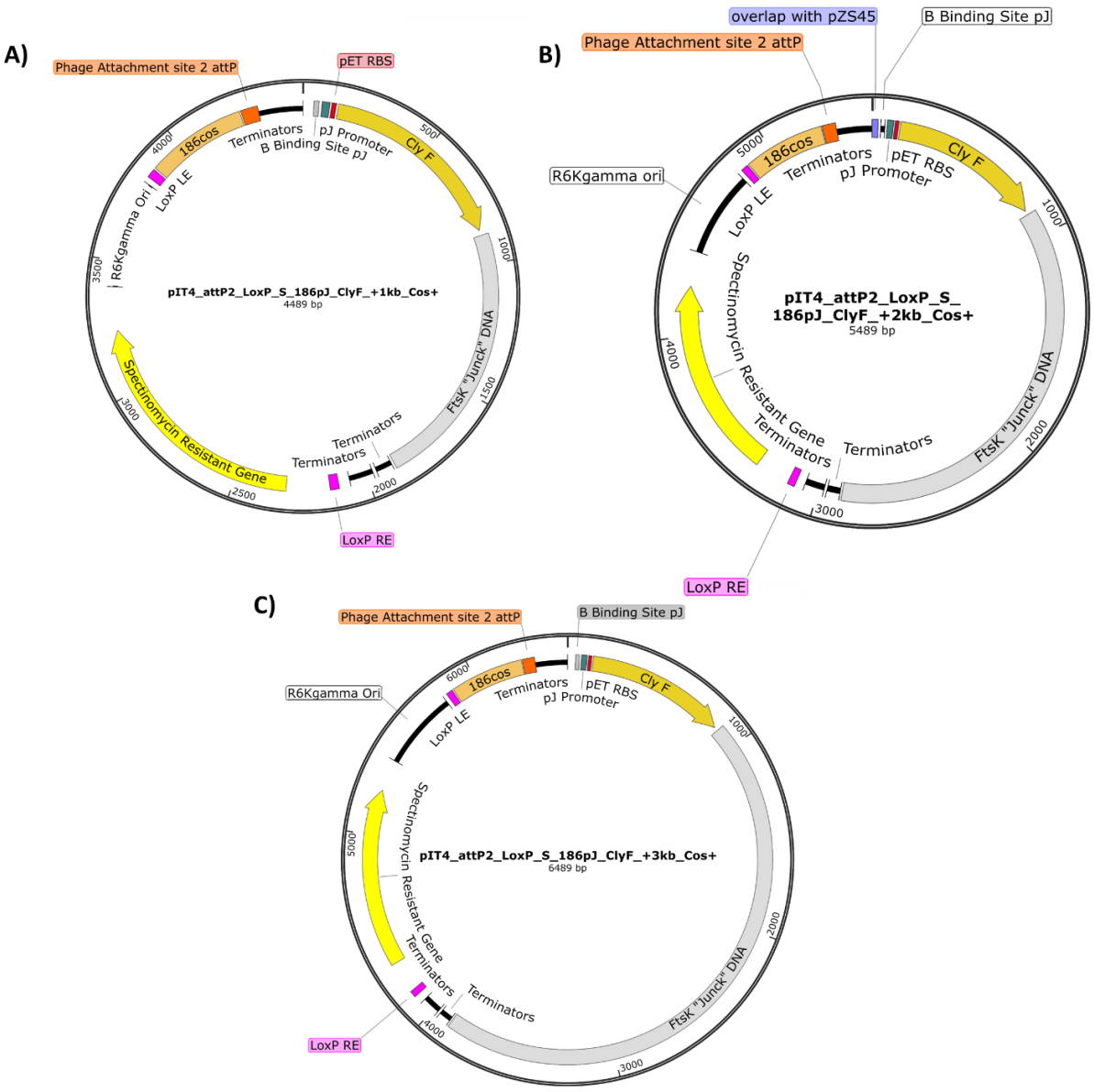


Figure 38: Maps for pIT4 plasmids used to increase the 186 genome size. *The basic pIT4 plasmid contains the same features as the previous pIT4 plasmids that contained the cos sequence (HB184). Plasmid size was increased using sections of a AH1606 gene fragment (FtsK gene). A) pIT4_attP2_LoxP_S_186pJ_ClyF_+1kb_cos+ (HB257) contains a 1 kb FtsK fragment and increased the genome size by 3.1 kb. B) pIT4_attP2_LoxP_S_186pJ_ClyF_+2kb_cos+ contains a 2 kb FtsK fragment and increased the genome size by 4.1 kb. C) pIT4_attP2_LoxP_S_186pJ_ClyF_+3kb_cos+ contains a 3 kb FtsK fragment and increased the genome size by 5.1 kb. Figure was generated using SnapGene.*

Due to the lack of the native 186 *cos* sequence in the final expression strain for the PHEALRESS V2.2 expression strain, the minimum genome change that could be tested was the empty pIT4 plasmid (pIT4_*attP*2_LoxP_S_186*pJ_cos*+). Observation of the plaque formation from the five strains was performed with a minimum of three biological replicates. Plate images for each strain shows that plaque formation numbers between the Empty +1.3 kb and the ClyF +2.1 kb strains are very similar.

The number of expression cells for both strains also matched reasonably well the number of plaques present for both strains. The + 3.1 kb phage strain produces plaques that are approximately half the size of the plaques present on the Empty + 1.3 kb plates. Although the size of the plaques of the + 3.1 kb phage strain is smaller, the number of plaques present is consistent with the cell count present on the LB plate.

The + 4.1 kb phage strain produced very little visible plaque formation. The corresponding cell count plates demonstrated that the dilutions contained the expected number of expression strain cells (~70 cells per plate). While there was a small number (3) of + 4.1 kb phage capable of producing plaques, clearly the majority of the phage produced are non-functional. The last expression strain tested, +5.1 kb, gave no plaques at all (not shown).

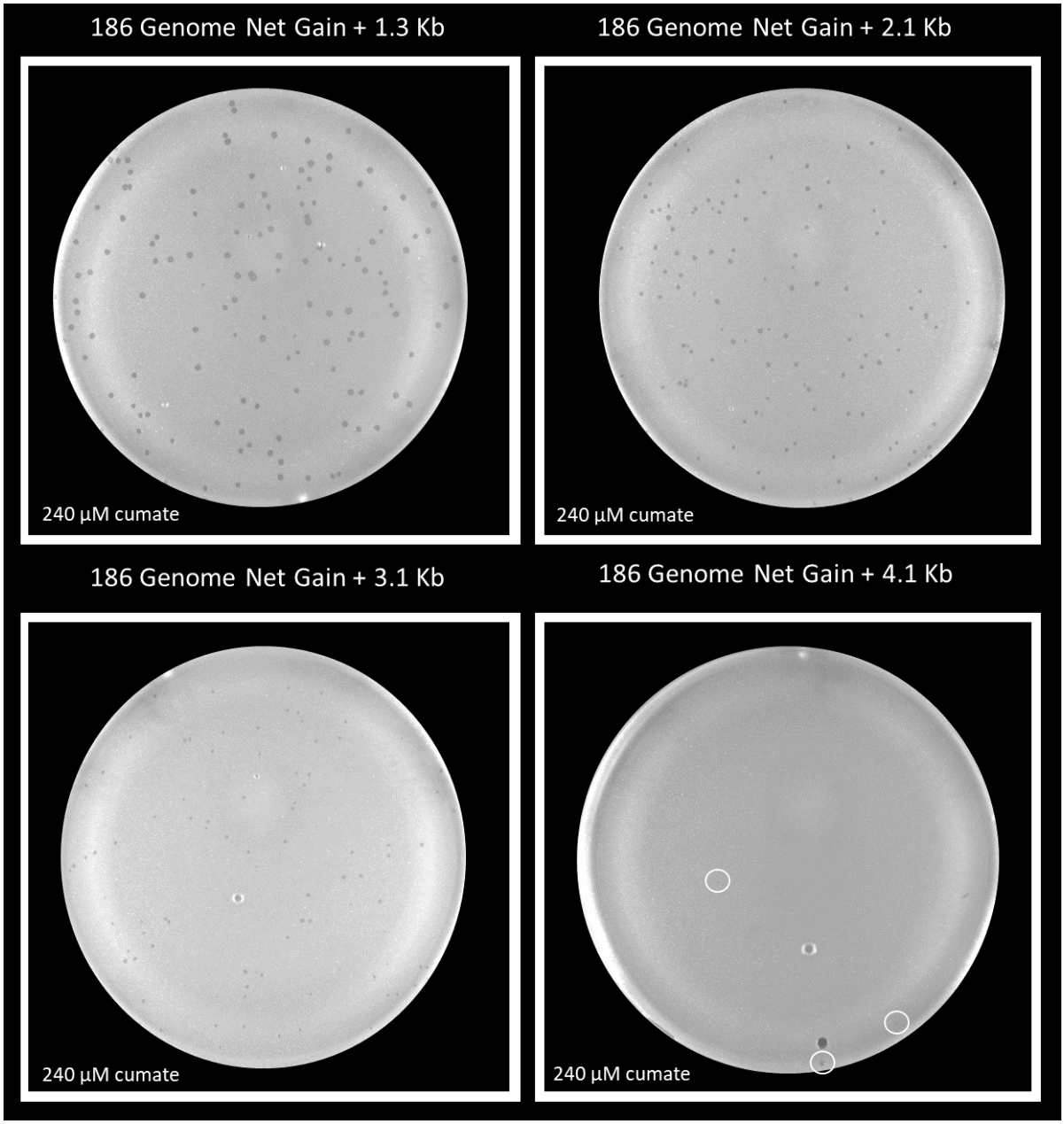


Figure 39: Phage Functional Assay: Testing Phage Propagation for Phage with Increased 186 Genome Size. *Expression strain cells were plated onto a lawn of QC024 to test for the production of functional phage particles, on plates containing 240 µM of cumate. Plaques are present on plates labelled 186 Genome Net Gain +1.3 kb (E4643 [pIT4-KT-cymR-pCym-Tum72]$^{\phi}$21 [186 (Δcos, Δtum, 186pJ, cos+)] (HB253)) and 186 Genome Net Gain +2.3 kb (E4643 [pIT4-KT-cymR-pCym-Tum72]$^{\phi}$21 [186(Δcos, ΔTum, 186pJ ClyF, cos+)] (HB189)). Plaques are visible but smaller in size on the 186 Genome Net Gain +3.3 Kb (E4643(pIT4-KT-cymR-pCym-Tum72] $^{\phi 21}$ [186(Δcos, Δ*t*um, 186pJ ClyF, 1Kb, cos^{+})]) (HB265)) plates. Only three of the plaques from the ~100 single expression cells plated on the final plate 186 Genome Net Gain +4.3 kb (E4643(pIT4-KT-cymR-pCym-Tum72] $^{\phi 21}$ [186(Δcos, Δ*t*um, 186pJ ClyF, 2Kb, cos^{+})]) (HB261)) could be observed. The last expression strain constructed with a +5.1*

*kb in 186 genome size (E4643(pIT4-KT-cymR-pCym-Tum72] ϕ^{21} [186(Δcos, Δ*tum*, 186pJ* ClyF*,* 3Kb*, cos⁺)]]) (HB263)) was not included since it contained no visible plaques.*

It can be concluded from the plaque assay that the "Rule of Thumb" for the genome size, -/+ 10%, is accurate. This does place a limitation on the current 186-based system, limiting it to genes that are smaller than about 3.5 kb. However, there are a number of non-essential 186 genes, which could be removed to increase the size limit for genes of interest (discussed later in Chapter 8.a.ii). While this size limitation will affect phage production and is a consideration when making libraries for screening, it was of interest to know whether the prophage based expression system could be used to test antimicrobial function for individual clones. Undiluted cultures of the five expression strains grown to OD_{600} 0.6 were tested against a lawn of the target strain, RN4220, to test if they would produce sufficient ClyF protein to lyse the target strain, as was the case in the plasmid-based system (Chapter 3: Design and construction of the PHEARLESS system).

Therefore, each of the expression strains containing the additional filler DNA, +3.3 Kb (HB265), +4.3 Kb (HB261) and +5.3 Kb (HB263) were tested for ClyF expression, compared to the ClyF (no filler DNA) control strain (HB189) and the empty (no ClyF) control strain (HB253) on a lawn of RN4220. All five expression strains were grown to OD_{600} 0.6 from overnight culture and 10 µL was spotted onto an agar plate containing a 0.7% agar top layer that contained 200 µL of the RN4220 target strain also grown to OD_{600} 0.6. Since these strains carried the *tum72* gene, only 40 µM of cumate was used to induce cell lysis. 40 µM of cumate were found to produce better ClyF expression and target lawn plaques then the 120 µM of cumate determined in Chapter 4.c PHEARLESS V2.1 Expression Strain Development and Optimization. The empty control strain (HB253) showed no killing of RN4220, unlike the positive control strain (HB189) which showed very obvious clearing. Each of the strains containing the extra "junk" sequence; +3.3 Kb, +4.3 Kb and +5.3 Kb, showed strong active ClyF expression, identical to the results for the ClyF positive control strain (Figure 40). Therefore, this expression strain could still be used for testing antimicrobial activity for genes that are larger than 3.5 kb in size.

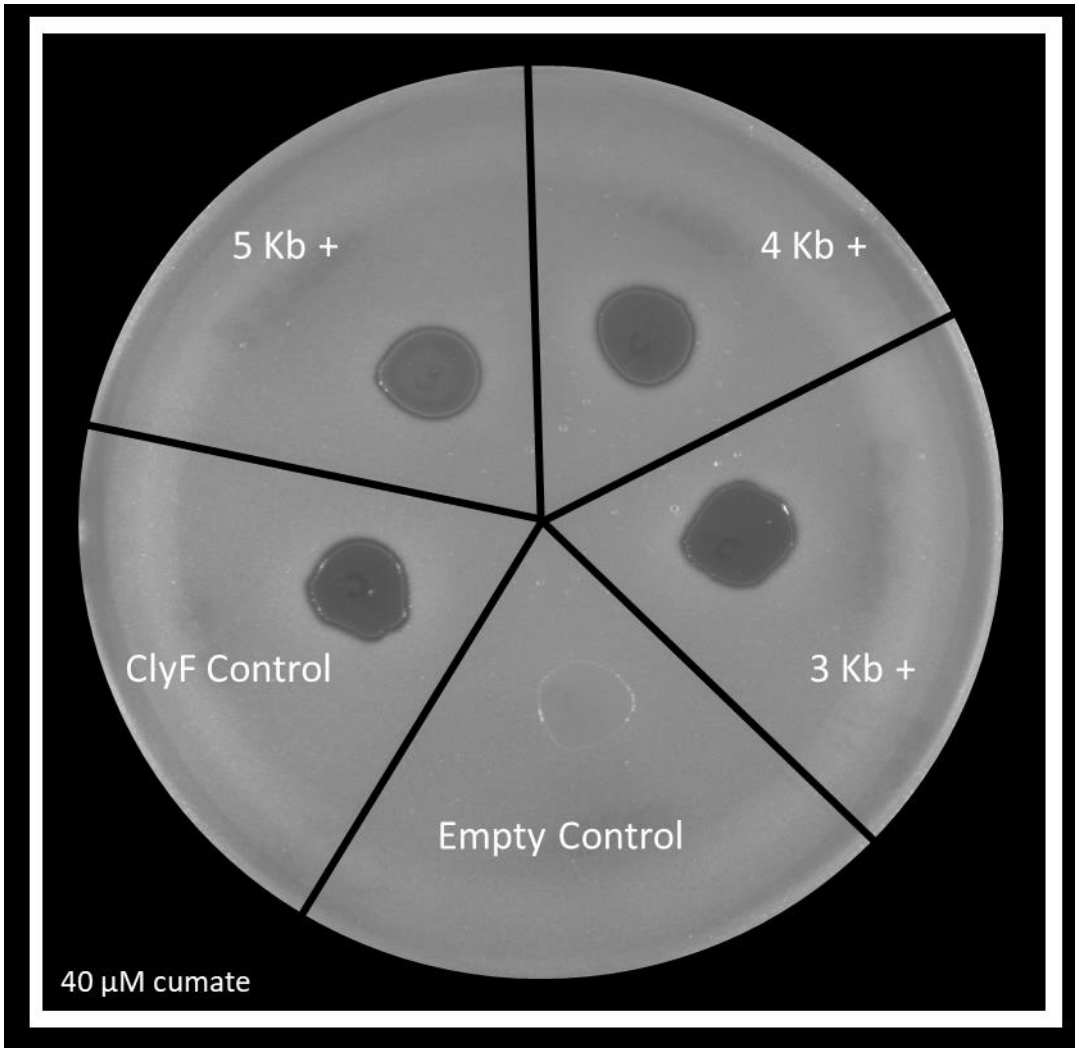


Figure 40: Production of Functional ClyF Protein From Strains With Increased 186 Prophage Genome Size. *Expression strains containing increased 186 genome sizes were tested to determine if ClyF controlled by the 186pJ promoter was still actively expressed. Expression strains were grown to OD600 0.6 and 10 µL spotted on a lawn of RN4220 on a 40 µM cumate plate. The positive control strain used was the ClyF expression strain controlled by the 186pJ promoter (HB189). The negative control used was the empty control strain which carried a pIT4 integration plasmid which lacked the ClyF gene (HB253). Each of the expression strains carried increased genome sizes; +3.3 Kb (HB265), +4.3 Kb (HB261) and +5.3 Kb (HB263) were capable of producing sufficient ClyF protein for clearing of the target strain.*

4.h PHEARLESS Mutant Protein Library Screening using the PHEARLESS V2.2 Expression Strain, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G))

With the improved PHEARLESS V2.2 expression strain, which uses the pIT4 plasmid containing the 186 *cos* sequence (Figure 38) and the *cos*- expression strain (HB186), the requirement for an overnight selection step to select for integrants is removed.

Only cells that have been successfully transformed and integrated with the pIT4 plasmid will produce functioning phage particles able to infect the propagation strain.

This improved approach was used in a proof of principle test to select for muralytic activity from a reverted ClyF gene, a member of an error prone PCR-generated library based on an inactive ClyF(C36G) fragment. Error prone PCR was used to generate a library ClyF(C36G) fragment (782 bp) containing one to four mutations per amplification. Details of error prone PCR mutation rates and calculations for the number of cycles used are found in Chapter 2.c.ii Error Prone PCR.

The PHEARLESS Protein Library Screening started with PCR, using the Phusion PCR protocols and reagents (Chapter 2.b.ii Polymerase Chain Reaction (PCR)) to amplify the plasmid backbone fragments used for the plasmid mutant library assembly. These pIT4 backbone fragments were amplified from the empty version of the pIT4 *cos*+ plasmid (HB194), since it lacks any protein genes that could potentially be carried through the library preparation into the transformation sample.

The ClyF(C36G) fragment was amplified from the pZS(^)45 186*pJ* ClyF(C36G) (HB177) template, noting that if any of this plasmid template carried through to the transformation step, it would not contribute to phage production. The primers designed for mutant library preparation from the ClyF(C36G) template (primers 3219/3220), are located immediately before the start codon and at the stop codon. These primers can be used to amplify the gene of interest located from either the pIT4 plasmid or the pZS(^)45 plasmid. The primers added sufficient overlap to the gene for assembly with the pIT4 backbone fragments which are amplified using primer pairs 3149/3053 and 3054/3150 (Table 3).

Following PCR, the fragments were cleaned up using a Monarch DNA Gel Extraction Kit to exclude the template (Chapter 2.b.iv DNA Clean up). The Phusion PCR generated fragment of the ClyF(C36G) was the template for error prone PCR (Chapter 2.c.ii Error Prone PCR), reducing the chances of template carryover. Error prone PCR was performed for eight rounds, aiming a minimum of two to three mutations per PCR product.

After the error prone PCR fragments were purified and the PCR product concentrations were measured using a Nano-Drop spectrophotometer, the pIT4 *cos*+ backbone fragments and the ClyF(C36G) error prone fragment was assembled using

Gibson assembly (Chapter 2.c.iii Plasmid Fragment Assembly - Gibson assembly). 56 ng of DNA each fragment was added to 15 µL of Gibson Assembly Master Mix in a total of 20 µL and incubated at 50°C for one hour.

Gibson assembly samples were cleaned up and transformed directly into the final expression strain (HB186), which carried the helper plasmid for pIT4 integration (AH6045), pre-induced to express the *att2* integrase. For the most efficient transformation and integration, ECC of the expression strain (HB186) were prepared the same day as the transformation (Chapter 2.c.v Preparing and Transforming Electrocompetent Cells (ECC)). Since the expression strain carries a temperature sensitive helper plasmid (*att*2 helper), cultures are grown to OD_{600} 0.4 at 30°C, before performing a heat induction step at 39°C for 30 minutes with shaking at 100 rpm to express the integrase. 3 µL of Gibson assembly was transformed into 40 µL of expression strain ECC, and left to recover and allow time for integration at 37°C of 90 minutes.

Before plating the transformation samples with the propagation and target strains, a drip assay was performed to confirm the transformation and integration efficiencies. A series of one in ten dilutions was done using the transformation sample. Then 10 µL of the dilutions were added to both a non-selection plate and a selection plate containing 20 µM of Spec (Figure 41). After adding a sample close to the top of a plate, the plate was tilted, allowing the sample to run down the plate. After drying, the plates were placed at 37°C overnight, meanwhile the transformation sample was kept at 4°C overnight for plating the next day.

The next day, the colonies that grew on the selection and non-selection plate were used to calculate the efficiency of the integration. The non-selection plate showed that the seventh dilution ($1/1x10^{-7}$) is the last dilution to give colony growth. The fifth ($1/1x10^{-5}$) and sixth ($1/1x10^{-6}$) dilutions contain samples that are diluted enough to be counted, containing 126 and 11 single cell colonies respectively. Using the cell count for the fifth dilution the number of cells present in the 1mL transformation sample is $1.26x10^{9}$. The selection plate only contains cells in the undiluted and the first (1/10) dilution sample, with a 271 and 37 cell count respectively. Using the undiluted diluted cell count the estimated number of cells that were successfully integrated was 27,100 cfu per 1 mL of transformation sample. Therefore, the percentage of cells that have been both transformed and integrated with the pIT4 plasmid is 0.002% allowing

screening a library of 27,100 clones. The assembly and integration steps are the limiting factors for the library size that can be tested, but could be scaled up. For example, trial was from a single transformation, using approximately one third of the assembly reaction. To increase the library size, the assembly reaction could be scaled up and all transformed into the final expression strain.

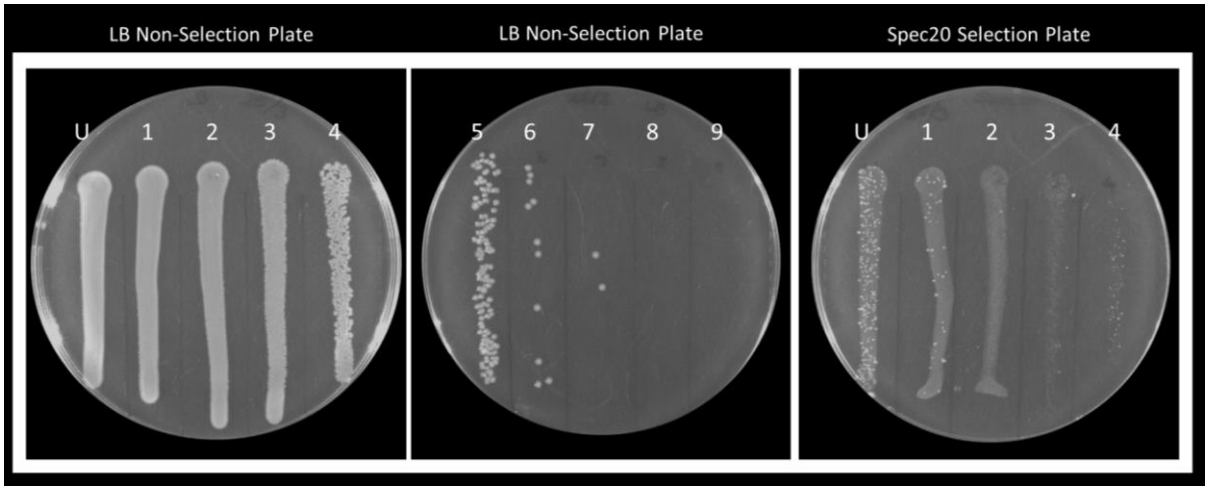


Figure 41: Drip Assay Showing Transformation and Integration Efficiency of the ClyF(C36G) Gibson Assembly Library. *To estimate the efficiency of the mutant library screen assay with the PHEARLESS V2.2 expression strain, a series of one in ten dilutions were made from the transformation sample after the recovery step, and 10 µL plated on a non-selective LB plate, and on an LB plate with 20 µg/µL Spectinomycin (Spec20). The non-selection plate produces countable single colonies in the fifth ($1x10^{-5}$) and sixth dilutions ($1x10^{-6}$), 126 and 11 single colonies, respectively. The Spec20 plate produced viable cell colonies in the undiluted and first dilution ($1x10^{-1}$), 271 and 37 cell count, respectively.*

After performing the drip assay to determine the integration efficiency, the remainder of the transformation sample was plated together with the target strain and the propagation strain, which were at a 50:1 ratio on plates containing 80 µM cumate. Again, the target strain was RN4220 and the propagation strain was QC024. The transformation sample was divided across two plates, each with 500 µL of the transformation mix added to the top 0.7% agar. For a detailed protocol for this experiment refer to Chapter 2.i PHEARLESS Protein Library Screening Protocols. The plates were incubated at 37°C overnight and observed the next day for the presence of plaques, representing possible ClyF revertants. Both plates contained a single plaque (Figure 42). These plaques were collected and purified by re-plating with the target strain and the propagation strain. Single plaques from this purification

step were collected, eluted in 50 µL of phage buffer solution, chloroform treated and stored at 4°C.

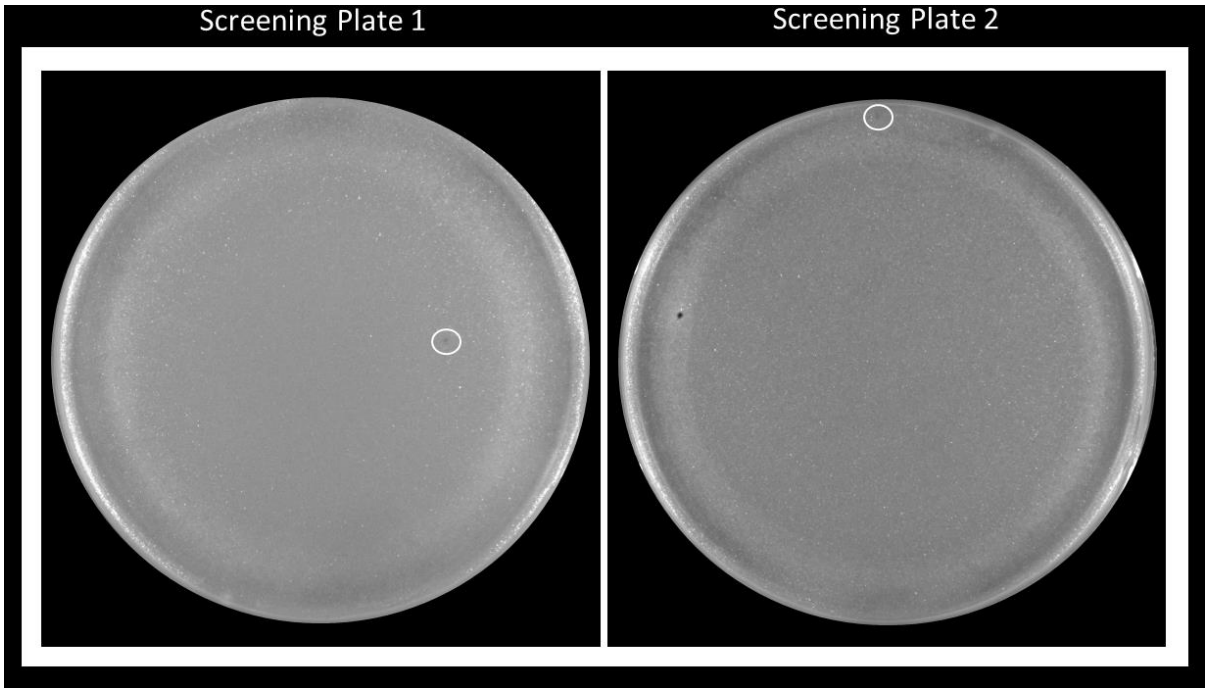


Figure 42: PHEARLESS Mutant Protein Library Screening using the PHEARLESS V2.2 Expression Strain, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G)). *Plates contain the propagation strain (QC024) and the target strain (RN4220) at a 1:50 ratio. The cumate concentrations used for inducing lysis and clyF production was 80 µM. The transformation sample was divided across two plates. Each plate produced one suspected ClyF revertant plaque shown by the white circle. These plaques were isolated using standard methods.*

The purification plates showed that the phage collected from the initial screening plates were capable of producing more plaques (Figure 43). The fluorescent image shows the same number of plaques, where the Td-tomato expressing propagation strain was been lysed (Figure 43). Any contaminating phage carrying inactive ClyF would be able to infect the propagation strain (and be visible in the fluorescent image) but not cause clearing of the target RN4220 strain. This was not the case, indicating good purification of the potential revertants.

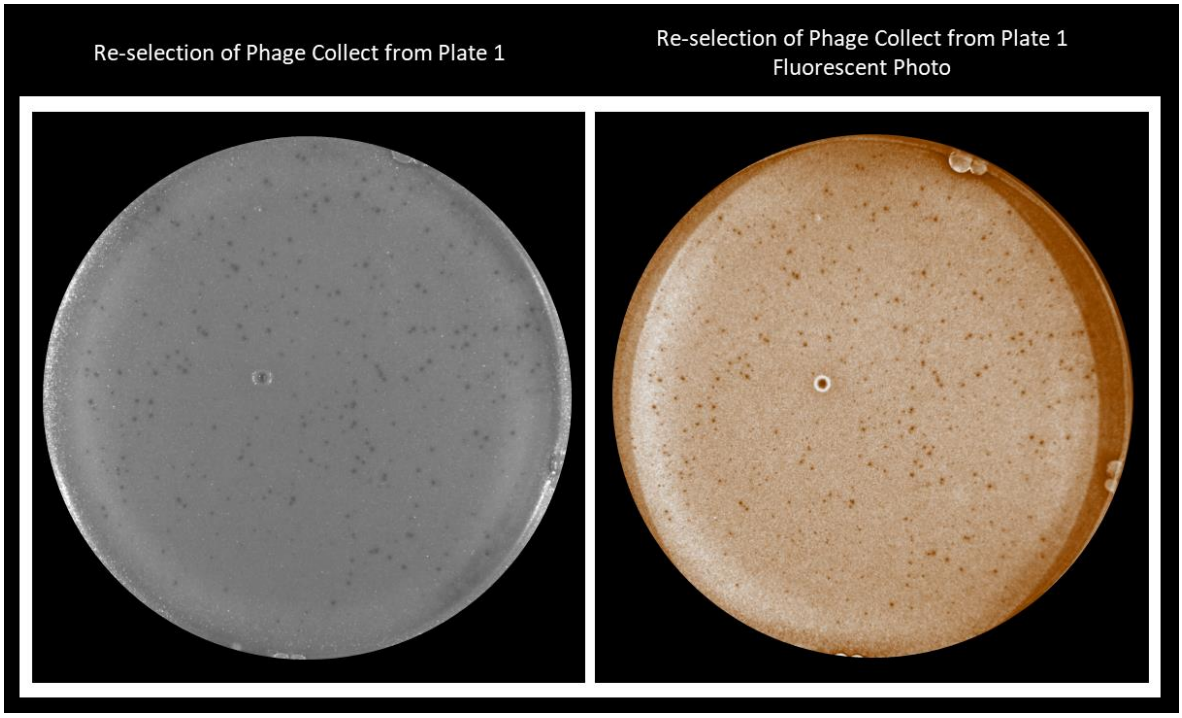


Figure 43: PHEARLESS Mutant Protein Library Screening, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G)), Phage Isolation Plate 1. *Phage plaques, isolated from the initial screening plate, phage were collected and treated with chloroform to kill any bacteria. Phage stocks were diluted and plated onto new lawns of Target: Propagation strain for purification. Grey scale and fluorescent photos were taken to determine if the phage samples were contaminated with non-active ClyF mutant carrying phage, which would be able to form plaques on the propagation strain but not clear the target strain. Both plaque isolates showed no contaminating phage. Representative result for one plaque shown here.*

The final isolated phage samples were used as template to generate PCR products for sequencing using primers 2948/2832, producing a 1kb band covering the ClyF gene (Figure 44, A). The PCR fragments were purified using Monarch PCR and DNA Clean-up kit and sent for Sanger sequencing (Chapter 2.b.iv DNA Clean up and Chapter 2.c.vi Sanger Sequencing)

Sequencing results showed that both phage isolates had the G to T nucleotide revertant mutation, turning the Gly back into a Cys. In addition, phage sample 2 also contained a second silent mutation at the 288 nucleotide, turning the A to a T. This result suggests that each isolate came from a separate pIT4 plasmid assembled and integrated during library preparation. It is theoretically possible however that the ClyF fragment of phage 1 was the template for phage 2 in later steps of the error prone PCR reaction.

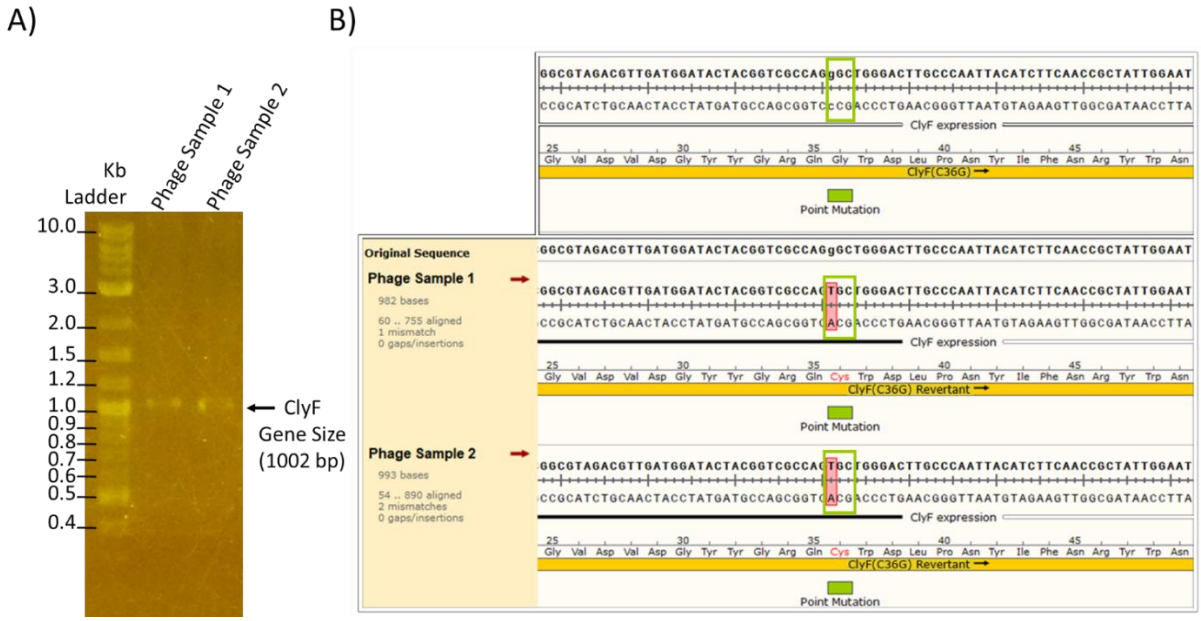


Figure 44: PHEARLESS Mutant Protein Library Screening, PCR amplification of the mutant ClyF gene from phage stocks and Sanger Sequencing results*. Isolated phage exhibiting muralytic activity were used as a template for PCR amplification, amplifying the boundaries of the ClyF gene. A) Both phage contained bands of the expected size for a ClyF gene in that location (1002 bp). B) Shows the Sanger sequencing results of the original ClyF(C36G) clone (top) and the two putative revertants. Each carries a mutation converting the Gly back to the wild type Cys. Phage 2 carried a second silent mutation within ClyF (not shown) Figure was generated using SnapGene DNA sequence alignment.*

4.h.i Repeated and Optimisation of PHEARLESS Mutant Protein Library Screening using the PHEARLESS V2.2 Expression Strain, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G))

To reconfirm that the revertant obtained in the initial PHEARLESS mutant library screening (4.h PHEARLESS Mutant Protein Library Screening using the PHEARLESS V2.2 Expression Strain, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G))) was a true revertant created from the error prone PCR, the assay was repeated. A fresh Phusion-amplified template of ClyF(C36G) was generated and a new error prone PCR library produced from this template. Changes made to the protocols will be listed below.

The sample size was changed in this experiment; the entire Gibson assembly mix was used in four separate transformations of electrocompetent cells. The transformation mixes were plated the same day, following the one-and-a-half-hour recovery step. The plate conditions were identical with propagation to target lawn strain ratios, but the cumate concentration was increased from 80 µM to 240 µM in

order to minimise the number of un-lysed expression cells. Finally, the recovery samples were plated across five plates instead of two to reduce the number of expression cells per plate, thus increasing the proportion of propagation cells available for infection and plaque formation

Out of the four transformation samples, samples one and three produced a putative hit on one of the five screening plates (Figure 45). A control sample, 5, was also taken from transformation sample 2 plate 2. This sample was of a phage that did not give region of target strain clearing and hence should show no revertant mutation. Re-selection was then performed on these three phage samples using the same protocols as before, but using plates with a higher cumate concentrations, 240 µM.

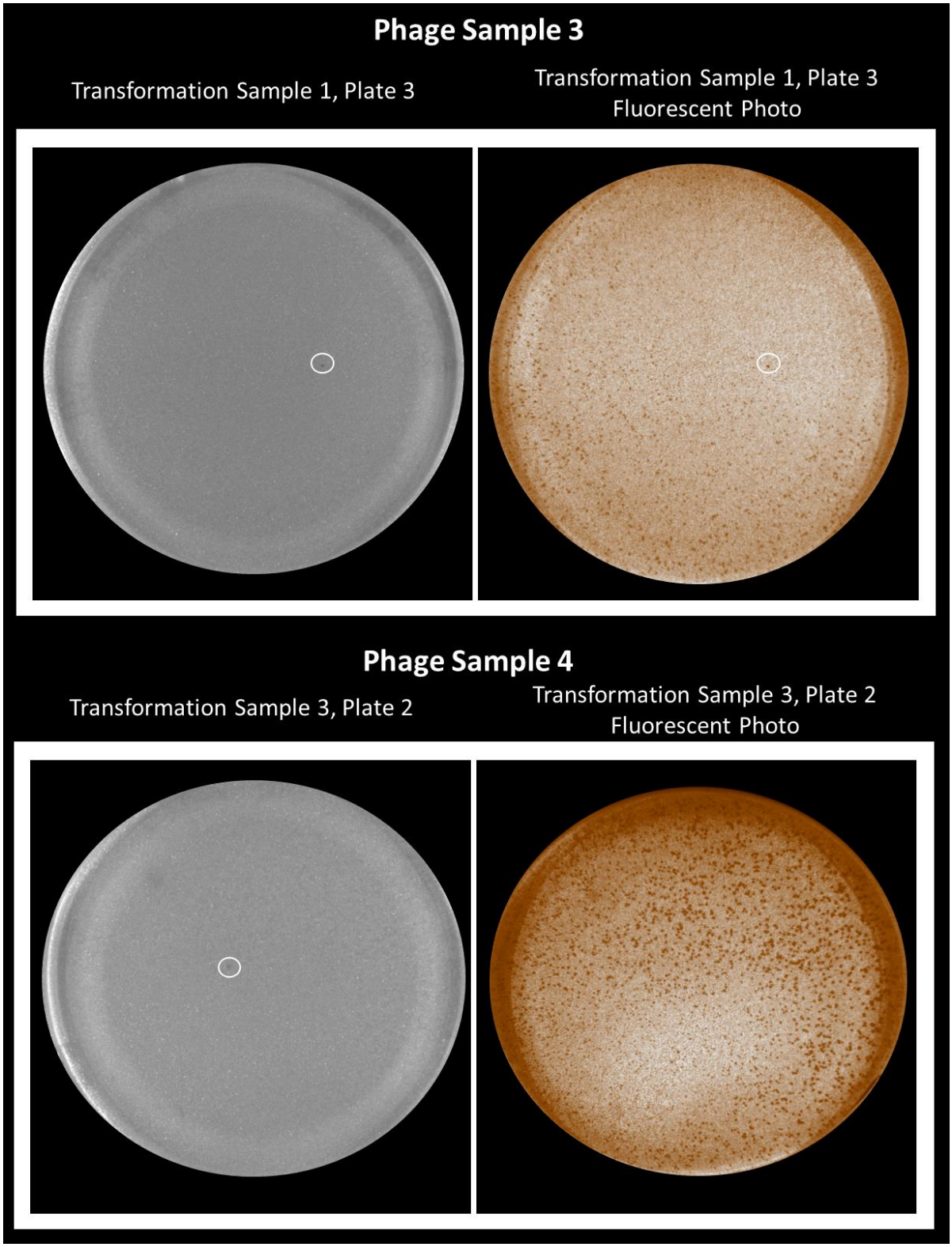


Figure 45: Repeated and Optimised PHEARLESS Mutant Protein Library Screening using the PHEARLESS V2.2 Expression Strain, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G)). *To maximize efficiency, all of the Gibson assembly was transformed into the expression strain (Three Transformations Samples) with the experiments performed in parallel. The transformation samples were divided between five plates contain the target strain (RN4220) and propagation strain (QC024) at a 50:1 ratio, with 240 µM of cumate. Two of the three transformation samples produced plaques (white circles), Transformation 1 (Plate 3) and Transformation 3 (Plate 2).*

Fluorescent photos were taken to assist in indicating how many successfully transformed and integrated expression strains were plated.

Phage samples 3, 4 and 5 were re-selected on new plates to confirm that the plaques observed on the screening plates were reproducible. The phage dilutions used for the re-selected plates were more diluted than expected but still produced between five to fifteen plaques (Figure 46). Due to the dilutions of the re-selection plates no addition re-selections were need to be performed. Phage sample 3 reproduced the plaques that occurred on the initial selection plate. Comparing the grey scale photo to the fluorescent photo shows that were are two plaques present that do not appear on the grey scale photo. This is phage contamination produced from a separate expression cell. Phage sample 5 (Figure 46) as expected did not produces any visible plaques present on the re-selection plate. There are however fluorescent plaques present on the fluorescent photo. Confirming that the sample does contain 186 phage but does not contain an active antimicrobial gene. Phage sample 4 (Figure 46) was determined to be a false positive on the initial screening plate and was unable to reproduces visible plaques on the re-selection plate. But due to the presence of the fluorescent plaques on the re-selection plate, 186 phage is present. Final phage samples for 3 and 4 was taken and PCR for the ClyF(C36G) gene present in the 186 genome. The PCR fragments for phage sample 3 and 4 were then sent in for sanger sequencing.

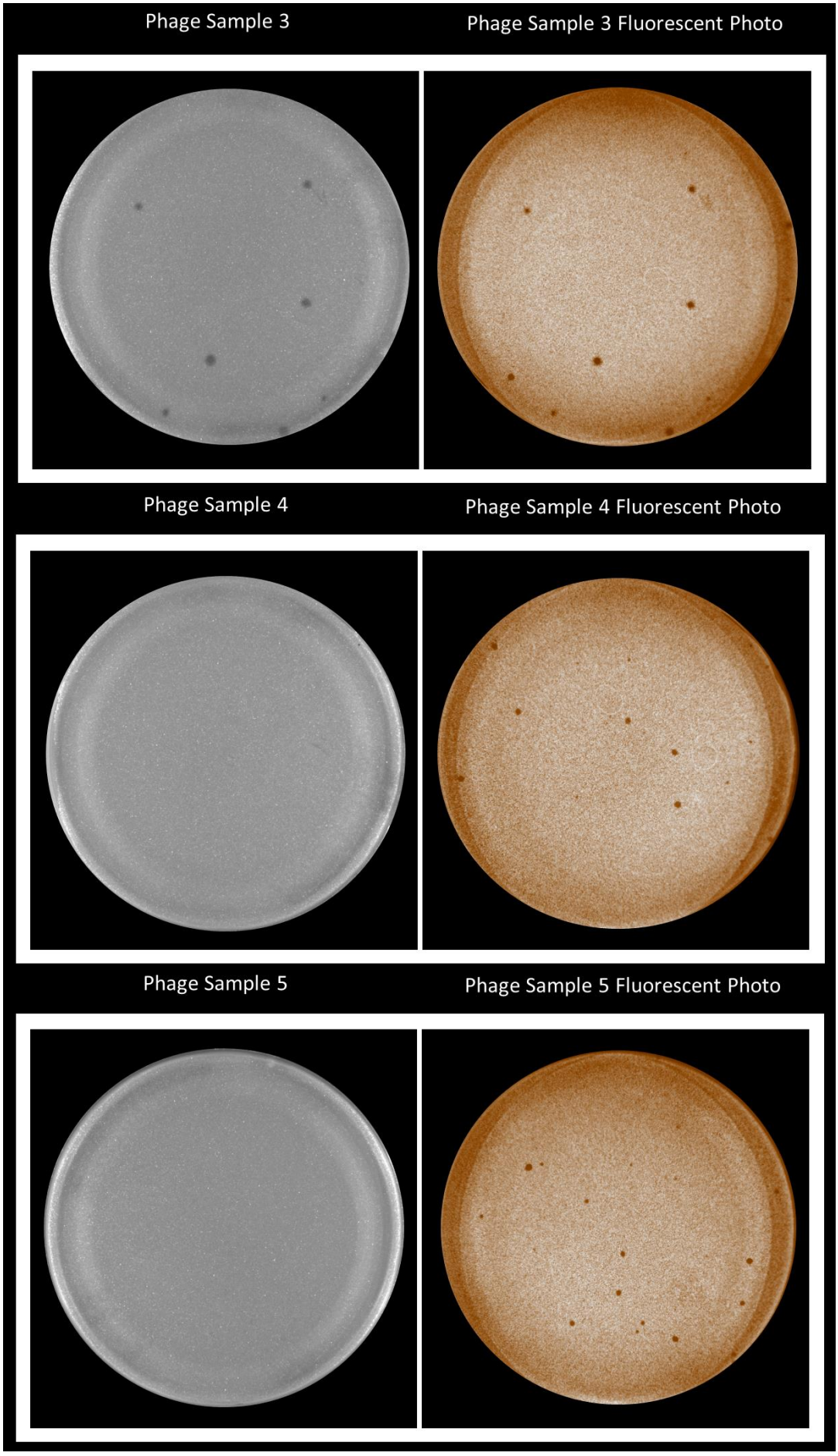

Phage Sample 3
Phage Sample 3 Fluorescent Photo
Phage Sample 4
Phage Sample 4 Fluorescent Photo
Phage Sample 5
Phage Sample 5 Fluorescent Photo

Figure 46: Repeated and Optimised PHEARLESS Mutant Protein Library Screening, Recovery of Muralytic Activity from Inactive ClyF Mutant (ClyF(C36G)), Phage Isolation Plates. *Phage isolated from the initial screening plates were collected and treated with chloroform to remove any bacteria. Phage stocks were diluted and plated onto new lawns of Target: Propagation strain (50:1) to verify the plaques observed on the initial plates were genuine. Grey scale and fluorescent photos were taken to determine if the phage samples were contaminated with non-active ClyF mutant carrying phage. An additional control sample (Phage Sample 5) from a fluorescent plaque which gave no clearing was isolated and reselected. Re-selection of phage sample 3 (top) gave a phage with lytic activity against RN4220. Phage sample 4 re-selection (centre) showed that the plaque observed on the transformation plate was likely a false positive, as no clearing was observed. Phage sample 5 (bottom) was selected as a negative control, unable to give clearing of the target strain.*

Sequencing results show that phage sample 3 does contain a mutation reverting nucleotide G back to T, turning the Gly back into the wild type Cys (Figure 47). Phage sample 3 also contains two more nucleotide mutations. The first was a silent mutation at the 175th amino acid (T 525 C nucleotide change), the second mutation changes amino acid 153 from Lys to Arg (nucleotide change). This is unlikely to disrupt activity since this is a conservative amino acid change. Sequencing of ClyF from phage sample 4 showed sample 4 does not contain any mutation, therefore the plasmid that was assembled included the wild type template fragment in this instance.


Original Sequence
Phage Sample 3
991 bases
41 .. 709 aligned
3 mismatches
0 gaps/insertions
Phage Sample 4
982 bases
35 .. 887 aligned
0 mismatches
0 gaps/insertions
ClyF expression
ClyF(C36G)
ClyF(C36G) Revertant
Point Mutation

Figure 47: Repeated and Optimised PHEARLESS Mutant Protein Library Screening, Sanger Sequencing results. *Phage Samples 3 and 4 were used as a template for PCR amplification, amplifying the boundaries of the ClyF gene. Sequencing of these PCR fragments showed that phage sample 4 that showed no lytic activity had no revertant mutation and no additional mutations. Phage sample 3 contained a revertant mutation, Gly to Cys, which reactivates ClyF activity. Phage sample 3 also contains two additional mutations, Lys153Arg and a silent mutation at the 175 amino acid.*

Summarising, in this chapter, the PHEARLESS system has been successfully adapted for high throughput analysis of a mutant library for proteins with antimicrobial activity. This adapted assay system has two major advantages over current systems used to test antimicrobial proteins and DNA libraries. The system's high throughput ability allows for a large number of clones to be tested in a single assay. The integration of the protein of interest into the 186 prophage provides a simple method of retrieving successful clone genomes.

With the generation of the two PHEARLESS assay systems, one is able to test for antimicrobials and the second to improve the protein's antimicrobial activity and possible host range. The following section is to search for proteins that can be tested in the system for antimicrobial function. The proteins that have been chosen for antimicrobial activity testing are phage proteins, mainly endolysin proteins, that target *S. aureus*. This allows *ClyF* to be used as a positive control in the assays since current protocols allow for three expression strains per plate when testing protein function. The putative proteins have been chosen through bioinformatics searches as well as literature searches. There will also be *S. aureus* phage proteins tested in collaboration with other laboratory groups in a later chapter (6.b Putative Endolysin Genes Provided by the Basil Hetzel Institute).

Chapter 5: Bioinformatic Pipeline for Putative Antimicrobial Phage Protein Discovery

The aim of this bioinformatics study is to search for antimicrobial proteins in bacteriophage, including proteins like endolysin, VALS and structural protein from predicted prophage genomes found in different bacterial *S. aureus* strains. Complete and incomplete *S. aureus* bacterial strains genomes were used in a web-based program called PHASTER to search for prophage genomes. Proteins annotated in the predicted prophage genomes were then formatted and uploaded to European Bioinformatics Institute (EMBL-EBI) (178) bio-sequence analysis tool (HMMSCAN). Domain results were then screened for domains of interest from a list of forty-four domains/phrases that are associated with antimicrobial activity. The proteins that contained one or more of the forty-four domains/phrases were then selected for a final table of predicted antimicrobial bacteriophage proteins.

5.a PHASTER: Web-based Phage Search Tool

PHASTER (**PHA**ge **S**earch **T**ool – **E**nhanced **R**elease) is a web-based server used for the identification and annotation of prophage sequences within bacterial genomes and plasmids (179,180). PHASTER uses BLAST (181) and a clustering algorithm for identifying prophage genomes. GenBank and FASTA formatted genomic sequence data are BLAST searched against a custom prophage/phage database. Phage proteins that either match phage or phage-like sequences that have a BLAST e-values less than 10^{-4} are saved as hits and further evaluated in a clustering program. The custom database was generated using two databases, the National Centre for Biotechnology Information (NCBI) phage database and a prophage database developed by Srividhya (182).

PHASTER uses a number of different search and annotation tools for prophage discovery. Genome-scale ORF predictions and translations are identified via GLIMMER 3.02 (183) which are then processed using BLAST, which aims to provide a predicted identify for the proteins. Additional identifications for common phage features and sequences are searched for using BLAST. RNA sites are also investigated since they can provide additional information on phage features like phage attachment sites, which are calculated using tRNAscan-SE (184) and

ARAGORN (185). After searching for phage-like genes a clustering program called DBSCAN (186) is used to cluster all the phage-like genes into predicted prophage regions. Predicted prophage regions are calculated using two parameters. The first is the cluster size (*n*) which is defined by the minimal number of phage-like genes to form a prophage cluster (187). The second parameter is the distance (*e*) defined by the maximal spatial distance between two genes within a cluster (187). A score is given to each predicted prophage regions out of 150; 150 being 100% complete, score between 70-90 being questionable and a score below 70 being considered incomplete.

PHASTER produces a web page containing the results (Figure 48) and a downloadable zip file of the results. The web pages contain a summary of the prophage regions present with a detailed list of the predicted proteins present in each region, including the CDS position, BLAST hit and E-values. The zip file contains the sequence of the phage region/s as well as a prophage region summary and a detailed list of predicted proteins. The detailed list of proteins was used in the pipeline of antimicrobial protein discovery.

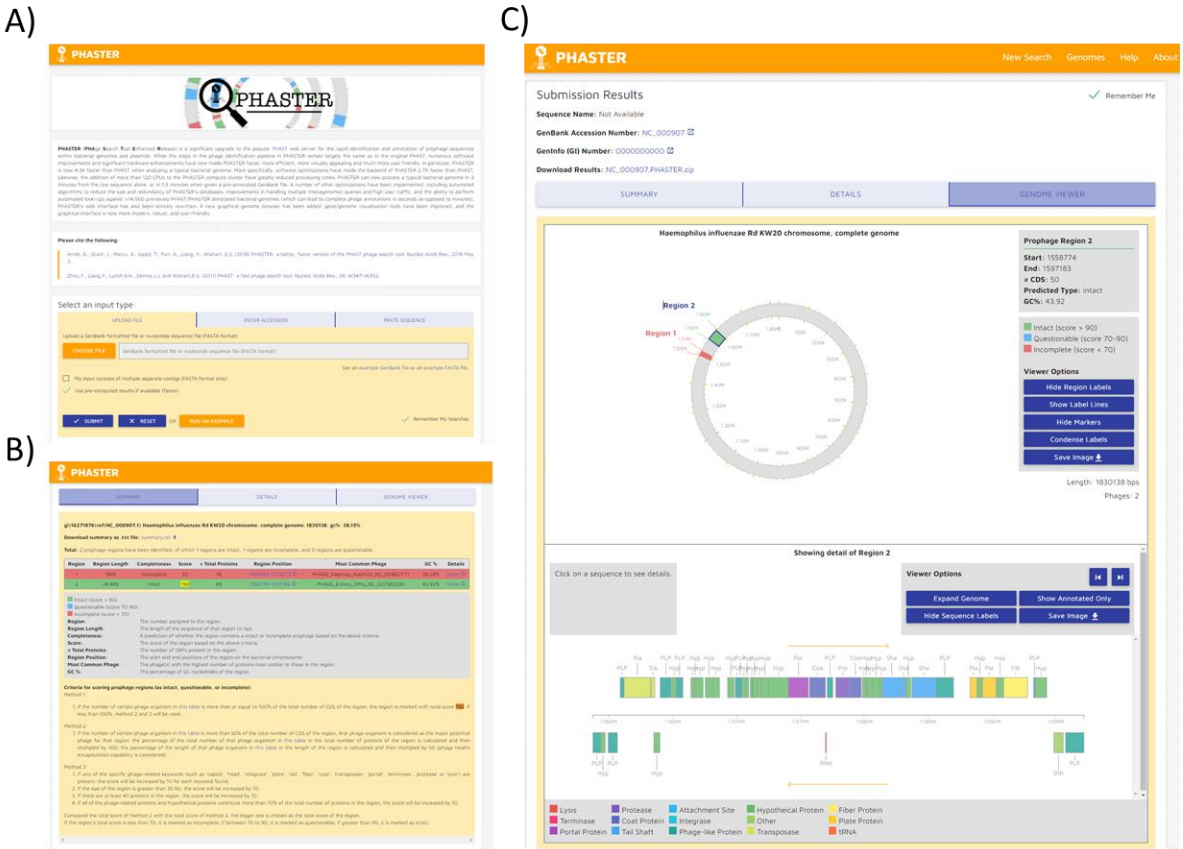


Figure 48: PHASTER (PHAge Search Tool Enhanced Release) *(179,180)****; Web based application for identification and annotation of prophage sequences within bacterial genomes and plasmids****.*

PHASTER interface for uploading bacterial genomes sequences or GenBank accession number. B) PHASTER Summary of predicted prophage, details about the completion score of the predicted prophage, region, length, number of proteins, GC% and most common phage it is related to. C) PHASTER Genome Viewer tab which produces a visual representation of the regions with further details about the predicted prophage protein identified and annotated.

5.b Data Formatting and Domain Search

After retrieving the detail file containing the predicted proteins for each predicted prophage, the list was formatted into eight columns. Column one contained the protein number, column two contains the protein location, columns three to six contains information on the predicted BLAST hit from PHASTER. Each protein entry contained a different number of details from the BLAST results separated using different tab separation. To simplify the results, not all the data from the detail file was included in the new list of proteins. The seventh column contains the protein sequence for the protein and the final column contains the name of the bacteria the phage region originated from.

HMMSCAN is a web-based tool found at EMBL-EBI (178) that is used for searching protein sequences against a database of protein profiles. The profile that was used to search against was the Pfam database generated by the EMBL-EBI. The version updated in May 2020 contained 18,259 different proteins families. The HMMSCAN was run using the default values for the search, E-Values default, 10, which will increase the number of false positives. But the final files will be evaluated separately, removing any protein that has a higher identity to non-antimicrobial proteins (188).

After retrieving the domain search results for each protein, a filter was then used to search for domains that have shown antimicrobial activity. Two versions of the domain list were generated (Table 5), the first list was the initial list and contains domains related more to well-known endolysins domains. This first list was used in the first domain search (Chapter 5.c *S. aureus* Genomes Collected from Patients with Chronic Rhinosinusitis Provided by Basil Hetzel Institute ENT Surgery Group), investigating bacterial *S. aureus* genome contig files provided by the Basil Hetzel Institute from patients with chronic rhinosinusitis (189).

The second list was generated for the second domain search which contain an increased number of domain families, from phage found in many different bacterial

species. The bacterial *S. aureus* genome used in this search were sourced from NCBI and belong to a study that sequenced many *S. aureus* genomes found in patients with cystic fibrosis lung infections (190) (Chapter 5.d *S. aureus* Genomes Collected from Patients with Fibrosis Lung Infections). Protein domain search results that contained a match for one or more of the words in the list were organized into a file for positive results. The protein domain results present in this positive result file were then organised into a table. The table contained the original information about the origin of the predicted protein, the protein sequence and PHASTER prediction. Each protein was re-investigated and details about the protein domains discovered were included in the table. A further parameter recorded for each protein was its presence in additional strains, searching for 100% identity. This showed that a large number of the proteins existed in additional prophage that were detected in PHASTER.

Table 5: Keyword Domain Name List from the domain search script. *Domain word is the key word used to search the results for the hidden Markov Models domain predictions. Two lists were generated a First and Second, with the Second Version containing a more extensive list. Definition of why each key word was chosen is included in the table, including any relevant literature reference.*

Domain Key word	**Definition**	**First or Second Version of the Domain Search List**
Amidase	Predicted Catalytic domain (124), family of enzymes that catalyses the hydrolysis of an amide (191).	Both Versions
Lysozyme	Predicted Catalytic domain (124), lysozyme is the name of a large family of lytic enzymes (191).	Both Versions
N-acetylmuramidase	Predicted Catalytic domain (124)	Both Versions
Endolysin	In some cases, HMMSCAN will detail a domain as a predicted endolysin	Both Versions
cell wall binding	In some cases, HMMSCAN will detail a domain as a CWB domain	Both Versions
hydrolases	In some cases, HMMSCAN will detail a domain as a hydrolases domain	Both Versions
peptidoglycan binding	In some cases, HMMSCAN will detail a domain as a peptidoglycan binding domain	Both Versions
CHAP	Predicted Catalytic domain (124), a domain family that is commonly found in endolysins and is predicted to mainly carry a peptidoglycan hydrolysis function (191).	Both Versions
SH3	Predicted Binding domain (124), SH3 is a well-studied, large family of CWB domains found in many types of endolysins.	Both Versions
Peptidase_	Predicted Catalytic domain (124)	Both Versions

Glucosaminidase	Predicted Catalytic domain (191), family of enzymes that hydrolyse the glycosidic bond between a carbohydrates and or carbohydrates and non-carbohydrate moiety.	Both Versions
Hydrolase	Common dictation for a catalytic domain where the domain is predicted to contain hydrolase activity/function.	Both Versions
lysis protein	Common dictation for a catalytic domain where the domain is predicted to contain a lytic function.	First Version Only
Lysis	Common dictation for a catalytic domain where the domain is predicted to contain a lytic function.	Second Version
Glyco_hydro	Predicted Catalytic domain (124)	Second Version
Transglycosylase	Predicted Catalytic domain (124)	Second Version
MltA	A domain type predicted to involved in peptidoglycan binding, original found in a murein degrading transglycosylase enzyme (191).	Second Version
PG_binding	Predicted Binding domain (124), In some cases, HMMSCAN will detail a domain as a PG binding domain.	Second Version
NAGLU	Domain belongs to the glycoside hydrolase family 89 which function as a glycoside hydrolases (191).	Second Version
GcnA	Annotation for a N-acetyl-β-D-glucosaminidase, included to identify Glucosaminidase domains (191).	Second Version
NAGidase	Domain that belongs to family that has beta-N-acetylglucosaminidase activity (191).	Second Version
WW_like	Predicted Catalytic domain, domain that belongs to a peptidoglycan hydrolase domain family (191).	Second Version
LytB_WW	Predicted domain that belongs to a Endo-beta-N-acetylglucosaminidase domain that is predicted to have a peptide binding function (191).	Second Version
LysM	Predicted Binding domain (124), the domain is predicted to bind to extracellular polysaccharides such as peptidoglycan (191).	Second Version
VanY	Predicted Catalytic domain (124)	Second Version
Dockerin_1	Belongs to EF_hand family, family includes domains that act as CWB domains (123,191).	Second Version
YkuD	Predicted Catalytic domain (124),	Second Version
CW_binding	HMMSCAN and PFAM annotation for CWB domain (191).	Second Version
NLPC_P60	Predicted Catalytic domain (124)	Second Version
ChW	Predicted Binding domain (124)	Second Version
CW_7	Predicted Catalytic domain, domain was original found in the C terminal moiety of a lysozyme encoded by the Streptococcus pneumoniae bacteriophage Cp-7 (191).	Second Version

LGFP	Predicted Binding domain motif, predicted to increase the efficiency of catalytic domain by directing it towards the target (124).	Second Version
SPOR	Predicted Binding domain motif, predicted to increase the efficiency of catalytic domain by directing it towards the target (124).	Second Version
SLAP	Predicted Binding domain motif, predicted to increase the efficiency of catalytic domain by directing it towards the target (124).	Second Version

5.c *S. aureus* Genomes Collected from Patients with Chronic Rhinosinusitis Provided by Basil Hetzel Institute ENT Surgery Group

The bacterial genomes that were provided by Associate Professor Sarah Vreugde of the ENT surgery group located at the Basil Hetzel Institute (BHI). The bacterial genomes sequenced were collected from patients suffering from chronic rhinosinusitis (189). Chronic rhinosinusitis is inflammation in the mucous membranes that line the sinuses and nasal cavity, caused by either viral or bacterial infection (192).

In the USA, acute and chronic rhinosinusitis affects one in eight adults with 30 million cases annually diagnosed. It is the fifth most common disease treated with antibiotic therapy, costing between ten and thirteen billion US dollar in direct costs per year in USA alone (193). *S. aureus* bacteria was collected from mucosal swabs and tissues collected from patients who underwent surgery to treat their chronic rhinosinusitis. From the mucosal swabs and tissues samples collected, 57 *S. aureus* genomes were sequenced. Each *S. aureus* genome uploaded to PHASTER produced at least one predicted prophage region, including all complete and incomplete predicted prophage.

The most common number of predicted prophage regions for the 57 strains is 3 regions, with 10 being the highest number of predicted prophage regions in a single strain. The total number of proteins screened for domains of interests is 7264. From the 7264 proteins screened, **453** were predicted to contain a domain that either has antimicrobial activity or contains a domain that is predicted to have antimicrobial activity (Table 6).

Table 6: Basel Hetzel Institute - Phage Predictions and Protein database V3*. Date set of all putative antimicrobial proteins identified from putative prophages discovered in S. aureus genomes provided by the ENT surgery group. Including details about the predicted domain and protein sequences. Use*

the provided link to gain access to the data set (https://drive.google.com/drive/folders/1a0wGVRV0VXZD8jTSJS7aJgr0jh66kZEq?usp=sharing).

5.d *S. aureus* Genomes Collected from Patients with Fibrosis Lung Infections

The second bioinformatics search (Table 7) was performed on prophage predicted in *S. aureus* genomes isolated from patients with cystic fibrosis (190). Cystic fibrosis (CF) is one of the most common lethal genetic diseases most commonly in people of Caucasian descent (194). Patients with CF are commonly afflicted with lung infections and inflammation eventually leading to respiratory failure. *S. aureus* is one of the major pathogens commonly found to cause infection in CF patients, as well as *Pseudomonas aeruginosa* and *Haemophilus influenzae* (195). The genomic data including the complete and incomplete assemblies, as well as the raw Illumina reads were made available through NCBI under the BioProject accession number PRJNA480016.

Sixty-five *S. aureus* genomes were sequenced during the study; one isolate is a completed genome while the remaining 64 genomes are incomplete contig files. From the 65 genomes, 63 were found to contain at least one predicted prophage region. 209 predicted phage regions were discovered, producing a predicted protein list of 6473 proteins. From the 6473 predicted proteins, the domain search produced **605** proteins that are predicted to have an antimicrobial ability.

Table 7: Cystic Fibrosis Lung Infections - Phage Predictions and Protein database V2*. Date set of all putative antimicrobial proteins identified from putative prophages discovered in S. aureus genomes isolated from patients with cystic fibrosis* (190)*. Including details about the predicted domain and protein sequences. Use the provided link to gain access to the data set (https://drive.google.com/drive/folders/1a0wGVRV0VXZD8jTSJS7aJgr0jh66kZEq?usp=sharing).*

5.e Domains Detected in the Bioinformatics

In addition to the domains that were searched for, a number of new domains were discovered that could contribute to antimicrobial activity (Table 6, Table 7) were detected. There was also a small number of domains that were found in only one of the searches. Below are the listed domains discovered in the search with a brief explanation of what function each of the domains have.

CHAP Domain

Proteins containing CHAP domains and amidase domains were detected in both bioinformatics searches. The CHAP domain (Pfam: PF05257) belongs to the endopeptidase group of proteins, these types of proteins and domains cleave the peptide bonds located at the crosslinking site between the two side peptide chains (108,196). Due to the diversity present in peptidoglycan polypeptide chains and the interpeptide bridge in different bacteria, endopeptidases particularly CHAP domains are a very diverse group (109). CHAP stands for cysteine, histidine-dependent amidohydrolase/peptidase and it is an extremely common catalytic domain found in bacteriophage. While they are a large family, CHAP domains are largely characterised by the conserved catalytic residues, either a histidine or cysteine residue which acts as a nucleophile (196).

The CHAPs general secondary structure belongs to the α + ß structural class of protein, with the N terminus of the protein consisting of primary of α-helices and the C terminus primary consisting of ß-strands (106). Novel CHAP domains have been shown to requires calcium ion for catalysis, showing that the CHAP domain family is quite broad and requires further study. Further evidence of why CHAP domains require further investigation is that a CHAP domain which contained a calcium binding site was not identified through sequence-based searches (123).

CHAP domains can be found in a number of different phage proteins but they are also commonly found associated with the SH3 CWB domains (196). Which is why CHAP domains are extremely common in phage endolysins that infect Gram-positive bacteria, predominantly in *Streptococcus* spp. and *Staphylococcus* spp. phage (124,197). In both bioinformatics searches a large number (>40) of detected proteins were identified as containing a N terminus CHAP domain. No domains were detected as the C terminus domain, but the proteins are large enough for a potential C terminus CWB domain. In addition, all the CHAP domains from both searches are exactly the same size (251 amino acids). While detected multiple times, only 7 unique protein sequences were detected. Six of these sequences show near identical homology, with one to two amino acid differences between them.

Amidase (N-acetylmuramoyl-L-alanine amidase) Domain

Amidase_3 (Pfam: PF01520) and Amidase_2 (Pfam:PF01510) were detected in both searches, both are classified as having N-acetylmuramoyl-L-alanine amidase activity, belonging to the family of hydrolases (124,198). They degrade the peptidoglycan layer by cleaving the link between N-acetylmuramoyl residues and L-amino acid residues (124). Both amidase domains contain metal-dependent activity, requiring a zinc metal ion for catalysis (124).

Amidase_2 is commonly found as the central catalytic domain in endolysin genes in *Staphylococcus*-like and *Mycobacterium*-like phage (124). Interestingly, the amidase domain commonly found in proteins predicted to have an endolysins function, usually located as a central domain (123,124). Amidase domains found in endolysin genes can produced mixed results. In some instances they can improve or act synergistically with the CHAP/first catalytic domain (123,124), but in other cases removing the amidase can result in increased activity (108,112,199). Secondary structural analysis of amidase_3 domains show they are generally comprised of six β sheets surrounded by five α helices, forming a pocket for zinc binding. Amidase_2 produces similar secondary structure, amidase_2 identified in phage lysin LysGH15 demonstrated a recessed area type structure formed by aββαaββαβααααα topology (123). The α helices form the top and the β form the bottom of the recessed area which contained the groove for the zinc ion binding site (123). A comparison of other amidase_2 domains demonstrated similar structural folds but low sequence homology. Further investigation into both these amidase (Amidase_2 and _3) domains is needed since there is less characterization compared to domains like the CHAP domain. In both searches the CHAP domain was found to be associated with amidase domain on the proteins predicted to function as endolysins.

Hydrolase_4 Domain

Hydrolase_4 (Pfam: PF12146) also identified as Serine aminopeptidase, S33, is a member of the AB hydrolase family (Pfam: CL0028). It is classified as a catalytic domain found in a wide range of enzymes. AB hydrolase stands for Alpha/Beta Hydrolase fold, a clan of which contains 73 families with the number of domains included at 605 (as of 2022). In both searches only one protein was listed that contained a Hydrolase_4 domain. Both searches also contained no other known domain structures with the Hydrolase_4 domain spanning the length of the domain. A JACKHMMER (200) search of the protein shows that the closest known proteins that

share homology are carboxylesterases, hydrolases that cleave a carboxylic ester group using water, producing a carboxylate and an alcohol molecule (201). The top ten hits for one Hydrolase_4 containing protein in JACKHMMER identified a range of bacterial proteins belonging to the *Staphylococcaceae* family; *Macrococcus lamae, Macrococcus caseolyticus, Staphylococcus auricularis, Staphylococcus microti, Staphylococcus massiliensis* and *Staphylococcus saprophyticus.*

To investigate a possible role that these proteins may have in a phage genome, InterPro web based protein family classification tool was used to identify proteins with similar domain architecture (202). InterPro identified one hundred sixty-two thousand proteins containing this domain architecture. This type of protein was only identified once in the CF search and four times in the BHI search, in phage that have very low intactness scores so it could be possible these are not phage proteins but rather bacterial protein. But since these proteins potentially contains hydrolase activity they were included in the final lists (203).

Peptidase

A number of different peptidase domains were detected in both searches, this section also contained protein domains with similar activity. The peptidase family is an incredibly large enzyme family, with peptidase found in every living organism on the planet (204). Certain peptidase (also called protease, proteinase and endopeptidase) found in phage are enzymes that uses hydrolases activity to cleave the peptide bonds located in the stem peptide and inter-peptide bridges, using water to break down the chemical bonds (204,205). The list of peptidase domains that were detected are sorted into studied and unknown sections. Since the peptidase family is a very broad family it is possible that some of the peptidase and hydrolases proteins discovered are not involved in PG degradation with potentially a different function internally during infection that is currently unknown (204). Of all the peptidase_ type domains discovered and detailed below, the CF search contained 96 hits that only contained a single peptidase domain that was the only identified domains. BHI search only produced 54 hits that contained this domain architecture. A number of different catalytic domains for endolysins have shown to contain different peptidases (124) including Peptidase_M23, which is why any Peptidase_ domains were included in the results.

Peptidase_S9 Domain

A domain that was only present in the CF search (7 hits) that belongs to this family is the domain Peptidase_S9 (Pfam: PF00326). Like Hydrolase_4, the proteins in the CF search that contained a Peptidase_S9 domain didn't contain any other known domains. All the proteins identified had a protein length of 234 amino acids, with the peptidase domain located at the N terminus in intact putative prophages. Therefore, it is a possibility that there is a potential unidentified C terminal domain which could be a novel CWB domain. Peptidase_S9 demonstrates serine-type peptidase activity, using a serine nucleophile that targets the peptide bonds in a polypeptide chain (206). A JACKHMMER (200) search of this domain produces very few proteins, which were classified as a protein with an unknown function. The top result however is a protein classified as a conserved hypothetical phage protein found in a *S. aureus* strain (Strain NCTC 8325).

Peptidase_M20 Domain

Peptidase_M20 (Pfam: PF01546) domains were found in both searches with CF containing a larger number of proteins. However, there were only two protein hits discovered in both searches which appeared in putative prophages, with higher intactness scores in CF and low scores in BHI. In both searches, they were the only identified domain in proteins of about 400 amino acid length spaning the entire protein. Peptidase_M20 belongs to a small clan of enzymes, Peptidase_MH (Pfam: CL0035), this clan contains a family of zinc metallopeptidases. The M20 domain belongs to a family of glutamate carboxypeptidases, which cleaves a C-terminal glutamate residue (207). The proteins found also contain a M20 peptidase dimerization domain, which consists of two alpha helices and four beta strands. InterPro (202) identifies two hundred and fifty-three thousand hits for proteins containing a Peptidase_M20 domains with M20 peptidase dimerization domain identified in the centre of the Peptidase_M20.

Peptidase_M23 Domain

Peptidase_M23 (Pfam: PF01551) is another zinc metallopeptidase, also classified as an endopeptidase which cleaves peptides and polypeptides. This domain belongs to the Hybrid (Pfam: CL0105) clan that is a superfamily that contains proteins that have

a hybrid motif. Endopeptidase is a catalytic domain that can be used to degrade the PG layer by cleaving the peptide bonds between the stem peptide and the inter-peptide bridges. Peptidase_M23 has a single Zn^{2+} active site with the zinc ligands identified His227, Asp231 and His318, demonstrating a HXXXD and HXH motif (208). The secondary structure of Peptidase_M23 identified in a *S. aureus* Peptidase_M23 shows a two-domain beta protein, generated by six beta sheets that contain the active sites (208).

This peptidase domain was discovered multiple times in both searches, 74 times in the CF search and 66 times in the BHI search. The majority of the protein hits identified were between 900 and 2000 amino acids long, each containing two to three domains. Five protein hits were found to be below the 900 amino acid threshold, with all but one containing only the peptidase_M23 domain. The five proteins include predicted prophage genomes that had a range of intactness scores and all but one was predicted as a tail length tape-measure protein. The proteins above the 900 amino acid length range all had two or three domains identified, a Peptidase_M23 domain located near the C terminal and either a PhageMin_Tail and/or SLT, with the PhageMin_Tail located at the N terminal of the protein and the Peptidase_M23 and SLT domains located at the C terminal. Most of these proteins were classified as phage tail tape measure protein by PHASTER prediction. From this data, protein Peptidase_M23 domain appears to be a common peptidase domain found on the phage tail tape measure proteins which may have a role in breaking through the PG layer of a bacteria to injects it DNA (209,210).

Peptidase_M23 has been identified as a catalytic domain found in phage endolysins the belong to a range of bacterial genera (124,211). CwlP is found in a *Bacillus subtilis* SP-β phage that contains very similar domain architecture, a N terminal PhageMin_Tail and a Peptidase_M23 and SLT domains and SLT domains located at the C terminal. However, the CwlP has the Peptidase_M23 and SLT domains swapped, with the Peptidase_M23 located closer to the C terminal (212). Functional studies performed on CwlP demonstrated that both the SLT domain and the peptidase_M23 hydrolase have activity, capable of breaking domain the *Bacillus subtilis* cell wall (212). Peptidase_M23 in the SP-β phage requires the presence of Zn^{2+}, digesting the D-Ala-diaminopimelic acid (A2pm) bond present in the cross-linkage (212).

Peptidase_M50B Domain

Peptidase_M50B (Pfam: PF13398) belongs to a clan, Peptidase_MA (Pfam: CL0126), of zinc-dependent metallopeptidases that contain the HEXXH motif. This domain appears is both searches but only one to two times. All three hits are identified in predicted prophage that PHASTER classifies as questionable. The proteins are not very large, 59 amino acids in length and the peptidase_M50B appears to cover the entire protein. It has not been thoroughly investigated with it only being noted as that it is found in bacteria and plants. There is one more domain located on the proteins that carried the Peptidase_M50B domain, SdpI (Pfam: PF13630). Sdpl belongs to the SdpI/YfhL protein family which is an integral membrane protein which is noted to protect a toxin-producing cell from being killed (213).

Peptidase_M78 Domain

Peptidase_M78 (Pfam: PF06114) also belongs to the Peptidase_MA (Pfam: CL0126) clan, also identified as IrrE N-terminal-like domain. This domain appears nineteen times in the BHI search and thirty-eight times in the CF search. Peptidase M78 appears on small proteins that are between ~150 to ~200 amino acids in length, with no other domains present. PHASTER predicts the majority of these proteins as unknown proteins, but some are identified as; ribosomal protein, DNA topoisomerase medium subunit and toxin-antitoxin system. There is more protein size and domain coverage variation then the previous peptidase domains but it appears that the domains cover the majority of the protein. Peptidase M78 is summarised as a IrrE N-terminal-like domain, JACKHMMER (200) of one of the peptidase M78 proteins produces a majority of uncharacterized proteins, with the majority of the results having originated from different staphylococcus phage.

Peptidase M78 domain family contains a metallopeptidase which contains a zinc-binding motif (HEXXH). This domain has been identified in a conjugative transposon protein called ImmA (214), suggesting that the peptidase M78 identified in this bioinformatic search could play a role in mobile gene element organization instead of muralytic activity.

Peptidase_S24 Domain

The Peptidase_S24 (Pfam: PF00717) domain was another commonly appearing protein domain (>50 in each search), in most cases accompanied by an N terminal helix-turn-helix (HTH_3 or HTH_19). Majority of the proteins containing these domains had a protein size of ~200 amino acids which were often predicted as unknown, putative repressors and cI repressor like proteins. The peptidase domain in all the protein hits is located at the C terminal with the proteins containing the HTH_3 domains located at the N terminus. Peptidase S24 also known as peptidase S24-like is part of the Peptidase_SF (Pfam: CL0299) clan which only contains two more members, with S24 being not well studied (215). It is noted that the S24 domains all contain a C terminal helix-turn-helix domain which is predicted to be important for DNA binding (216). It is more probable that these proteins are involved in DNA regulation of some kind due to the presence of the C terminal helix-turn-helix domain, known for being important for DNA binding (217).

Peptidase_S49_N Domain and Peptidase_S41

Both Peptidase_S49_N (Pfam: PF08496) and Peptidase_S41 (Pfam: PF03572) belong to the same clan, ClpP_crotonase (Pfam: CL0127). Peptidase_S49_N (Pfam: PF08496) appears regularly in both searches while Peptidase_S41 (Pfam: PF03572) only appear once in the CP search. Peptidase_S49_N appears regularly in both searches, 30 hits in the BHI search and 43 in the CF search. While this domain commonly appears in both searches, it is present in a range of predicted prophage genomes from incomplete, questionable and intact prophage.

The majority of the proteins are listed as unknown proteins, with three proteins being predicted as phi PVL ORF 17 homologue in the BHI search. While appearing multiple times there are also only three different sequences in each search, showing that the protein sequence is most likely tightly conserved. The protein sequences are either 99 or 98 amino acids long, with only the Peptidase_S49_N domain present covering the entire protein sequence. The only Peptidase_S41 (Pfam: PF03572) detected is much larger at about 500 amino acids in length and also contains a C terminal PG_binding_1 domain (Pfam: PF01471). This PG binding domain is a member of the PGBD (Pfam: CL0244) clan which contains two more PG binding domains, 2 and 3. PG_binding_1 was found to act as a CWB domain for a Streptomyces phage

endolysin (218). While the single protein containing a Peptidase_S41 domain also carries a PG_binding_1 domain which suggests it could potentially have muralytic activity. The protein is only discovered once and in a prophage genome that is classified as incomplete. Peptidase_S49_N is more likely involved in the maturation of the pro-head of the phage and acts as a phage pro-head protease gene (219).

Trypsin and Trypsin_2 Domain

A set of serine proteases that were detected in both searches was Trypsin (Pfam: PF00089) and Trypsin_2 (Pfam: PF13365), both belonging to the Peptidase_PA (Pfam: CL0124) clan. Serine proteases cleave peptide bonds using a Serine residue which functions as a nucleophilic amino acid at the protein's active site (220). The majority of the proteins detected are predicted to function as a exfoliative toxin A. Exfoliative toxin are extracellular protein commonly found in *S. aureus* which it can gain from bacteriophage by horizontal gene transfer (221). The toxins are released during bacterial cell lysis causing *Staphylococcal* scalded skin syndrome (SSSS). Clinical features of SSSS range from localised to severe blistering, largely affecting neonates and young children (222). Exfoliative toxins are major pathogenic features that can originate from bacteriophage and makes treatment of the bacterial infection difficult.

CLP_protease

Another serine peptidase domain detected in both of the searches was CLP_protease (PF00574) which belongs to the Clp_crotonase (Pfam: CL0127) clan. The average protein length between both searches is between 200 and 250 amino acids with the domain covering the majority of the protein. No other domains are present on the protein. The PHASTER prediction produced a number of different predictions for the proteins carrying this domain, including; unknown, pro-head protease, putative Clp protease, protease/scaffold protein and S14 family endopeptidase ClpP. The JACKHMMER (200) search for one of the proteins produces a large list of significant matches with the first forty-five matches for phage/virus genes, top results for *S. aureus* phage.

Comparing the UniPro (223) entries for a large number of the matches show that all the proteins are classified as unreviewed with protein function inferred from

homology, with their molecular function predicted as “ATP-dependent peptidase activity” and “serine-type endopeptidase activity”. A protein that belongs to the CLP peptidases family is ClpP, a CLP peptidases protein found in *E. coli*. These Clp proteases are constructed of two components: a number of regulatory ATPase component and a proteolytic component (224). Clp proteins in *E. coli* as well as Lon proteins account for approximately 80% of protein degradation (225). This family of proteins is believed to participate in protein quality control in the cell, due to the Clp ATPase components participation in protein degradation, as well as folding and remodelling.

Comparing the function in *E. coli* it is likely that the proteins containing a CLP_protease domain do not have a catalytic function against its bacterial host. However, due to the novelty of the protein it could still be investigated, since it could potentially have a role in the quality control of phage proteins or degradation of host protein. Bacteria do evolve ways to inhibit phage infection that the phage must develop methods to get around.

Glucosaminidase

There are two types of catalytic domains that will target glycosidic bonds between the N-acetylglucosamine and the N-acetylmuramic acid in the PG layer. Both cleaved the glycosidic bonds using hydrolase activity, each cleaving either the left or the right side of the N-acetylmuramic acid. Glucosaminidase (Pfam: PF01832) are well-studied catalytic domains which target the left side of the N-acetylmuramic acid, part of the Lysozyme (Pfam: CL0037) clan, belonging to the Glycoside hydrolase family 73, this domain cleaves the β (1-4) linkage between the glycans using glycoside hydrolases. The Glucosaminidase (Pfam: PF01832) domain was detected in both searches over twenty times, all in addition to an N terminal CHAP domain (except for one protein).

While this bioinformatics domain search was able to detect the Glucosaminidase domain, the PHASTER BLAST search prediction produced a number of results. Many names included “tail tip” or “tail-associated” as well as “cell wall hydrolase”. All the proteins are about 600 amino acids in length, with an N terminal CHAP domain and the C-terminal Glucosaminidase domain. The most likely function of these proteins are phage structural proteins located on the tail tip area, which is in line with

the PHASTER BLAST predictions. Glucosaminidase (Pfam: PF01832) belongs to the glycoside hydrolase family 73, a large family of enzymes that hydrolyse a glycosidic bond between either two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety (226).

Glucosaminidase domains are a known catalytic domain they appear less frequently in staphylococcus-like endolysins (124), and are more likely to be found in structural proteins. Some examples of previous studies of this domain involve it's use as a lytic enzyme (227), or its function as a structural component (228). Estrella and Quinones article (228) on *S. aureus* phage virulence showed that some of the phage investigated required β-N-acetyl glucosamine modification of cell wall teichoic acids for host cell infection. This is significant since it has been observed that modification of cell wall teichoic acid can remove MRSA resistance to β-lactam as well as other antibiotics (229).

Cell Wall Binding Domain

This bioinformatics was originally designed for phage endolysin discovery, but since has been increased to cover different catalytic domains. CWB domains were not the focus of the bioinformatics search, but a few domain key words like "SH3" and "PG_Binding" were still included in the search. The second bioinformatics search on the protein list generated from the CF bacterial genomes included a number of additional CWB domains. Unfortunately, apart from one PG_binding_1 domain (Pfam: PF01471) none of the other CWB domains search for were detected. This is not unexpected since phage that target *S. aureus* strains more commonly contain SH3 CWB domains in their endolysin, as previously mentioned (99,113).

The Src Homology-3 (SH3) CWB domain is a commonly found CWB domain in *S. aureus* phage endolysins and autolysins, belonging to either the SH3_3, SH3_5 and SH3b type (124). The SH3 (CL0010) clan of proteins can be found in a number of other organisms, from prokaryotic to eukaryotic (119). The general secondary structure for SH3 domains, previously described in the Introduction SH3 domain, are made up of five to seven β-strands that are organised as antiparallel β-strands, connected by linkers (120).

These domains are known to favour binding to sequences that are proline rich (230). One of the first characterised SH3 domains was a lysostaphin SH3b domain present

on a ALE-1, a lysostaphin isolated from a *Staphylococcus capitis* EPK1 bacteria (231). Two types of SH3 domains were detected in both searches, SH3_5 (Pfam: PF08460) and SH3_3 (Pfam: PF08239), SH3_3 only being detected twice in the BHI search. It is believed that the SH3 domain binding target is in the majority cases located in the glycine-rich bridges which is one of the reasons they are commonly observed in *Staphylococcal* phage (232). PG_Binding domains were included since proteins that are capable of binding to the PG layer of the bacterial cell would be of interest. The section on the PG_Binding domain detected was investigated above due to its presence on a protein containing a Peptidase_S49_N (Pfam: PF08496) domain.

Potential Domain of Interest

An additional column was added to the table to indicate any domains found in proteins containing domains of interest. While these domains may not be of interest directly as antimicrobial proteins, they can potentially shed some light on the overall function that the protein has.

Prophage_tail and PhageMin_Tail

Two phage tail domains were detected in both searches. Prophage_tail (Pfam: PF06605) and PhageMin_Tail (PF10145), each located at the N terminal of a protein. Prophage_tail (Pfam: PF06605) is actually classified as a peptidase, belonging to the Phage_barrel (CL0504) clan which consists of a variety of bacteriophage structural proteins that consist of beta barrels. A JACKHMMER (200) search produced a large list of significant matches for phage genes. Of the top 80 matches viewed, all are only predicted proteins with no experimental evidence of their function.

In both searches, the proteins that contained the Prophage_tail (Pfam: PF06605) also contained a Lipase_GDSL_2 (Pfam: PF13472) domain. This domain is located at the C terminal domain and belongs to the SGNH_hydrolase (CL0264) clan. This domain belongs to the GDSL-like Lipase/Acylhydrolase family and the domain is predicted to act as a lipase.

GDSL lipases have been found to be abundant in other microbes and plant species. These lipases contain a GDLS motif GxSxxxxG, containing a Ser active site near the N terminus of the domain (233). With reports that GDSL lipases found in *Arabidopsis*

possesses anti-microbial activity, expressed in responses to abiotic, biotic stress and defence against pathogens (234).

The second phage tail domain discovered was a PhageMin_Tail (Pfam: PF10145), a phage-related minor tail protein. This domain is less defined than the Prophage_tail (Pfam: PF06605) domain since it does not even belong to a clan. All the proteins detected in both searches that contained a PhageMin_Tail (Pfam: PF10145) also contained a C Terminal Peptidase_M23 (Pfam: PF01551) domain. A JACKHMMER (Potter et al., 2018) search for the protein containing this domain produces a large list of significant matches for phage genes like PhageMin_Tail (Pfam: PF10145). Also like PhageMin_Tail (Pfam: PF10145) all of the hits are of only predicted proteins without evidence of the function. One protein hit for a Lysostaphin protein found in *S. aureus* (UniProtKB: A0A380EA25) is predicted to have hydrolysis activity that targets the -Gly-|-Gly- bond found in a pentaglycine inter-peptide link in the *Staphylococcal* PG layer. This hydrolysis activity would be performed by the Peptidase_M23 (Pfam: PF01551) domain. In addition to the Peptidase_M23 (Pfam: PF01551) domain, a third domain was detected in about half the proteins discovered; SLT (Pfam: PF01464). SLT (Pfam: PF01464) belongs to the Transglycosylase SLT domain family as part of the Lysozyme (CL0037) clan. The Lysozyme (CL0037) clan is a large clan of proteins found in a large number of organisms including phage, fungi and animals. All proteins found in the clan have a polysaccharides hydrolysing function (235). The amino acid sequence can share little homology but all share a conserved core structure that consists of two helices and a three-stranded beta-sheet, forming the substrate-binding and catalytic cleft (235). As stated above in the Peptidase_M23 section the SLT domain has been shown to demonstrate muramidase activity, digesting the linkage between NAG and NAM (212).

Domain GPW_gp25 (Pfam: PF04965) was detected only once in the BHI search. It was included in the results although it has a very high E value. A JACKHMMER (200) search of the protein displays a list of unknown proteins, all originating from *S. aureus* phage. The domain detected is predicted to belong to the baseplate wedge protein family, with no known clan. Another gp25 protein belonging to the same family is the T4 phage gp25 protein which is structural protein part of the type VI secretion system. It is one of the many components that construct the wedge, an inner section of the baseplate (236). This protein likely has no catalytic activity since it likely a portion of the baseplate, important for sheath assembly and contraction of

the tail for attachment and host cell penetration. It was detected in the search because it shared the highest homology with, “Gene 25-like lysozyme”. This protein was still included in the table due to its value of being an important structural protein. The T4 phage is a Gram-negative *E. coli* virus and the JACKHMMER (200) produced many proteins that were un-identified. Studying and identifying this protein could provide important structural information for baseplate formation in *S. aureus* phage.

A single domain that appeared in the CP search is a Lysozyme_like (Pfam: PF13702) domain that appears on a protein carrying a NLPC_P60 (Pfam: PF00877) domain. The Lysozyme_like (Pfam: PF13702) domain belongs to the large Lysozyme-like superfamily (Lysozyme, Pfam: CL0037). While a number of the protein families belonging to the Lysozyme-like superfamily have been thoroughly investigated it also contains a number of emerging families of proteins that have only recently been discovered. The Lysozyme_like (Pfam: PF13702) domain is one of these. At least one of the proteins belonging to this family has been crystallised, 4HPE (237), a putative cell wall hydrolase from *Clostridioides difficile* a Gram-positive bacteria. The domain attached to the Lysozyme_like domain is NLPC_P60 (Pfam: PF00877), belonging to the Peptidase_CA (CL0125) clan, a family of catalytic domains that target membrane-associated lipoprotein NlpC in *E. coli* (238).

Another protein that only appears once in the CP search is a protein containing two of the same domain, Stap_Strp_tox_C (Pfam: PF02876), belonging to the Ant-toxin_C (CL0386) clan of toxins. This is a small clan containing two members, Stap_Strp_tox_C and the MAP domain family. This domain is predicted to function as bacterial super-antigen toxins found in *S. aureus*. Like exfoliative toxin A listed above this protein was included in the table due to its pathogenic ability. One of the domains recorded was PDZ_2 (Pfam: PF13180) which belongs to a family of peptide binding site domains. This domain was only found in the CP search on two individual proteins, in one protein it appeared twice. On one protein it was located directly before the Peptidase_S41 (Pfam: PF03572) domain and the PG_binding_1 domain (Pfam: PF01471). On the other protein it is located at the C terminal of the protein after Trypsin_2 (Pfam: PF13365) domain.

A domain that was only found on proteins containing Peptidase_S24 (Pfam: PF00717) was the helix-turn-helix domain HTH_3 (Pfam: PF01381), which belongs to a large family of a diverse range of mostly DNA-binding domains. This could

suggest that the proteins containing the Peptidase_S24 (Pfam: PF00717) domains are actually involved in genome regulation due to the HTH_3 (Pfam: PF01381) domain. In addition to HTH_3 (Pfam: PF01381), HTH_19 (Pfam: PF12844) was also detected. Like HTH_3, HTH_19 also belongs to the same super family of DNA-binding domains but is found in a family the contains antitoxins from bacterial toxin-antitoxin systems which are predicted to be DNA binding (239).

Summarising, bioinformatics tools are excellent resources for identifying specific types of proteins. From 13,737 putative phage proteins in this bioinformatic pipeline, the pipeline was able to filter the list down to 1,058 potential antimicrobial proteins. This number can be further reduced by removing proteins that have high sequence homology to each other. These data sets (Table 6, Table 7) will be used as a source of putative antimicrobial phage proteins for testing in the PHEARLESS assay system. With the additional information about the predicted domains provided in the data sets, informed decisions can be made when choosing proteins for investigation. Allowing us the choice to excuse proteins that could require additional proteins to assemble and function or choosing a unique protein that contains domains with low identification scores (Domain E-values scores).

From the combined protein results for both bioinformatic searches, four proteins were chosen and investigated for antimicrobial activity. The four proteins were selected using the criteria of each protein showing medium to low homology with one another, since many of the proteins identified in both searches shared high homology. In-depth literature searches were also performed on each to eliminate proteins that have a high homology to identified phage proteins that have already had functional studies performed.

Chapter 6: Selected Proteins Tested for Antimicrobial Activity

6.a Proteins Investigated using the PHEARLESS System

After completing the optimised versions of the PHEARLESS V1 expression strain (Chapter 3: Design and construction of the PHEARLESS system) and PHEARLESS V2 expression strains (Chapter 4: PHEARLESS V2 Assay) for testing antimicrobial proteins for activity, a small selection of putative proteins were chosen for testing. Seven proteins were chosen from the bioinformatics results presented in Chapter 5: Bioinformatic Pipeline for Putative Antimicrobial Phage Protein Discovery and a third bioinformatics search performed by Oliveira and Sampaio (232). The bioinformatics search aimed to provide insight into the evolutionary relationship and genomic diversity between different phage genomes in a library of 200 *Staphylococcal* phage. In addition to providing a better evolutionary understanding of phage genomes, a dataset of endolysin genes was generated. The broad criteria for selection was that the protein is likely to be derived from a phage, and contains at least one catalytic domain, meaning that phage structural proteins may be included.

A literature search was performed to determine if any of the protein candidates had been studied in terms of catalytic activity. Each of the significant matches above 50% similarity were investigated for evidence of catalytic activity and each of the UniProt (223) hits were investigated for links to functional studies. In many cases the annotation score of the protein was classified as “protein predicted”, as part of a genome annotation article. The molecular function of theses unknown genes are in some cases predicted using the Gene Ontology: tool for the unification of biology (GO) (240,241). The protein name as well as the phage names for each of the protein were investigated.

6.a.i Proteins Discovered in Bioinformatics Searches

Three of the seven protein candidates were chosen from the BHI bioinformatics search (Chapter 5.c *S. aureus* Genomes Collected from Patients with Chronic Rhinosinusitis Provided by Basil Hetzel Institute ENT Surgery Group). One of the proteins (P1) was predicted to function as an endolysin protein due to the presences of CHAP, Amidase and SH3 binding domains, in that order. The second protein (P2)

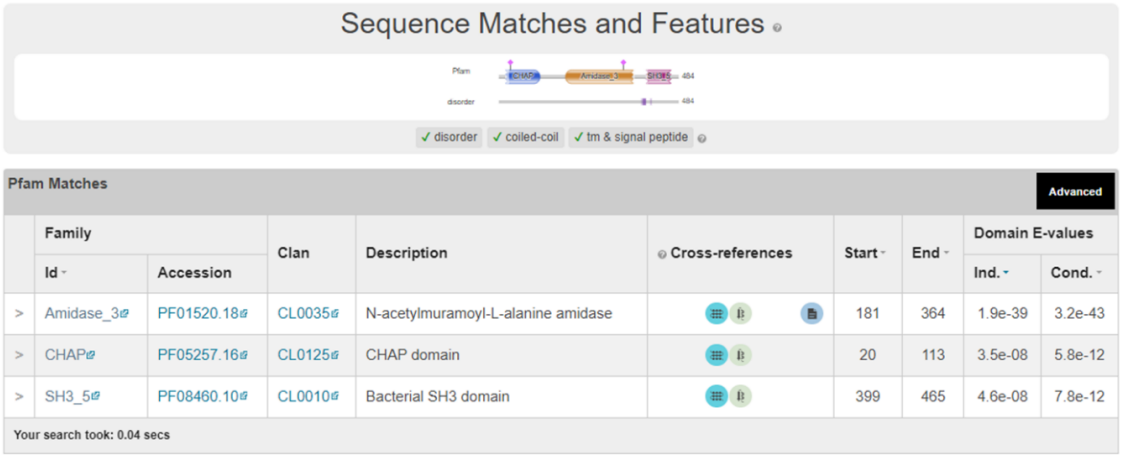
contained a single CHAP domain at the N terminal with no domain detected at the C terminal. The third protein (P3) is predicted to act as a structural protein, located in the tail region due to the presence of a CHAP domain and glucosaminidase domain.

Putative Phage Protein with Predicted Antimicrobial Activity: PE_BHI_P1

The first protein was from the BHI bioinformatics search (Chapter 5.c *S. aureus* Genomes Collected from Patients with Chronic Rhinosinusitis Provided by Basil Hetzel Institute ENT Surgery Group) labelled as PE_BHI_P1 (P1) (Figure 49) (53.86 kDa) was found in an intact predicted phage region (PHASTER score 150) in strain BLAC_32, contig 17. The phage is predicted to be 29.7 Kb in length. The predicted annotated ORFs in this phage all corresponded to known phage genes, with 24/42 found in *Staphylococcus* phage ϕJB (NCBI Reference Sequence: NC_028669.1).

HMMSCAN results (Figure 49. A) show three different domains that are commonly found in endolysins: An N-terminal CHAP domain with domain boundaries from amino acids 20 to 113, a second Amidase_3 (N-acetylmuramoyl-L-alanine amidase) domain, predicted to start at residue 181 and end at the amino acid 364, and a final domain at the C terminal, predicted to be a CWB domain, SH3_5 (Src homology-3 domain), from amino acids 399 to 465.

A)



Pfam Matches — Advanced

	Family		Clan	Description	Cross-references	Start	End	Domain E-values	
	Id	Accession						Ind.	Cond.
>	Amidase_3	PF01520.18	CL0035	N-acetylmuramoyl-L-alanine amidase		181	364	1.9e-39	3.2e-43
>	CHAP	PF05257.16	CL0125	CHAP domain		20	113	3.5e-08	5.8e-12
>	SH3_5	PF08460.10	CL0010	Bacterial SH3 domain		399	465	4.6e-08	7.8e-12

Your search took: 0.04 secs

B)

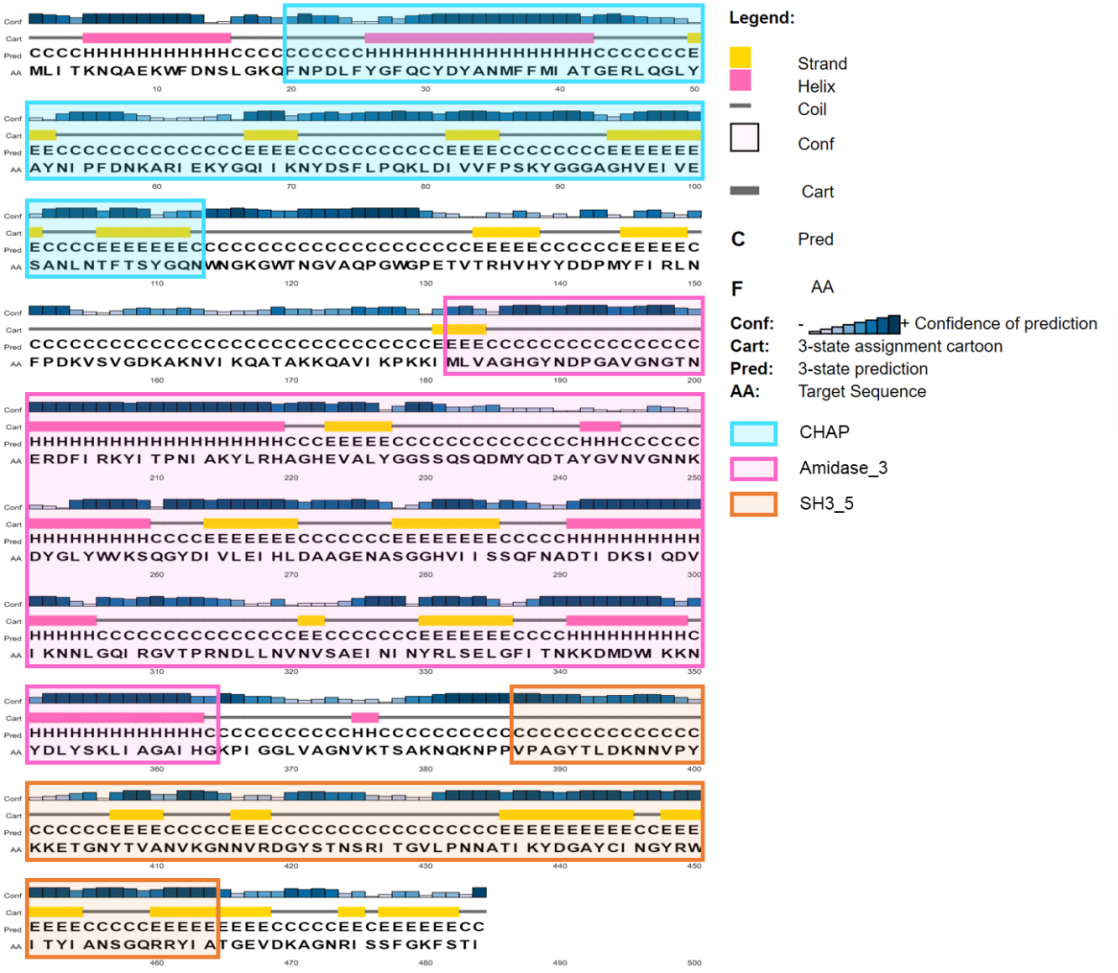


Figure 49: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_BHI_P1, chosen from the BHI bioinformatics search. *A) Shows figure adapted from HMMER: Biosequence analysis using profile hidden Markov Models online web tool (200). HMMER identifies three domains; N terminal CHAP domain (20 – 113 amino acids), Amidase_3 (181 – 364 amino acids) and a C terminal SH3_5 (399 – 465 amino acids). B) Shows figure adapted respectively from PSIPRED protein structure prediction online tool (242). PSIPRED predicts that the CHAP domain contains one α helix followed by five β sheets (α β β β β β), the Amidase_3 follows an β α β α α β β α β β α α*

architecture and the SH3_5 contains five predicted β sheets (β β β β β). PSIPRED also includes a legend with further information about the predicated secondary sequence.

Matches from the JACKHMMER (200) search show hits for proteins identified as amidases from different species of *S. aureus* phage, with the top thirty hits for proteins with an identical % similarity. One of the proteins from the list of top thirty hits have been investigated experimentally.

The P1 protein hit that was investigated was probable autolysin PH (UniProt: PH_STAA8) which was isolated from *S. aureus* NCTC 8325 (NCBI: txid71366). Unlike peptidoglycan hydrolase (UniProt: Q2FYD8 · PH_STAA8) this study was able to purify full length protein, but the protein showed no activity against the *S. aureus* strain it was tested against (SA113) (243). However, a chimeric endolysin protein constructed using the CHAP domain from probable autolysin PH with the CBD from LytM, was constructed and was capable of lysing *S. aureus*.

The problems encountered with protein aggregation at high expression levels are less likely to occur in the PHEARLESS system because expression levels are considerably lower. In addition, since the putative proteins originated from strains sequenced by our collaborators, it is possible to test for activity against the original isolates, in addition to our standard test strain (RN4220). This putative endolysin (P1) was also found in three other putative phage identified in different isolates in the BHI strain; BROCK_27, COUG_54 and FRAN_8.

Putative Phage Protein with Predicted Antimicrobial Activity: PE_BHI_P2

The second protein from the BHI bioinformatics search (Chapter 5.c *S. aureus* Genomes Collected from Patients with Chronic Rhinosinusitis Provided by Basil Hetzel Institute ENT Surgery Group) was labelled PE_BHI_P2 (P2) (Figure 50) (29.06 kDa), this protein is predicted to have catalytic activity due to the presence of the CHAP domain. Whether it acts as a endolysin or a structural protein is unknown. P2 was discovered in an intact predicted phage (PHASTER score 130) found in the ARC1 bacterial strain, contig 10. The predicted phage genome is 51. Kb in length, containing 64 predicted genes, 63 were predicted as phage genes. Half of the genes detected were found in *Staphylococcus* phage tp310-3 (NCBI: txid445517). A gene from this phage was also detected with 100% identity in the JACKHMMER (200) search for the P2 protein.

HMMSCAN results (Figure 50. A) only detects the presence of a CHAP domain located at the N terminal of the gene. Lowering the significance bit scores for the HMMSCAN search to the lowest values available (Sequence: 7, Hit: 5) does not produces any additional domains (Figure 50. A). The C terminal annotated in Figure 50. B is a confirmed CWB domain identified in a protein that has high homology, explained further bellow Figure 50. The secondary structural predictions (Figure 50. B) for this protein by PSIPRED (244) predict that the CHAP domain contains two helix structure followed by four β stand structures.

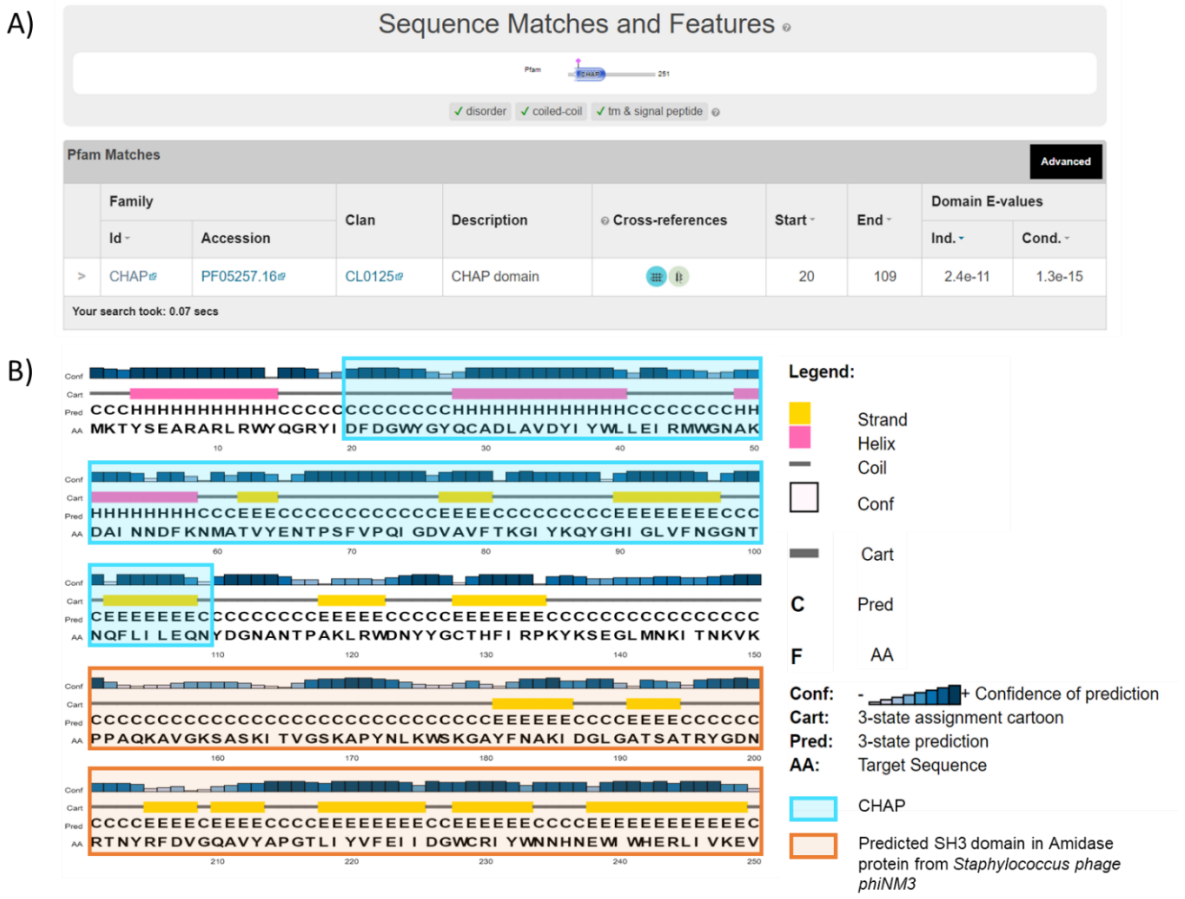


Figure 50: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_BHI_P2, chosen from the BHI bioinformatics search. *A) Shows figure adapted from HMMER: Biosequence analysis using profile hidden Markov Models online web tool (200). HMMER identifies only identified one domain, a CHAP domain located at the N terminal (20 - 109). PE_BHI_P2 has a C terminal, a SH3 domain, which was not detected by HMMER. This SH3 domain is a novel SH3 CWB domain that was identified in an S. aureus phage φNM3 (245). B) Shows figure adapted respectively from PSIPRED protein structure prediction online tool (242).*

The JACKHMMER (200) search of this protein produced a large list of matches, forty with above 50% identity that were searched for in the literature. All these proteins are identified as *S. aureus* phage genes with less than half of them with a predicted GO function on UniProt. The proteins that possess a predicted molecular function by GO (240,241) are all predicted have N-acetylmuramoyl-L-alanine amidase activity. As stated above, while classified as an endopeptidase, CHAP can also function as an amidase (63). Structural phage proteins involved in the tail structure region can contain catalytic domains (246), but for P2 there is a possibility that it functions as an endolysin.

HMMSCAN was unable to detect a domain at the C terminal domain, but a study done on generating a chimeric endolysin using a protein that is identical to P2 suggested there is an SH3 domain. A protein classified as Amidase (UniProtKB: A0EX11_9CAUD) found in *Staphylococcus* phage ϕNM3 (NCBI:txid387909) was identified and predicted to carry a novel CBD.

P2 CWB domain has been investigated in an article by Daniel and Euler (245). This article had three aims; 1) making a chimeric endolysin using the P2 CWB domain, 2) identification of a unique CWB domain in *Staphylococcus* lysins, and 3) testing whether the chimeric endolysin would act synergistically with antibiotics against MSRA. The article hypothesised that the novel CWB domain, instead of binding to the peptide cross bridges, would bind to alternative epitopes like the cell wall-associated carbohydrates (245). It was predicted that this modification to the binding site would reduce the possibility of the host strain becoming resistant to binding. This was confirmed to be a novel CBD which shares no known homology with any known CWB in the GeneBank database. This makes P2 of particular interest as a candidate for testing in the PHEARLESS system since the CWB is still considered to be novel. Identical genes were also found in a large number of other predicted phage genomes, well over half in the BHI collection.

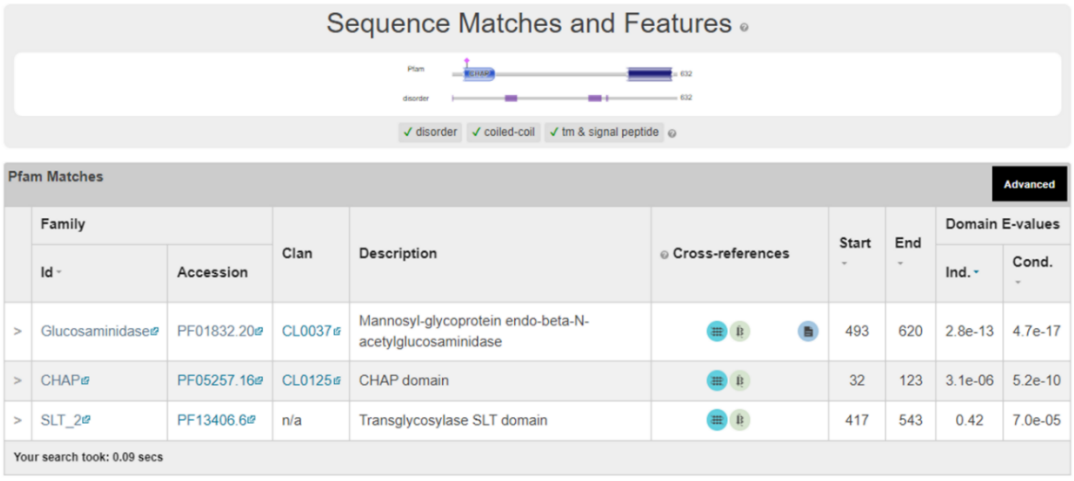
Putative Phage Protein with Predicted Antimicrobial Activity: PE_BHI_P3

The third protein selected from the BHI bioinformatics search (Chapter 5.c *S. aureus* Genomes Collected from Patients with Chronic Rhinosinusitis Provided by Basil Hetzel Institute ENT Surgery Group) was labelled PE_BHI_P3 (P3) (Figure 51) (72 kDa). This protein is not a predicted endolysin protein but likely a structural protein

located in the tail region, involved in PG layer degradation due to the CHAP domain and glucosaminidase domain. This protein comes from a predicted intact phage region (PHASTER Score 150) in the KUD_41 bacterial strain, contig 11. The KUD_41 strain is also the carries the only phage containing this exact protein sequence. This phage genome is 34.2 Kb in length containing 46 predicted genes. This phage has 16 genes in common with *Staphylococcus* phage 53 (NCBI: NC_007049.1). This phage appears in the JACKHMMER (200) search of this protein with 100% identity.

HMMSCAN results (Figure 51. A) also suggest that the protein contains two domains as well as a potential third domain. The first domain, CHAP, is located at the N terminus, with a glucosaminidase domain located at the C terminal. There is a possible third domain detected, a transglycosylase SLT domain, but this is questionable.

A)



Pfam Matches — Advanced

	Family		Clan	Description	Cross-references	Start	End	Domain E-values	
	Id	Accession						Ind.	Cond.
>	Glucosaminidase	PF01832.20	CL0037	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase		493	620	2.8e-13	4.7e-17
>	CHAP	PF05257.16	CL0125	CHAP domain		32	123	3.1e-06	5.2e-10
>	SLT_2	PF13406.6	n/a	Transglycosylase SLT domain		417	543	0.42	7.0e-05

Your search took: 0.09 secs

B)

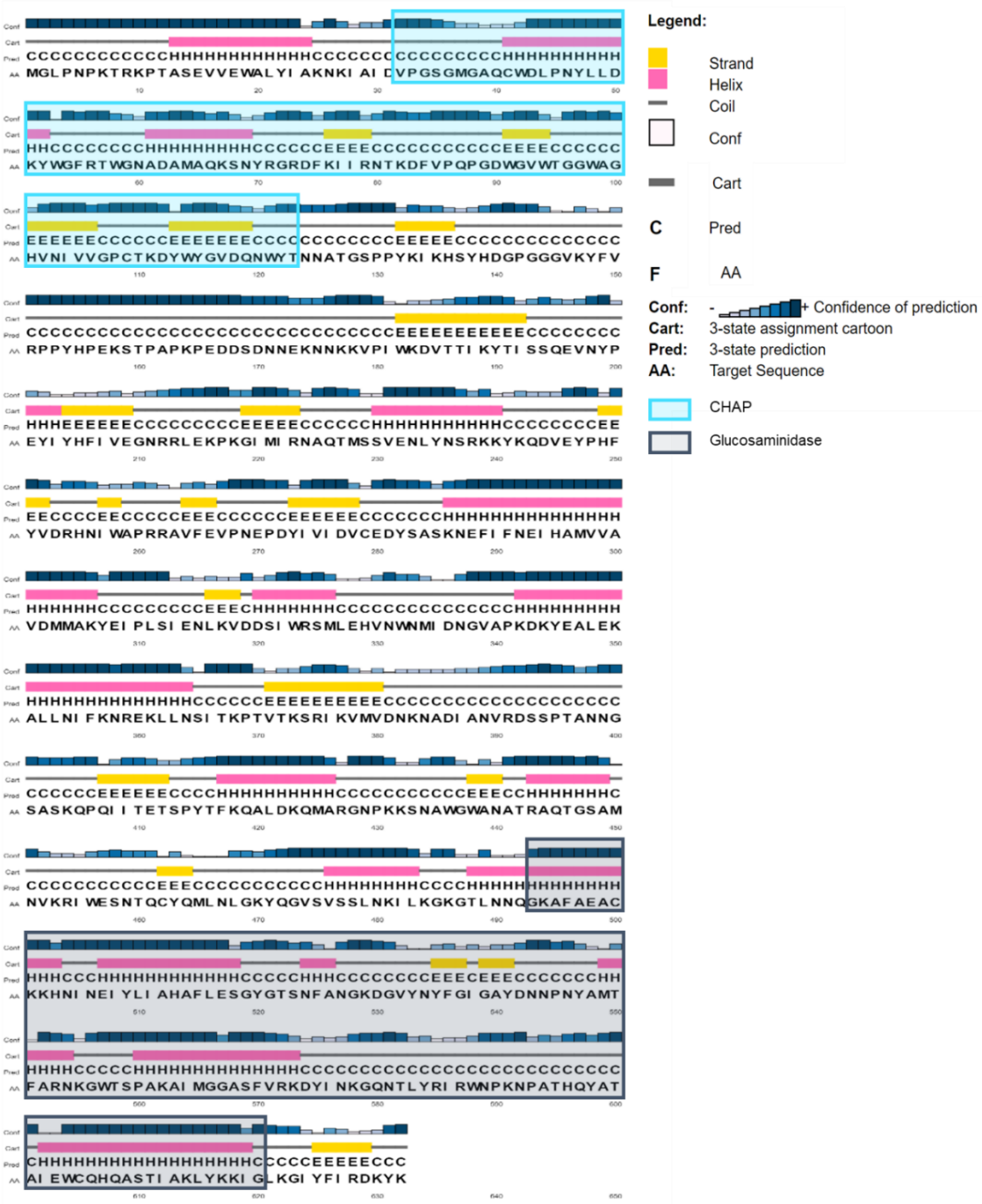


Figure 51: Putative phage protein (P3) with predicted antimicrobial activity against S. aureus, PE_BHI_P3, chosen from the BHI bioinformatics search. *A) Shows figure adapted from HMMER: Biosequence analysis using profile hidden Markov Models online web tool (200). HMMER identifies two domains, a N terminal CHAP domain (32 – 123 amino acids) and a C terminal Glucosaminidase (493 – 620 amino acids). There is a third domain detected, SLT_2 domains, but this domain is located within the Glucosaminidase domain which has a higher E-Value then the SLT_2 domains. B) Shows figure adapted respectively from PSIPRED protein structure prediction online tool (242).*

The first fifteen hits in the JACKHMMER (200) search have >90% identity with P3. This includes a hit for the BLAST protein results found in the PHASTER results, Endo-Beta-N-acetylglucosaminidase found in *Staphylococcus phage* SA12 (NCBI: txid1347760).

The JACKHMMER (200) search produced a number of proteins (~800) with many of the top forty labelled cell wall hydrolase. Twenty of the top forty proteins are predicted to have a Glucosaminidase (Pfam: PF01832) molecular function using the GO prediction tool (240,241). Of the top forty hits with above 50% identity none show any further investigation or characterization apart from protein predictions.

Two proteins which have 99% and 59% identity have originated from studies that have investigated the phage's infection capability. The first protein (99% Identity) is predicted to function as a putative tail-associated cell wall hydrolase protein (UniProtKB - B2ZZ02_BPMR2) originating form *Staphylococcus* phage phiMR25 (247). The second protein (59% Identity) is also a predicted tail tip protein (UniProtKB - A7VMY9_BPMR1) that originates from *Staphylococcus* phage phiMR11 (248) which is morphologically similar to phiMR25. Both of these phage have the advantage of having a wide host range and carrying no known genes for antibiotic resistance or toxins, making them potentially suitable for phage therapy.

Putative Phage Protein with Predicted Antimicrobial Activity: PE_GDT_P4

PE_GDT_P4 (P4) was not found in either the bioinformatics searches performed in Chapter 5: Bioinformatic Pipeline for Putative Antimicrobial Phage Protein Discovery. This protein as well as PE_GDT_P5 (P5) and PE_GDT_P6 (P6) were sourced by a separate bioinformatics search. Oliveria and Sampaio (232) investigated evolutionary relationships between different phage clusters. In addition, this study also

investigated the evolutionary relationship between the predicted endolysin genes, producing a list categorized into different clusters.

P4 is a single gene found in *S. aureus* phage vB_SauS_IMEP5. belongs to the B cluster phage. Cluster B was found to be the most diverse and largest cluster (232). P4 is predicted to contain a N-terminal CHAP domain, an Amidase_3 domains and a C terminal SH3_5 domains. These domains were detected using HMMSCAN, but to detect the SH3_5 domain the sensitivity need to be lowered with a significance bit scores (Sequence score: 15.0, Hit: 10.0) (Figure 52).

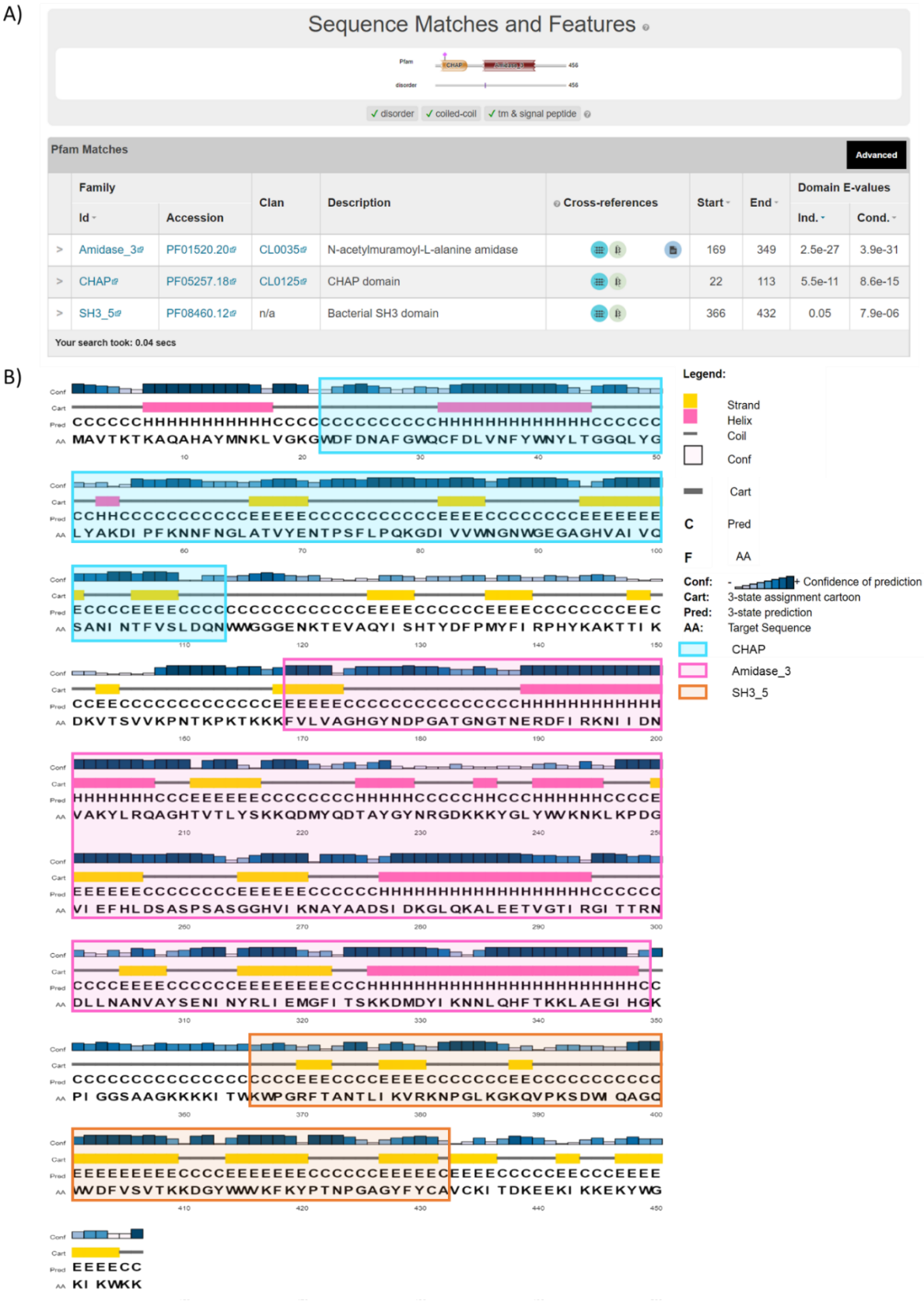


Figure 52: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_GDT_P4, chosen from a separate bioinformatics search by Oliveria and Sampaio (3). *A) Shows figure adapted from HMMER: Biosequence analysis using profile hidden Markov Models online web tool (1). HMMER identifies identified three possible domains, a N terminal CHAP (22 – 113 amino*

acids), Amidase_3 domains (169 – 349 amino acids) and a predicted C terminal SH3_5 (366 – 432 amino acids). The C terminal predicted SH3_5 domains is not present in the diagram due to the low E-Value. B) Shows figure adapted respectively from PSIPRED protein structure prediction online tool (2).

The first hit for this protein in JACKHMMER (200) produced an entry for a lysin gene from the original host phage vB_SauS_IMEP5 (NCBI Taxonomic identifier: 1852565). This phage was identified and characterised as a novel virulent phage collected from a dairy farm in Shihezi, Xinjiang, China (249). The top forty hits produced proteins that were predicted to contain N-acetylmuramoyl-L-alanine amidase activity by GO prediction tool, involved in the peptidoglycan breakdown process. None of the top forty hits which contained an above 50% identity have any functional studies performed, with the exception of one. The only protein that has had antimicrobial activity investigations performed is the only protein present in the list that is not a phage protein. Probable autolysin PH (UniProt: PH_STAA8) is an autolysin found in a *S. aureus* genome called *S. aureus* subsp. aureus NCTC 8325 (NCBI: txid93061). Functional studies were performed on this protein, the full length protein did not show any detectible activity (243). However, a chimeric endolysin created in the same studies using the catalytic domains of Probable autolysin PH did contain activity.

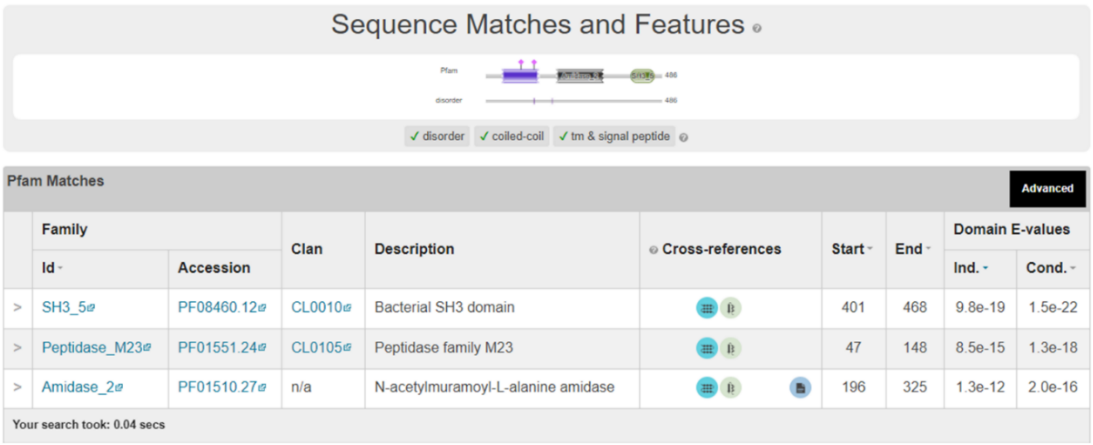
Putative Phage Protein with Predicted Antimicrobial Activity: PE_GDT_P5

PE_GDT_P5 (P5) was the second protein chosen from the separate bioinformatics search performed by Oliveira and Sampaio (232). P5 was also found in cluster B, originating from *S. aureus* 2638A (250). This endolysin is unique compared to the other endolysins contained in this list because it contains a secondary translation start site located in the gene. It is not revealed if the secondary translation site was detected using a feature of the bioinformatics search pipeline (232). It was likely discovered while investigating the proteins identified in the pipeline. The secondary translation site of the phage 2638A endolysin gene was discovered during protein purification (250). The bioinformatics results in this study did not detect any overlapping protein sequences. The full length protein produces a endolysin containing a N terminal peptidase domain (Peptidase_M23), a central amidase domain (Amidase_2) and a C terminal CWB domain. The truncated protein consists of a N terminal amidase domain and C terminal CWB domain (SH3_5) (Figure 53. A).

What makes this endolysin unique compared to the other typical endolysins found in all three bioinformatics searches is that the N terminal catalytic domain is a

Peptidase_M23. For *S. aureus* phage, CHAP_Amidase_SH3 is the common domain order observed (246). Peptidase_M23 was detected in other proteins in both bioinformatics studies performed in this project (Chapter 5.e Domains Detected in the Bioinformatics), but in those cases the domain was found in combination with a PhageMin_Tail domain, and in most cases an SLT domain. This suggests that the Peptidase_M23 proteins identified in Chapter 5 are most likely structural tail tip proteins, unlike P5 which has a domain layout more similar to an endolysin protein.

A)



B)

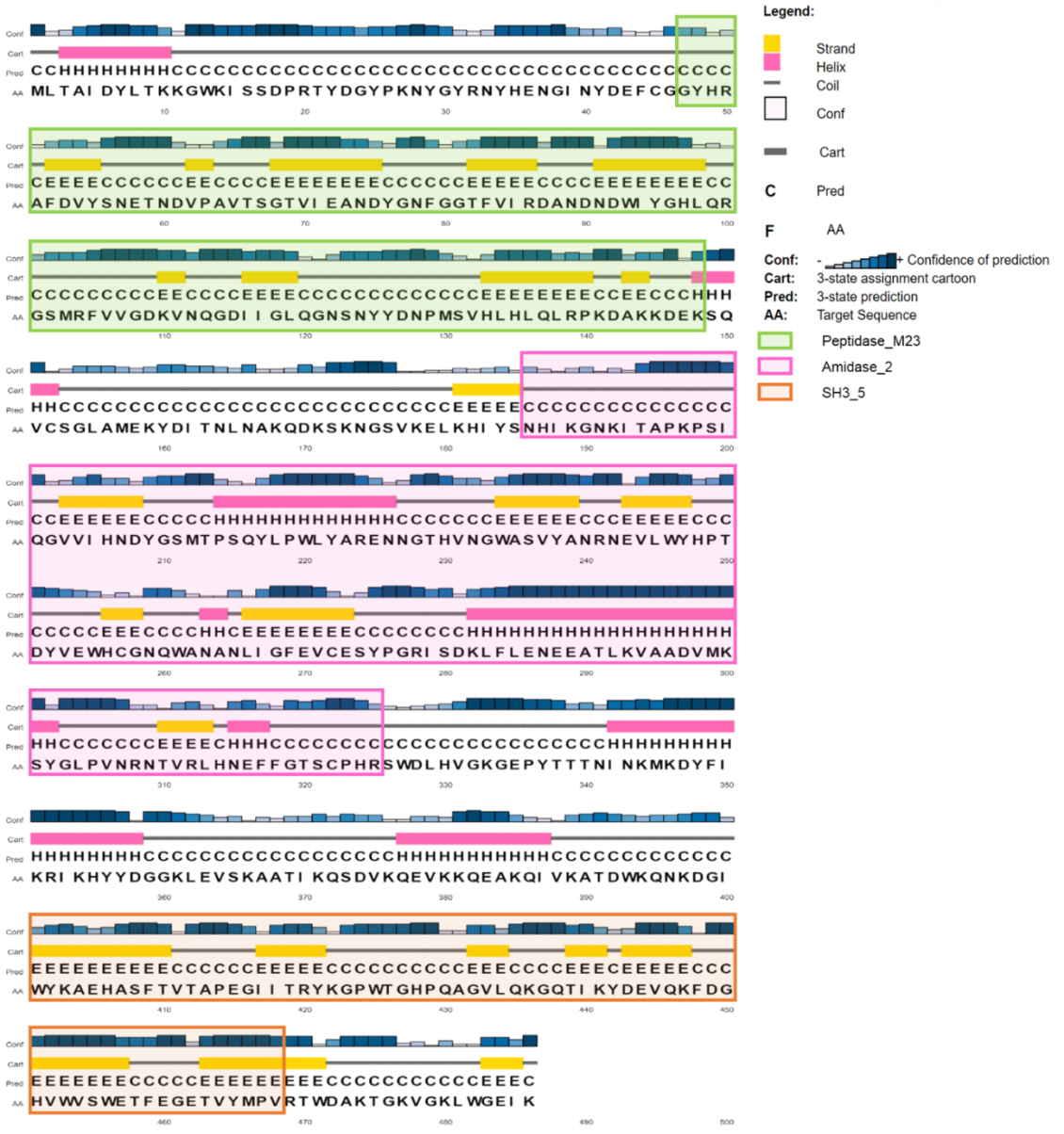


Figure 53: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_GDT_P5, chosen from a bioinformatics search by Oliveria and Sampaio (3). *A) Shows figure adapted from HMMER: Biosequence analysis using profile hidden Markov Models online web tool (200). HMMER identifies a N Terminal Peptidase_M23 domain (47 – 148 amino acids), Amidase_2*

domains (196 – 325 amino acids) and a C terminal SH3_5 domains (401 - 468). B) Shows figure adapted from PSIPRED protein structure prediction online tool (242).

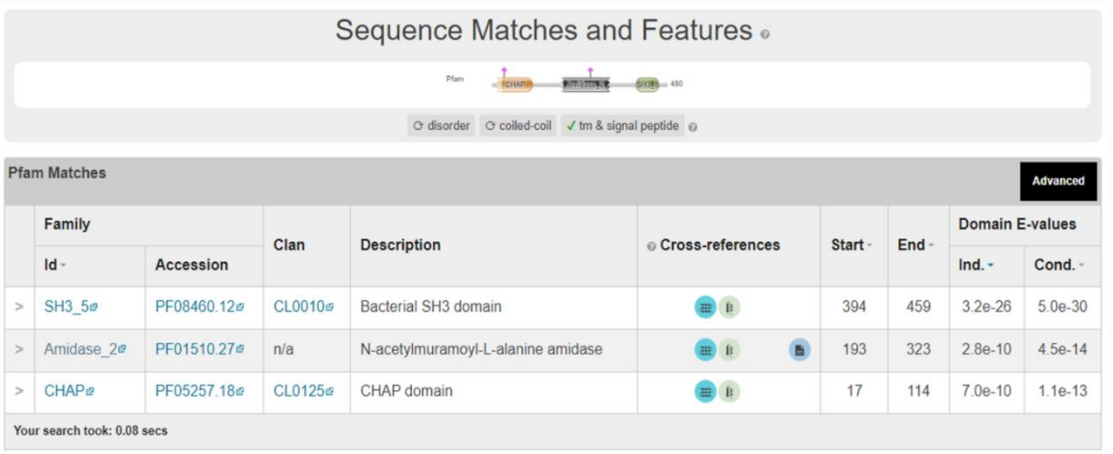
Putative Phage Protein with Predicted Antimicrobial Activity: PE_GDT_P6

PE_GDT_P6 (P6) was the last protein chosen from the separate bioinformatics search performed by Oliveira and Sampaio (232). Cluster D consists of two lytic phages, P6 originated from *S. aureus* phage vB_StaM_SA2. vB_StaM_SA2 is predicted to act as a lytic phage, but it does contain a tyrosine recombinase gene. It is also noted that vB_StaM_SA2 has a 10% higher G+C content then other *S. aureus* phage (232). P6 is strongly predicted to act as a endolysin gene due to its common endolysin domain architecture, N terminal CHAP domain followed by an Amidase_2 and a C terminal SH3_5 domain (Figure 54. A).

The secondary structure (Figure 54. B) of the CHAP domain contains one α helix structure and four β stands, similar to the secondary structures present in the previous CHAP domain containing proteins listed above (P1, P2, P3 and P4). The Amidase_2 domain contains a similar secondary structure to the Amidase_3 domains that is present in P1 and P4.

Both P1 and P4 organization shows an initial short single β sheet followed by a single long α helix followed a single short β sheet. After containing two to three short α helix followed by two small β sheets, finally one long α helix then two short β sheets ending in a final long α helix. P6 is classified as an Amidase_2 domains, it shows a similar organization compared to Amidase_3 but is much shorter. Containing an initial β sheet followed by a small α helix, two small β sheets, one long α helix ending in a short β sheet. SH3_5 has the same secondary structural organization in present in the previous proteins (P1, P2, P4 and P5) that contained a SH3 domain, with it consisting of five to seven β sheets.

A)



B)

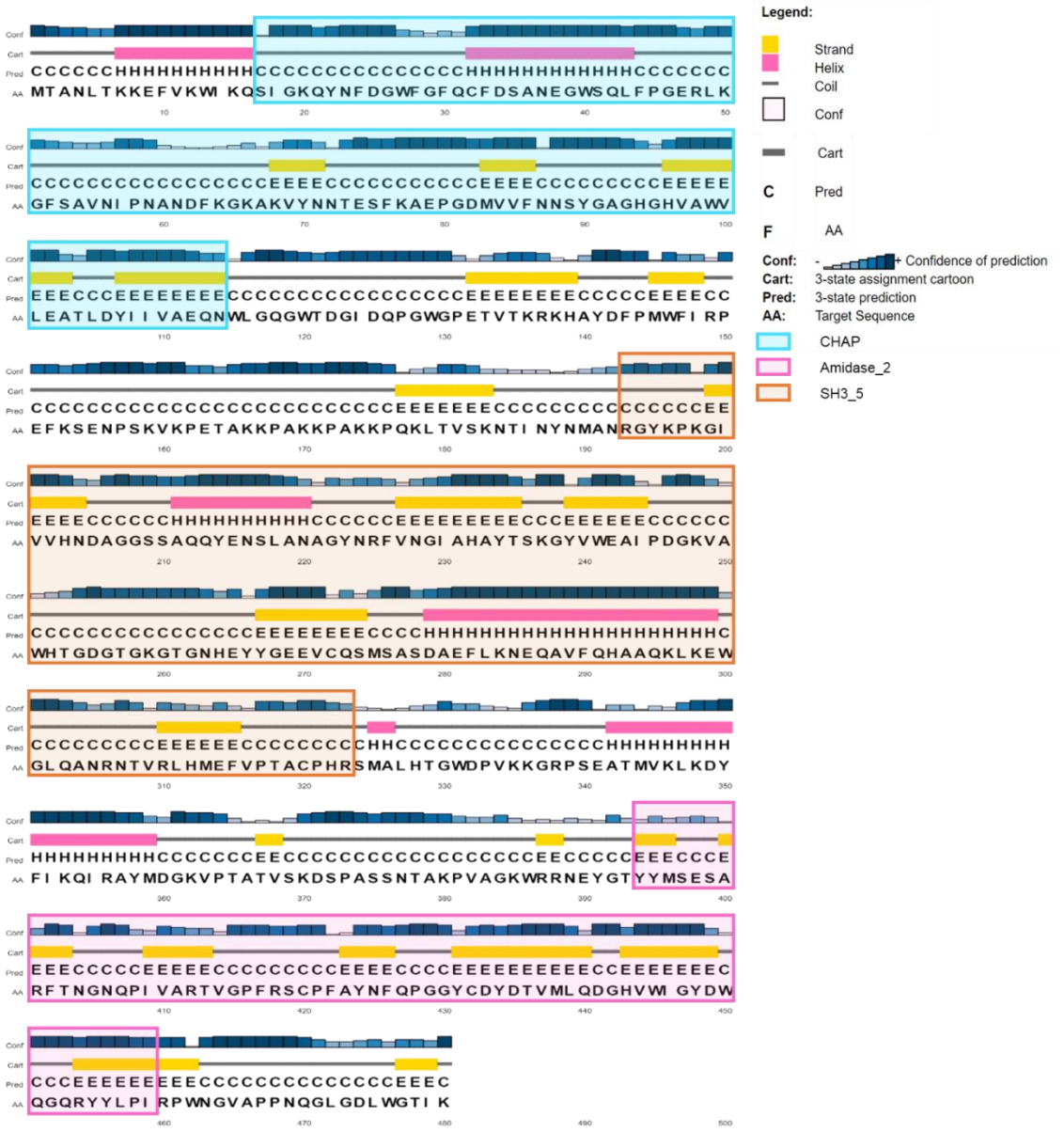


Figure 54: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_GDT_P6, chosen from a separate bioinformatics search by Oliveria and Sampaio (3). *A) Shows figure adapted from HMMER: Biosequence analysis using profile hidden Markov Models online web tool (200). HMMER identifies a N terminal CHAP domain (17 – 114 amino acids), Amidase_2 domains*

(193 – 323 amino acids) and a C terminal SH3_5 domains (394 - 459). B) Shows figure adapted from PSIPRED protein structure prediction online tool (242).

A literature search for homologs of protein P6 investigated the top thirty hits detected in JACKHMMER (200) which had a >56% identify and 70% similarity. The top thirty hits contain a number of different descriptions for the proteins detected including; endolysin, ORF##number, N-acetylmuramoyl-L-alanine amidase, putative cell wall hydrolase, Lysin, Amidase, autolysin and CHAP domain protein, with all of the hits originating from *S. aureus* phage.

Out of the top thirty protein hits by JACKHMMER, only one hit had been investigated for lytic activity, with experimental evidence performed at the protein level. Probable autolysin LytO (UniProtKB: LYTO_STAA8) was the first hit with the highest identity (65%) and similarity (76%). Probable autolysin LytO is a chimeric protein generated from an autolysin present in a *S. aureus* strain which was fused to an SH3 domain (243). This chimera was created with the aim of improving the activity of the autolysin. The wild type LytO gene, containing CHAP and Amidase_3 domains was discovered in a *S. aureus* strain (*S. aureus* strain NCTC 8325) (NCBI: txid93061). Autolysins present in bacteria have a critical function in growing and maintaining the bacterial cell wall, while also functioning in recycling of cell wall components (251). LytO did show lytic activity against *S. aureus*, but it was found to be weaker than other chimeric protein constructed (243).

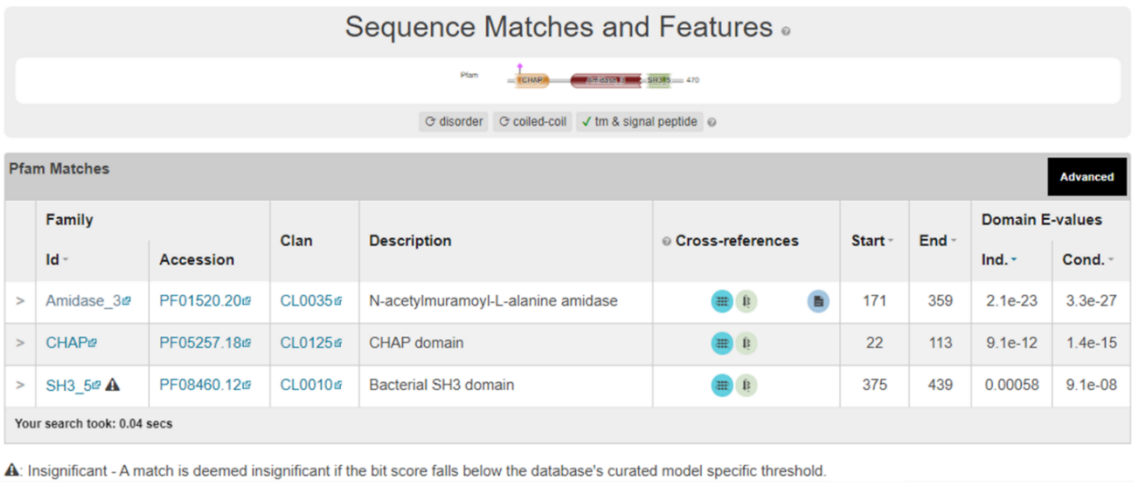
Putative Phage Protein with Predicted Antimicrobial Activity: PE_ CFLI_P7

The last protein in the list was discovered in the CF bioinformatics study (5.d *S. aureus* Genomes Collected from Patients with Fibrosis Lung Infections). PE_CFLI_P7 (P7) was found in an intact predicted phage detected in *S. aureus* strain CFBR_EB_Sa105. The phage has the top score for completeness (150) with sixty-one predicted phage proteins present. This phage appears to be novel with its closest relative, PHAGE_Staphy_SA97 (NC_029010) only containing twenty-five shared genes.

P7 is predicted to act as an endolysin protein due to its common endolysin domain architecture, CHAP-Amidase-SH3. HMMSCAN results (Figure 55. A) detects an N-terminal CHAP domain followed by an Amidase_3 domains. A SH3_5 domain was also detected at the C terminal, but the significance bit scores need to be lowered

significantly to detect it (Sequence score: 15.0, Hit: 10.0) (Figure 55. A). Even with lowering the sensitivity the SH3 domain match was still considered insignificant, falling below the database's curated model specific threshold used in HMMSCAN. The secondary structure (Figure 55. B) shows that the CHAP domain and Amidase_3 domain show a similar secondary structure present in previous CHAP and amidase domain. The CHAP domain following the same large α helix structure followed by four to five β stands. The amidase domain in P7 is an Amidase_3 domains, it follows the same α β pattern as P1 and P4 which also contain Amidase_3 domains. P7 is predicted to contain a C terminal SH3_5 domains, however, it was detected well below the detection threshold used by HMMSCAN. Therefore, it could be a false match or more likely given the protein's other domains, a novel SH3 domain/novel CWB domain. The weak SH3_5 domain predicted has a different predicted secondary structure compared to previous SH3 domains, as it contains two large α helices. SH3 domains are more commonly recognized by being organized with multiple β stands, seen in each of the SH3 domains (P1, P2, P4, P5 and P6) listed above.

A)



B)

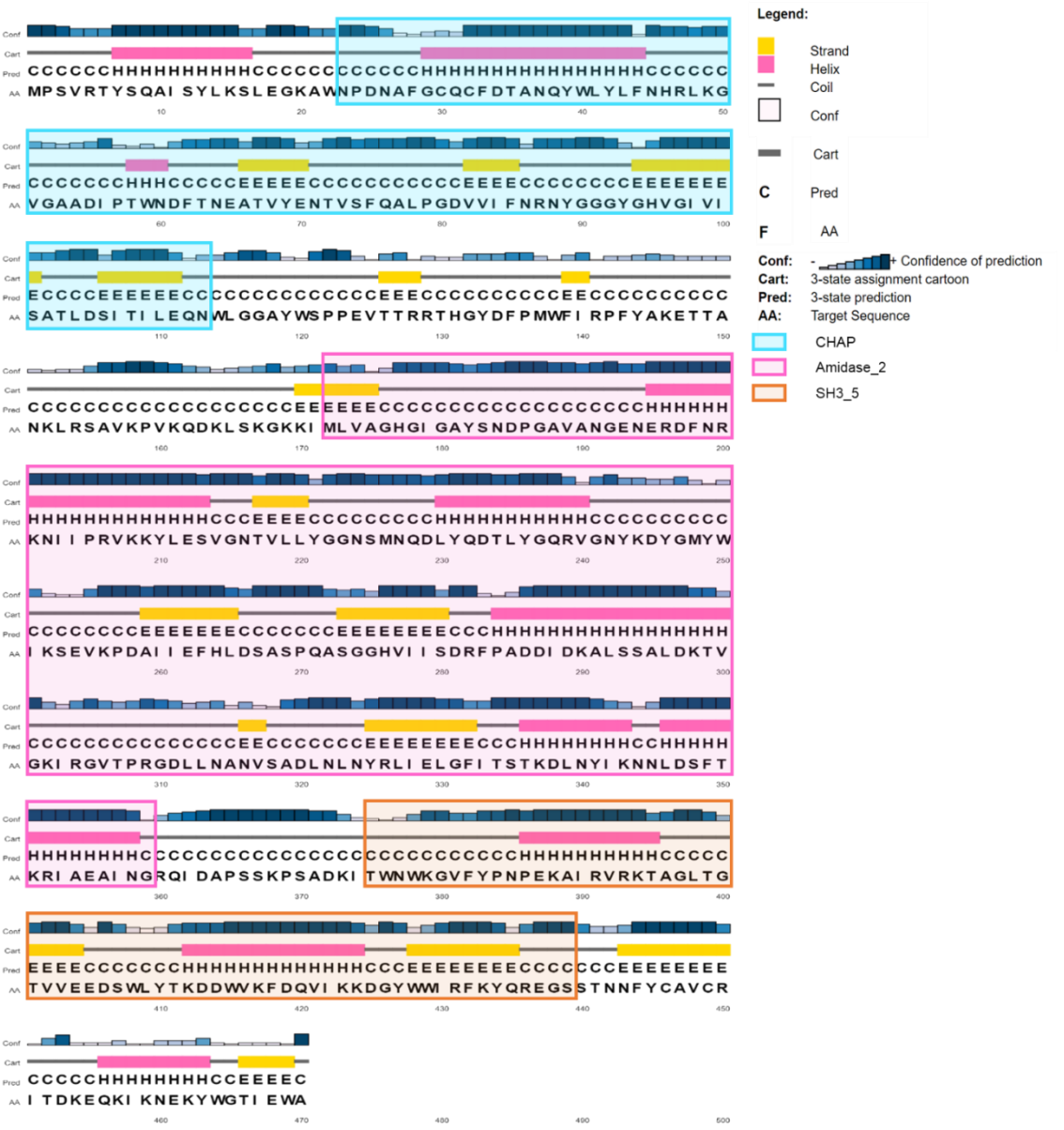


Figure 55: Putative phage protein with predicted antimicrobial activity against S. aureus, PE_ CFLI_P7, chosen from the CF bioinformatics search. *A) Shows figure adapted r from HMMER: Biosequence analysis using profile hidden Markov Models online web tool (200). HMMER identifies a N terminal CHAP domain (22 – 113 amino acids), Amidase_2 domains (171 – 359 amino acids) and a*

C terminal SH3_5 domains (375 - 439). While the C terminal domain has been identified as a SH3 domain, it has very low E value, suggesting that it is a new novel SH3 type domain or a novel domain type that shares homology with the SH3 family. B) Shows figure adapted from PSIPRED protein structure prediction online tool (242).

While P7 was chosen from the CF bioinformatics search, the protein was also present in a bioinformatics search performed by Oliveira and Sampaio (232). In the Oliveira and Sampaio study, P7 originated from *S. aureus* phage SA97, the same phage that PHASTER detected as the closest relative.

A literature search for P7 using JACKHMMER (200) shows that P7 is quite uncommon with only twenty hits having identify above 50%. The first hit in the JACKHMMER (200) results is the endolysin gene for the *S. aureus* phage SA97 (NCBI: txid1498171) (252). SA97 was shown to be capable of infecting RN4220 therefore it must have an active endolysin (252).

6.a.ii Testing selected proteins for activity using PHEARLESS version 2.2

Each of the seven proteins described in Chapter 6.a.i Proteins Discovered in Bioinformatics Searches were tested for lytic activity against the laboratory *S. aureus* strains, using the PHEARLESS V2.2 expression strain (Chapter 4.g PHEARLESS V2.2; Improved Version of the Expression Strain) with the PHEARLESS Protein Screening Protocols (Chapter 2.h PHEARLESS Protein Screening Protocols). The pIT4 integration plasmid carrying the *cos*+ sequence was integrated into the Δ*cos* 186 prophage, in the cumate inducible *tum72*-based expression strain. The pIT4 plasmids carrying one of the seven putative proteins were generated using Gibson assembly (Chapter 2.c Cloning Strategy). After the integration, PCR was performed to confirm that the pIT4 plasmids had been integrated at the correct site (Appendix Fig. 5). PCR fragments of the integrated proteins were amplified and sent for Sanger sequencing.

The DNA sequences for proteins of interest were codon optimised and ordered from Twist Biosciences (https://www.twistbioscience.com/). Sequences were provided pTwist Cm, a high copy number expression plasmid. The plasmid contains a pUC origin of replication that has a copy number between 150 – 200 and a

chloramphenicol resistance gene (Figure 56). These plasmids were initially tested as a possible expression plasmid to use in the PHEARLESS V1 expression strain (Chapter 3: Design and construction of the PHEARLESS system) since in addition to the protein of interest the designed sequences also carried the 186*pJ* promoter. However, the high copy number origin of replication (pUC) was found to make the expression strain very unstable, with difficulty in transformation, growth and retention of the plasmid even under selection. Due these issues, the genes were amplified from these plasmids and assembled into the pIT4 integration plasmids. The PHEARLESS V2.2 expression strain was chosen over the PHEARLESS V1 expression strain for these reasons; (1) The pIT4 system genes act as a single copy until expression, this is an advantage in cases where leaky protein expression is toxic. (2) The pIT4 system is integrated into the expression strain so there is a lower chance of losing the protein expression and removes the need for antibiotic selection.

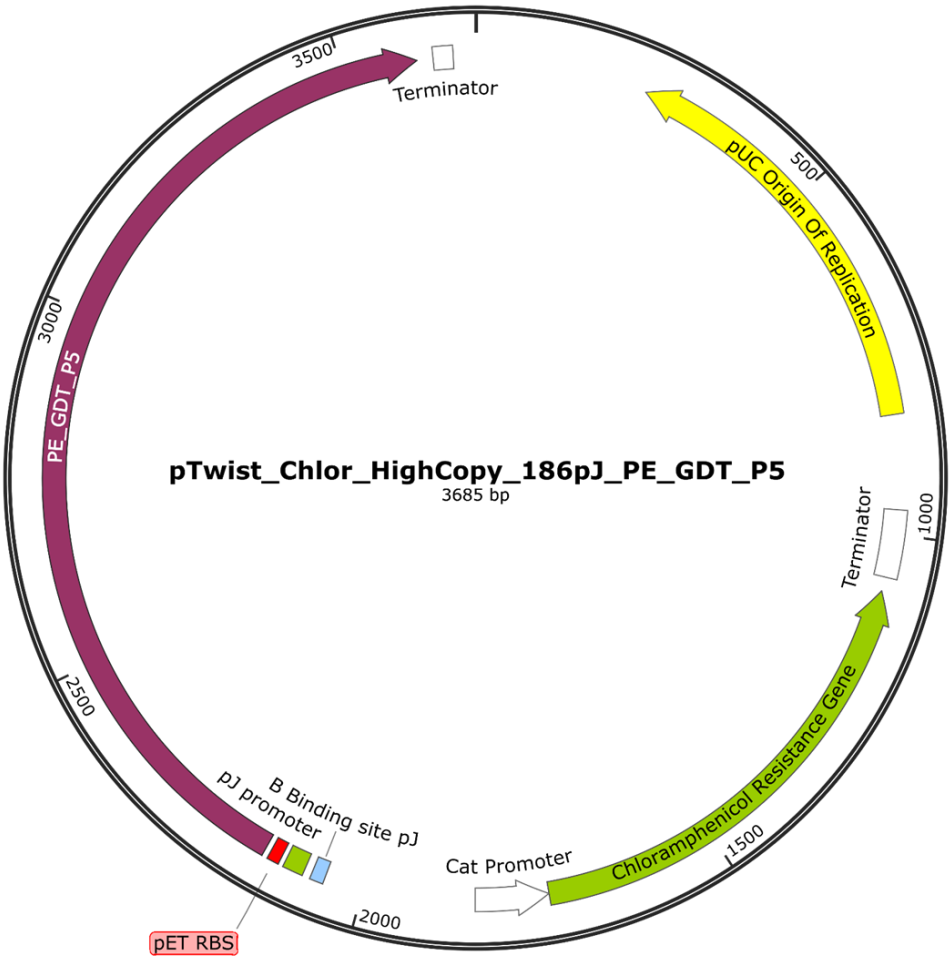


Figure 56: Twist Biosciences pTwist Cm High copy number Expression Plasmid Containing the PE_GDT_P5 gene controlled by the 186pJ promoter. *pTwist Cm plasmid map as an example of the plasmid organisation. The plasmid contains a Cm antibiotic resistance gene and a pUC origin of replication producing 150 to 200 plasmid copies per cell. Figure was generated using SnapGene DNA maps.*

Since the proteins are only being tested initially for antimicrobial activity, the protocols follow the PHEARLESS Protein Screening assay (Chapter 2.h PHEARLESS Protein Screening Protocols) All seven proteins were tested against two laboratory *S. aureus* strains, RN4220 and HER1049. The plates also included the propagation strain, QC024, to take advantage of phage propagation to increase protein production. All strains were grown to OD_{600} 0.6 and plated at a 50:1 target strain to propagation strain ratio. Expressions strain were diluted in a 10 fold dilution series up to 10^{-3}. Plates were divided into three sections for the two control strains (ClyF and empty) and the expression strain testing the putative proteins. 10 µL of sample were spotted onto the plates, plates were then incubated at 37°C overnight and observed the next day for activity.

The positive control expression strain containing ClyF (HB189) protein expression shows cell lysis occurring at each dilution. With the propagation strain advantage and the antimicrobial strength of ClyF, even the largest dilution (1/1000) was capable of producing enough protein to completely clear RN4220.

The negative control expression strain had the pIT4_*attP*2_LoxP_S_186*pJ_cos*+ integrated into the expression strain (HB253). This strain contains all the components of the positive control strain including the 186*pJ* promoter but lacking the ClyF gene. This control is important, since simply spotting samples onto the lawn can slightly disrupt lawn growth. The negative control samples on all the plates (Figure 57) show little disruption on the lawn strain and no clearing. There is a slight physical appearance difference where the samples have been added, helping to indicate where the samples were added. This is likely due to the physical difference between *S. aureus* (target strain) and *E. coli* (propagation stain). In the lawn mixture, the *E. coli* which is at a much lower concentration than *S. aureus* provided a "*speckled*" effect to the lawn. When the samples are spotted, the phage released kill the vast majority of the propagation stain, leaving the target strain to grown in across the small gaps created. This provides contrast to the "*speckled*" effect that the propagation strain gives.

Initial results for each of the putative proteins; P1 (HB243), P2 (HB245), P3 (HB250), P4 (HB239), P5 (HB241), P6 (HB247) and P7 (HB237), showed no lytic activity against the two laboratory *S. aureus* strains, RN4220 (Figure 57) and HER1049 Figure 58). Each protein shows the same result as the negative control strain for each of the plates. At least four of the proteins (P1, P2, P3 and P4) came from clinical *S. aureus* isolates, and so target strain specificity could be responsible for lack of activity.

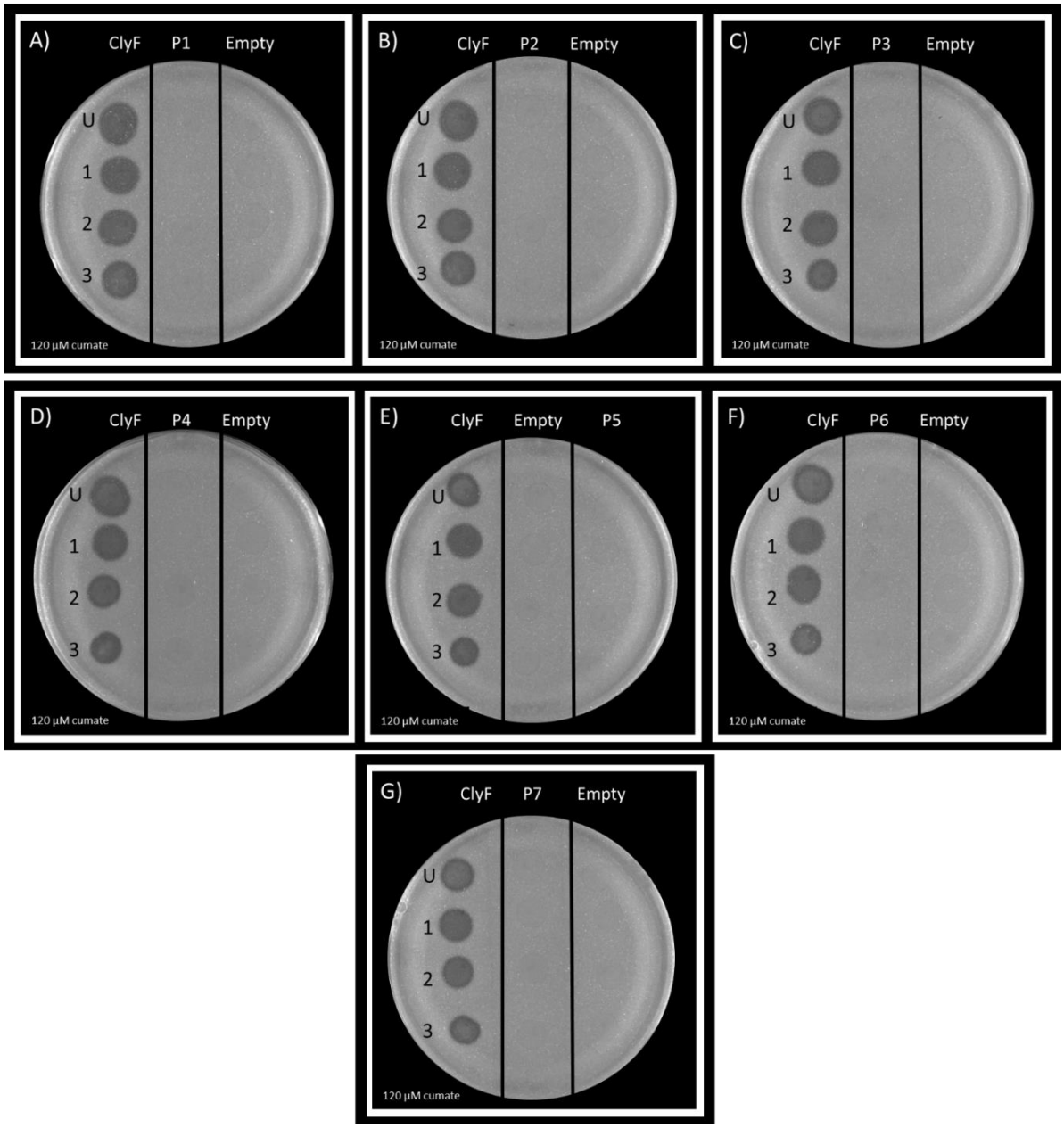


Figure 57: PHEARLESS Protein Screening Assay using the PHEARLESS V2.2 Expression Strain, Testing Seven Putative Proteins for Antimicrobial Activity Against S. aureus Strain RN4220. *Each plate contains a Target strain (RN4220) to Propagation strain (QC024) ratio of 50: 1, and 120 µM cumate. Each plate was spotted with a positive ClyF expression strain control (HB189, ClyF) which produce target cell lysis in each of the samples added to every plate (left series of spots). The negative control expression strain (HB253, Empty) which shows no target cell lysis in any samples on each plate (spots on the right of each plate. A-G) figures show results for proteins 1 to 7 (centre spots). U= undiluted, followed by serial 10 fold dilutions.*

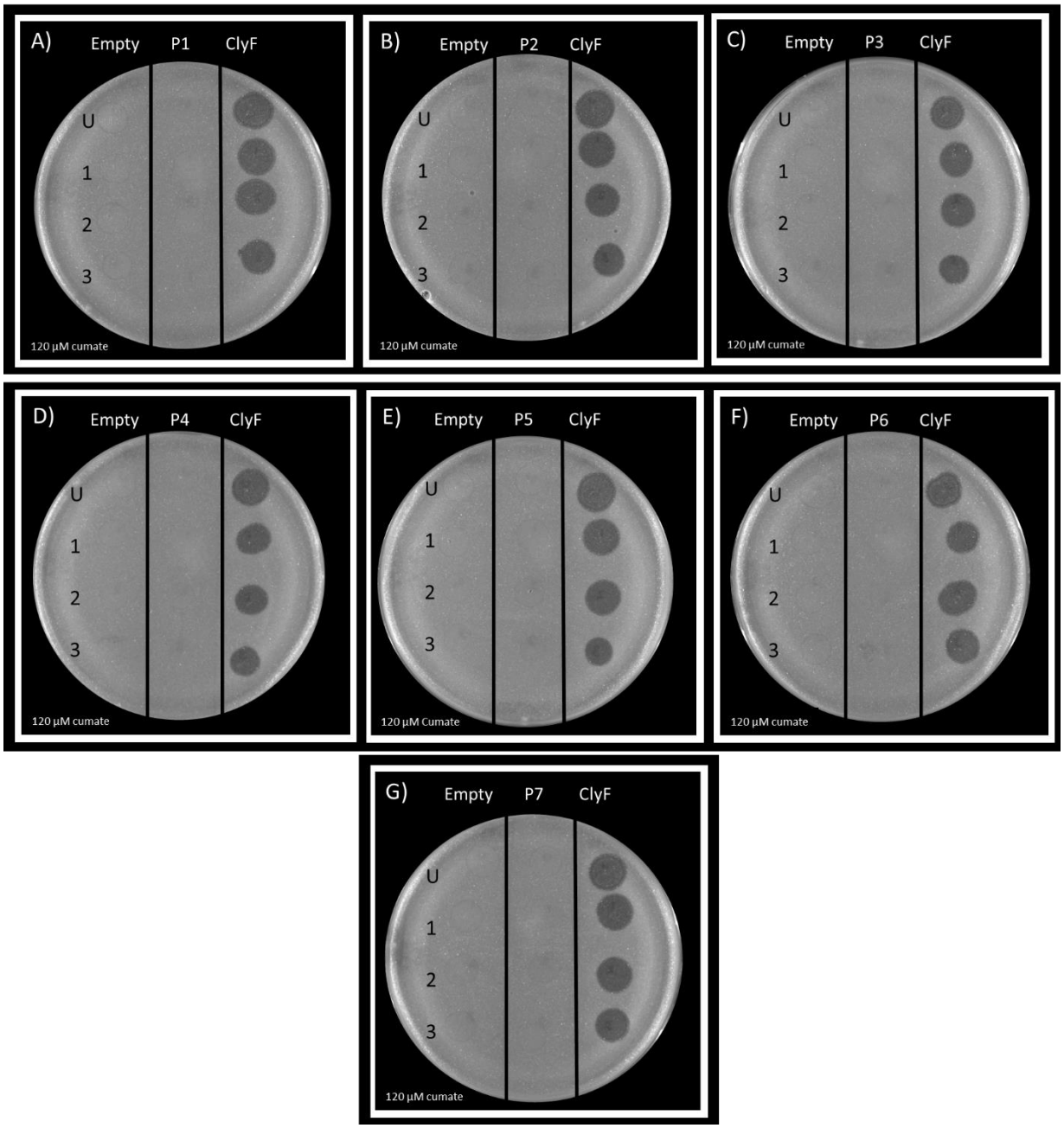


Figure 58: PHEARLESS Protein Screening Assay using the PHEARLESS V2.2 Expression Strain, Testing Seven Putative Proteins for Antimicrobial Activity Against S. aureus Strain HER1049. *Each plate contains a Target strain (HER1049) to Propagation strain (QC024) ratio of 50: 1, and 120 µM cumate. Each plate was spotted with a positive ClyF expression strain control (HB189, ClyF) which produce target cell lysis in each of the samples added to every plate (series of spots on right of plate). The negative control expression strain (HB253, Empty) which shows no target cell lysis in any samples on each plate (spots on the left of each plate. A-G) figures show results for proteins 1 to 7 (centre spots). U= undiluted, followed by serial 10 fold dilutions. None of the proteins showed activity.*

None of the proteins initially showed any antimicrobial activity against either laboratory *S. aureus* strain, RN4220 and HER1049. After completing the assay, plates were stored at 4°C, so they could be re-examined at a later time points. The

reasoning was two-fold. First, that enzyme activity may be weak, and take time to become obvious to the naked eye. Second, that because not all of the expression cells will lyse, un-lysed cells may be able to persist on the plates, with a portion of the cells lysing to releasing more enzyme. One of the advantages of endolysins and chimeric endolysins are that they will continue to diffuse and retain activity until degraded. This is consistent with the observation that the lysis spots that were created using strains containing ClyF would continue to expand over time (Figure 59).

When re-examining the proteins tested using the pIT4 integration plasmids in the PHEARLESS V2.2 expression strain, plates with P5 expression showed some evidence of activity and were periodically photographed. After the initial test to observe antimicrobial activity after a one-night incubation at 37°C, the plates were wrapped in cling wrap and stored at 4°C. Photographs of the plates were taken at three-to-four-day periods for up to one month. Photos taken on day one (Figure 59, Day One shows samples containing P5 expression appeared to be indistinguishable from the results produced by the empty control. The empty control samples do not change between day one and day thirty, with no killing formation over that time. At the thirty-day point, faint clearing had appeared in each of the P5 samples (Figure 59), suggesting some weak activity had appeared after one month.

After testing for antimicrobial function, it was discovered that the homologous protein discovered in JACKHMMER was current with the most recent publication. In addition, there was a second phage identified as 2638A, its endolysin protein also shared homology but had yet to be functionally tested (UniPro: Q4ZD58_BP263). P5 has had previous functional test performed (250), which also identified during purification that phage 2638A endolysin contained a secondary start codon, leading to the expression of a truncated endolysin protein during cell lysis.

P5 was tested against multiple *S. aureus* strains, including MRSA strains (CSA #175, SRCAMB collection). This study purified P5 using His tag chromatography, and lytic activity was assessed using plate and turbidity assays. Plate cell lysis results showed lytic activity for the full-length protein, producing a "*very broad, ill-defined region of clearing that grows with time*" (250).

Although P5 antimicrobial activity has already been confirmed, the original strains tested included MRSA strains. Due to albeit slow activity against our non-clinical

RN4220 strain, it could indicate a broad host range, making it a good candidate for both mutant library screening and chimeric endolysin design to increase antimicrobial activity.

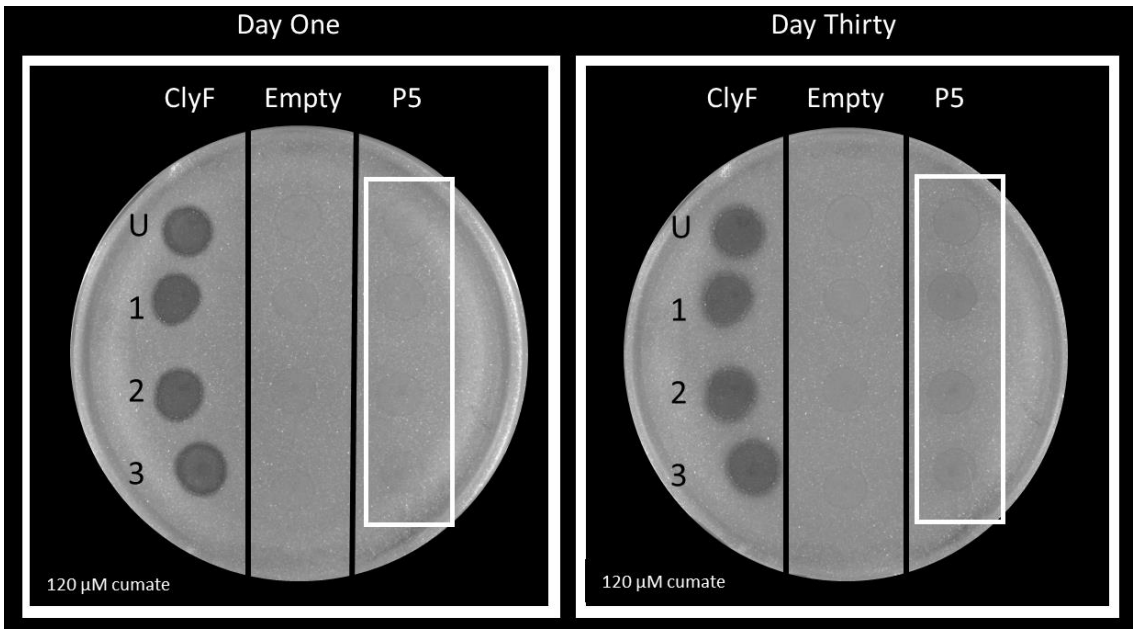


Figure 59: Times Lapse of PHEARLESS Protein Screening Assay using the PHEARLESS V2.2 Expression Strain, Testing PE_GDT_P5 for Antimicrobial Activity Against S. aureus Strain RN4220*. Plate contains a Target strain (RN4220) to Propagation strain (QC024) at a 50: 1 ratio with 120 µM cumate. Control strains include the positive control ClyF expression strain (HB189, ClyF) which produces target cell lysis. The negative control expression strain (HB253, Empty) shows no target cell lysis against target strain. Day One photo taken on day one following overnight incubation. Plates were stored at 4°C and the second photo, Day Thirty, was taken thirty days later. White box indicates where cell lysis of the target strain has begun to appear. This effect is not observed for the empty expression stain.*

6.b Putative Endolysin Genes Provided by the Basil Hetzel Institute

A collaborative project between the ENT surgery group at the Basil Hetzel Institute, Adelaide and the Shearwin Laboratory identified two unusual endolysin proteins. A set of studies were performed by Dr. Sha Liu in the ENT surgery group, testing *S. aureus* killing using a combination therapy of phage and antibiotics. Bacteriophage insensitive mutants (BIMs) of *S. aureus* were found to be re-sensitized to phage infection by the presence of ½ minimal inhibitory concentration (MIC) of clindamycin. The phage that were produced following infection in the presence of ½ MIC clindamycin (exit phage) were sequenced, and found to differ from the original phage (entry phage) only in their endolysin gene (253).

Two exit phages selected from the combination therapy were found to behave differently, with the exit phage now capable of infecting the same host without the need for ½ MIC clindamycin treatment. The entry phage endolysin contains a common domain structure, that is, an N-terminal CHAP domain, a central amidase domain and a C-terminal SH3 CWB domain (Figure 60, A). The parent endolysin protein is identical to an endolysin protein found in *Staphylococcus* phage K (phage K), identified as ORF30/ORF32 (LysK) (254) (NCBI - GenBank: AAO47477.2) (Figure 60, E) (254).

The protein is identified as ORF30/32, this dual assignment of open reading frames being due to the presence of a homing endonucleases domain, HNH endonucleases, located within the amidase domain. Adjacent to the HNH endonucleases domain is a DNA-binding motif identified as a NUMOD4 motif. This type of DNA-binding motif is commonly found in putative HNH endonucleases in bacteriophage (255).

Aligning the nucleotide sequence for the entry phage endolysin protein and the phage K ORF30/32, including the sequence between the ORF30 and ORF32, produces an identical sequence. Earlier studies have shown that when the endolysin gene (LysK) is expressed, the HNH endonuclease/NUMOD4 motif is removed via a splicing mechanism, forming a final protein with a functional amidase domain (254). The architecture for the final protein for ORF30/32 is an N terminal CHAP domain, Amidase domain and a C terminal SH3 domain, 495 amino acids in length. Homing endonucleases are enzymes that facilitate horizontal transfer of genetic elements, via cleavage of a recognition sequence (12-44bp). In phage, the HNH homing endonucleases enzymes are often found encoded within group I intron mobile elements, and can also be involved in DNA packaging (256,257). Group I introns are more commonly found within structural RNA genes, but when inserted within a protein-coding gene, it is evidence that the gene may be selected against (256). Insertion of the HNH endonuclease into the LysK related endolysin gene results in the generation of multiple stop codons within the endolysin reading frame (Figure 60, A).

There are also identical versions of this endolysin protein in the Genbank database that do not contain the HNH homing endonucleases. One example of this is an endolysin gene phiSA012_ORF51 that originates from a Lys-phiSA012 phage. This protein (phiSA012_ORF51) is identical to the protein sequence for the final

ORF30/ORF32 protein except for a single amino acid (83 amino acids). phiSA012_ORF51 (Figure 60, E) has demonstrated a broad host range with strong lytic activity (258).

The two exit phages recovered in the experiments described above, which have lost the HNH insertion, are likely to be a result of selection for an active endolysin. Instead of retaining the amidase domain that is reformed in splicing of the wild type parent phage endolysin, Clinda_3 (Figure 60, C) and Clinda_8 (Figure 60, D) both have the amidase domain completely removed when the NUMOD4 motif is lost, producing an endolysin that only contains a N-terminal CHAP domain and a C-terminal SH3 CWB domain. It has been noted in some studies that removal of the amidase domain results in an increase in lytic activity of the enzyme (108,112,199).

To test the enzymatic activity of the new endolysins produced from the modified phage, the proteins were tested in the PHEARLESS V1 expression strain, using the PHEARLESS Protein Screening protocols. The protein variants tested included the two mutant progeny endolysins Clinda_3 (Figure 60, C), Clinda_8 (Figure 60, D). Three variants of the original phage endolysin were also tested. (i) A first translation product (truncated by a stop codon generated by the presence of the HNH module) (Figure 60, B) which includes only the CHAP domain and a small section of the amidase domain up to the first noted stop codon, (ii) a LysK(HNH-H77A) mutant which targets the catalytic center of a HNH homing endonuclease (Figure 60, F) , and (iii) the full parental LysK (Figure 60, E) that lacked the NUMOD4 motif.

The HNH H77A mutation was chosen based on both sequence alignments and alpha fold structural predictions of the HNH protein. This LysK mutant was generated by Dr Nan Hao and was assembled into the protein expression plasmid, pZS(^)45 186*pJ*. Notably, while generation of the H77A variant was straightforward, the endolysin carrying the wild type HNH insertion could not be cloned. Only a few colonies were observed and when sequenced, each clone carried additional mutations. This strongly suggests that the HNH insert is active and in some unknown manner leads to *E. coli* cell death when expressed. PHEARLESS V1 expression strain was chosen over PHEARLESS V2.2 expression strain for initial testing due to ease of cloning and testing.

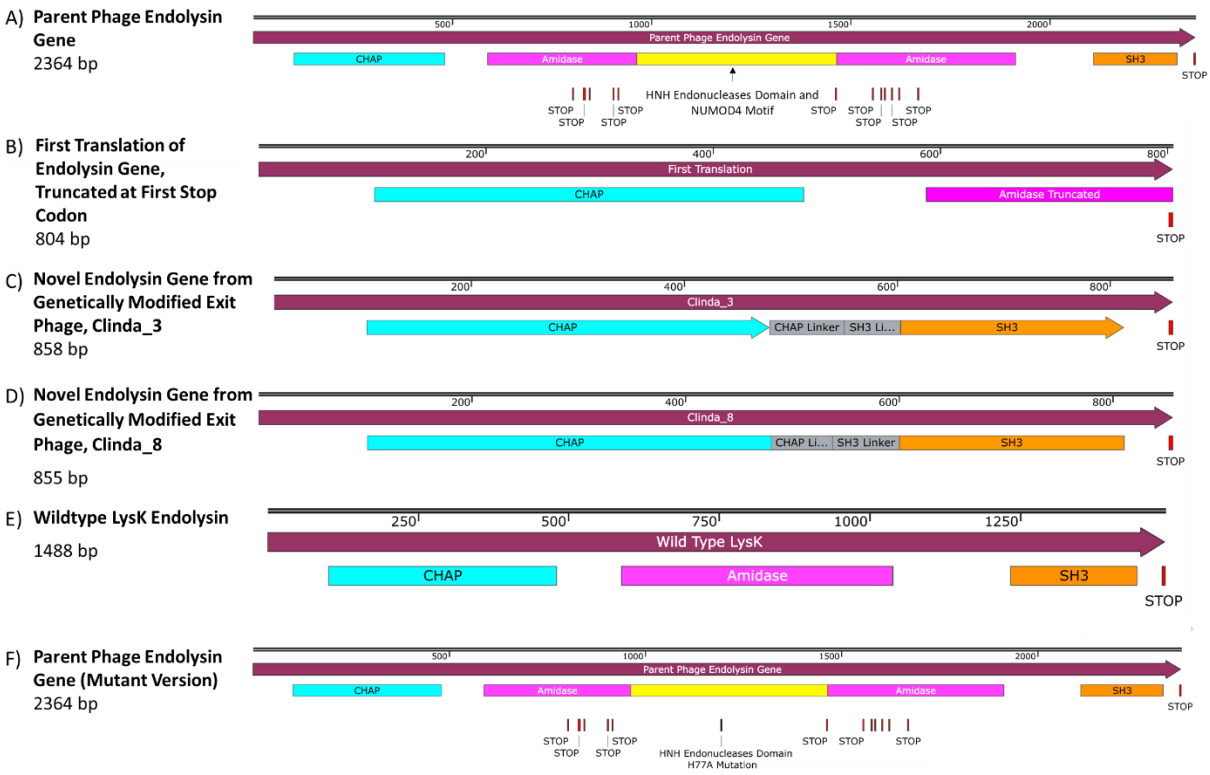


Figure 60: Domain Maps for Endolysin Proteins Discovered by the ENT surgery group at the Basil Hetzel Institute*. Progeny (exit) phage were isolated that had gained the ability to infect a bacteriophage-insensitive mutant (BIM) of S. aureus. A) The parental (entry) phage can only infect the BIM in the presence of ½ MIC of specific antibiotics (clindamycin in this case). The entry phage endolysin gene contains a CHAP domain, SH3_5 (SH3) domain and a central Amidase_2 (Amidase) domain containing an HNH endonuclease domain and NUMOD4 motif. B) A version of the first translation of the parent gene A) up until the first stop codon. This protein will contain a CHAP domain and a truncated Amidase_2 (Amidase Truncated). C) Clinda_3, novel endolysin gene identified in one of the two exit phage. The gene has had the Amidase_2 domain removed, and had rejoined the linker regions in frame. E) CLinda_8, novel endolysin gene identified in the second example of an exit phage. Like Clinda_3, Clinda_8 also has had the Amidase_2 domain removed but carries six amino acid differences at the junction point. E) A "wild type" LysK endolysin gene that contains the CHAP, Amidase_2 (Amidase) and SH3_5 (SH3) domains without the NUMOD4 motif was also generated. F) The final version of the entry phage endolysin gene created, containing the CHAP domain, SH3_5 (SH3) domain and an Amidase_2 (Amidase) domain with an inserted HNH endonuclease domain and NUMOD4 motif, where the HNH endonuclease domain had been mutated (His to Ala mutation at the 77th amino acid) to inactivate HNH activity. Figure was generated using SnapGene DNA maps.*

The six protein sequences inserted into the pZS(^)45 186*pJ* expression plasmid and expressed in the PHEARLESS V1 expression strain (HB70). For cloning purposes, gene fragments were ordered from IDT and assembled with the pZS(^)45 186*pJ* plasmid, using Gibson assembly.

6.b.i Testing Putative Endolysins Provided by the Basil Hetzel Institute using PHEARLESS assay

The five proteins; First Translation (Figure 60, B) (HB301), Clinda_3 (Figure 60, C) (HB298), Clinda_8 (Figure 60, D) (HB294), LysK (Figure 60, E) (HB296) and LysK(H77A) (HB306) (Figure 60, F) were tested against *S. aureus* six strains, C319, C244, C330, C43, C259 and ATCC 25923. Strains C319 (MSSA), C244 (MRSA), C330 (MRSA), C43 (MSSA), C259 (MRSA) are clinical isolates originally harvested from the nasal cavity of patients suffering from chronic rhinosinusitis. Samples were collected between 2011 and 2015 from patients in Adelaide, South Australia. ATCC 25923 (MRSA) (*S. aureus* subsp. *aureus* Rosenbach), a clinical isolate isolated in Settle 1945, was purchased from ATCC (American Type Culture Collection).

These assays were performed using protocols detailed in Chapter 2.h PHEARLESS Protein Screening Protocols, with minor modifications. Two to three biological replicates were performed for each protein on each target strain. Target strains were streaked out on LB agar plates and incubated overnight at 37°C. A single colony was picked and resuspended in 3 mL of LB agar, incubated overnight shaking at 180 rpm. After the overnight incubation, 600 µL of overnight culture was added to initiate a fresh 50 mL culture (LB plus spectinomycin 50 µg/mL) for the PHEARLESS assay.

Plates contained 120 µM cumate and three expression strains were tested on each plate, two proteins of interest and one control strain. Control strains were either a negative control, being the empty control strain (HB302) or a positive control, being the expression strain producing ClyF (HB305). Initially, it was unknown whether ClyF would be effective against any of the clinical *S. aureus* strains. The C319 lawn strain was the strain from which each exit phage (Clinda_3 and Clinda_8) was isolated, with the parent (entry) phage unable to infect without the presence of clindamycin.

The Empty control strain produced no killing against any of the target strains tested, confirming that the expression strain cell lysis and release of 186 phages has no effect on the target *S. aureus* strains. There was the lawn disruption effect in some of the undiluted (U) samples for the empty control strain but this effect was not seen beyond the undiluted sample. The positive control, ClyF, was found to be effective against each clinical *S. aureus* strains tested here; C319 (Figure 61), ATCC 25923 (Figure 62), C43 (Figure 63), 259 (Figure 64), C330 (Figure 65) and C244 (Figure

66). ClyF produced clear zones up to the third dilution (3) for all strains, identical to the clearing seen against laboratory strain RN4220 (Figure 16).

Both Clinda_3 (Figure 60, C) and Clinda_8 (Figure 60, D), showed activity against each of the clinical samples they was tested against; C319 (Figure 61, C), ATCC 25923 (Figure 62, C), C43 (Figure 63, C), 259 (Figure 64, C), C330 (Figure 65, C) and C244 (Figure 66, C). In all cases, these variants were not as effective at clearing the target strain as ClyF, but they did produce obvious killing in the undiluted (U), first dilution (1) and second dilutions (2). Identical results were observed in each biological replicate (minimum 3 each target strain) for Clinda_3 and Clinda_8.

LysK (Figure 60, E) and LysK(H77A) (Figure 60, F) was predicted to have little or no activity against any of the phage resistant *S. aureus* strains, with perhaps the exception of ATCC 25923 (Figure 62), given the parent phage is capable of infecting this strain. FT (Figure 60, B) is the first translation product of the LysK protein, from the ATG up until the first stop codon present in the parent protein as a result of the HNH insertion. The FT protein contains the CHAP domain and section of the truncated amidase. By testing FT, it can be determined if this truncated protein has any catalytic activity.

Against strains C319 (Figure 61, A, B), 259 (Figure 64, A, B), C330 (Figure 65, A, B) and C244 (Figure 66, A, B), LysK, LysK(H77A) and FT showed no evidence of any lytic activity, producing results identical to the empty control. ATCC 25923 (Figure 62, A, B) and C43 (Figure 63, A, B) produced inconclusive results for LysK, LysK(H77A) and FT.

Against target strain C43 (Figure 63, A, B) and ATCC 25923 (Figure 62, A, B), LysK, LysK(H77A) and FT appear to show some faint killing activity. This appears to be real killing rather than the lawn disruption effect because there are faint indications in the first dilution spots for each strain. The lawn disruption effect can be observed in the Empty control strain on the ATCC 25923 lawn containing the Clinda_3 and Clinda_8 expression strain (Figure 62, C), an effect that is not present in the first dilution. What makes these results inconclusive are biological replicates which do not exhibit the sample faint killing in the first dilution (1) sample. Variations of the assays where the expression strain is concentrated 10x before making dilutions were performed to try and determine activity but those results were also not conclusive, in that there were

indication of where the samples where added for the 10x concentrated undiluted samples and first dilutions, but they were still very faint.

Thus, the Lysk, LysK(H77A) and FT proteins may have some very weak activity against C43 (Figure 63, A, B) and ATCC 25923 (Figure 62, A, B), however it is so weak that the PHEARLESS V1 assay system is not sensitive enough to detect this reliably. Consistent with the phage infection data, the Clinda_3 and Clinda_8 results clearly show that, following HNH removal (irrespective of the mechanisms responsible for removal), the resulting CHAP-SH3 endolysins have enzymatic activity against the target strains.

Some possible solutions to improve the sensitivity of the assay would be place the genes in the PHEARLESS V2.2 expression strain, so that the propagation strain could be used to increase the amount of protein expressed. Long periods of incubation like performed when testing P5 could also be used to test for activity but that would also be dependent on the half-life of the protein. If these proteins have a short half-life, the proteins would degrade before enough target cell death occurs for observation. A final option could a liquid version of the assay system. Using liquid culture and following cell growth via optical density could potentially provide a more sensitive assay, with the advantage of providing data on the kinetics of enzyme activity. Experiments would be performed in a similar manner to the growth rate experiment protocols found in Chapter 2.b.vii Optical Density Reading of 186 Tum Variants in Liquid Cultures, with mixtures of target lawn to expression strain. This liquid culture assay is being pursued by other members of the Shearwin laboratory, following a recent successful bid to fund a specialised, temperature-controlled plate reader.

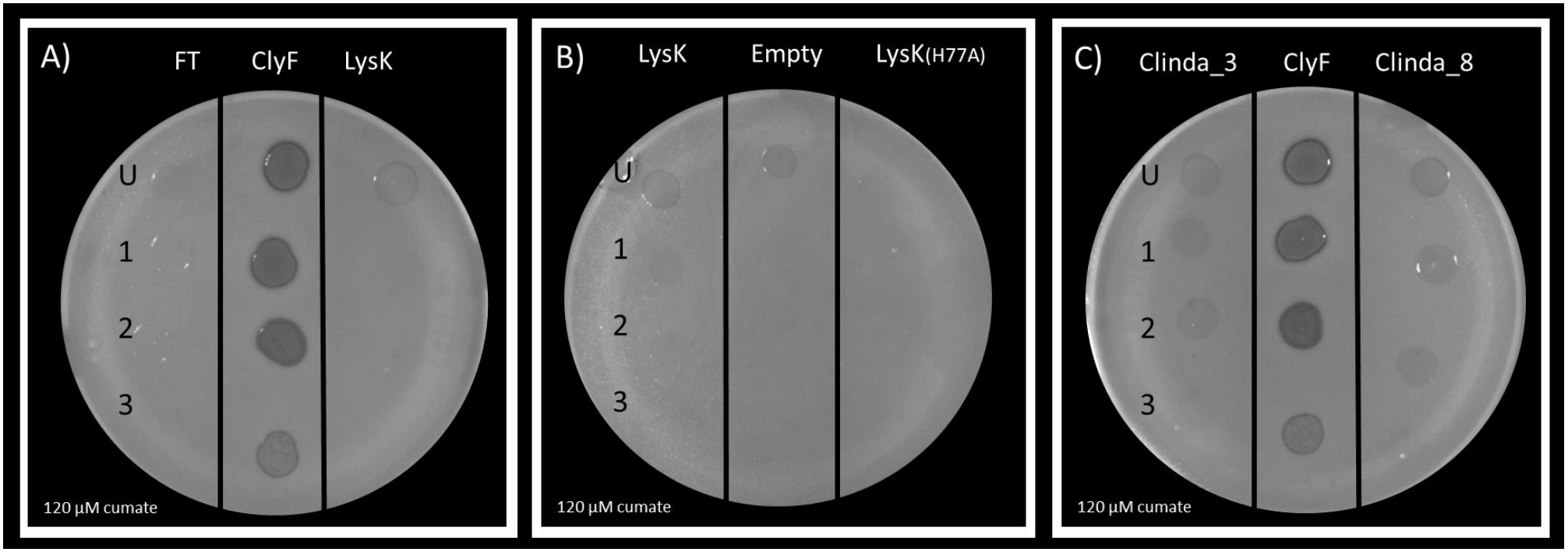

A)
FT
ClyF
LysK
U
1
2
3
120 µM cumate
B)
LysK
Empty
LysK(H77A)
U
1
2
3
120 µM cumate
C)
Clinda_3
ClyF
Clinda_8
U
1
2
3
120 µM cumate

Figure 61: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain C319. *Each plate contains the target strain MSSA S. aureus Strain C319, 120 µM cumate. Plates are divided into three section with either the empty control strain (HB302) or the ClyF expression strain (HB305), as labelled. Expression strains were grown to* OD_{600} *0.6 and a series of one in ten dilutions were made. Samples for each expression strain includes the undiluted sample (U), first (1/10) dilution (1) sample, second (1/100) dilution sample (2) and the final third (1/1000) dilution sample (3). A) Shows the plating results for the First Translation (FT) (HB301) expression strain and LysK (HB296) expression strain compared against the ClyF expression strain (ClyF) (HB305). ClyF shows killing in each of the samples added to the plate. Both FT and LysK show no killing in any of the samples. LysK does however show the lawn disruption effect in the undiluted sample, no disruption or cell death are observed in any of the dilution samples. B) Shows the plating results for the LysK(H77A) (HB306) mutant compared against another biological replicate of LysK (HB296). Plate also includes the empty (Empty) expression control strain which demonstrates the lawn disruption effect in the undiluted sample. This effect is not seen in the additional dilution samples. Both LysK and LysK(H77A) show no killing in any of the samples added to the plate. C) Shows the plating results for the Clinda_3 (HB298) and Clinda_8 (HB294) expression strains compared to the ClyF expression strain. ClyF reproduces the same killing seen in A), in each of the undiluted and dilution samples. Both Clinda_3 and Clinda_8 produces killing in the undiluted samples (U), up to the second dilution (1/100) samples (2).*

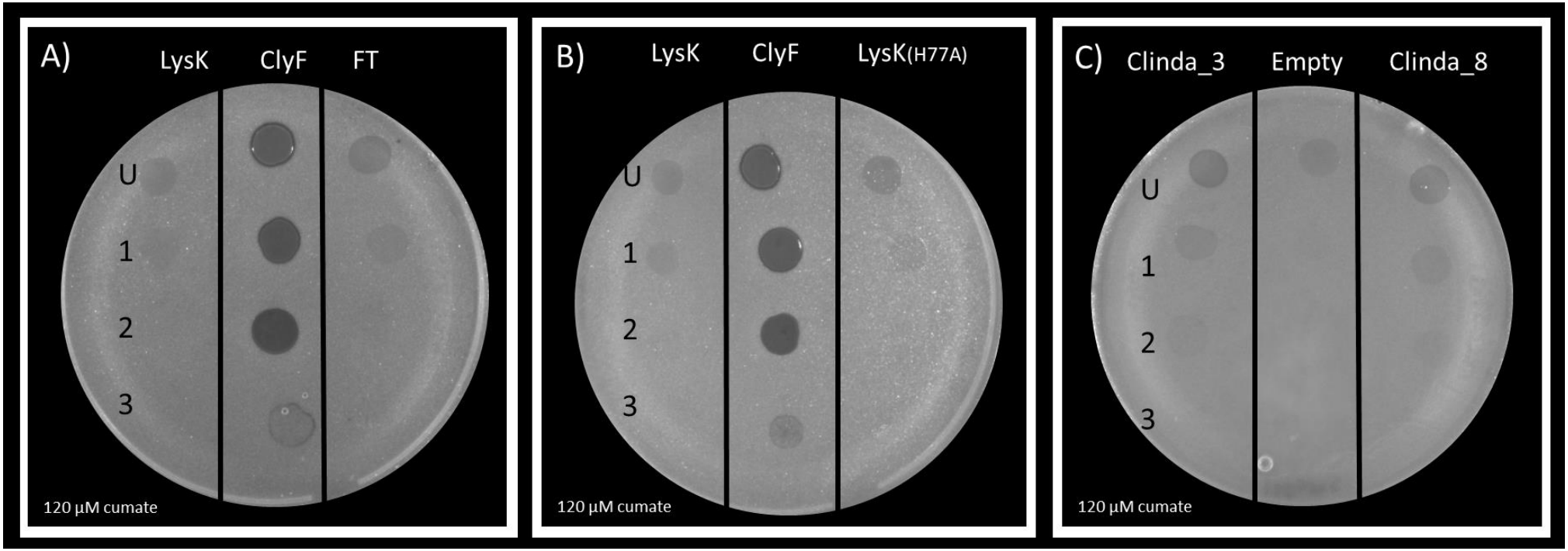

A)
LysK
ClyF
FT
U
1
2
3
120 µM cumate
B)
LysK
ClyF
LysK(H77A)
U
1
2
3
120 µM cumate
C)
Clinda_3
Empty
Clinda_8
U
1
2
3
120 µM cumate

Figure 62: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain ATCC 25923. *Each plate contains the target strain MRSA S. aureus Strain ATCC 25923, 120 µM cumate. Plates are divided into three section with either the empty control strain (HB302) or the ClyF expression strain (HB305), as labelled. Expression strains were grown to* OD_{600} *0.6 and a series of one in ten dilutions were made. Samples for each expression strain includes the undiluted sample (U), first (1/10) dilution (1) sample, second (1/100) dilution sample (2) and the final third (1/1000) dilution sample (3). A) Shows the plating results for the First Translation (FT) (HB301) expression strain and LysK (HB296) expression strain compared against the ClyF expression strain (HB305). LysK and FT results appear to be inconclusive, appearing to show a small amount of plaque formation. But, the killing observed in the first dilution (1) is so faint that it is not conclusive that LysK has activity against ATCC 25923. It is so faint that visual effects like lens glare, which can occur when taking photos, can distort the images. B) Shows the plating results for the LysK(H77A) (HB306) mutant compared against another biological replicate of LysK (HB296). Both LysK and LysK(H77A) show the same appearance of killing up to the first dilution (1). However, like the previous plate, the killing in the first dilution (1) is so faint it is not conclusive, and its appearance could be due to lens glare. ClyF shows killing up until the third dilution (3). C) Shows the plating results for the Clinda_3 (HB298) and Clinda_8 (HB294) expression strains also compared to the Empty control strain. Both Clinda_3 and Clinda_8 produce clear killing in the undiluted samples (U) up to the second dilution (1/100) samples (2).*

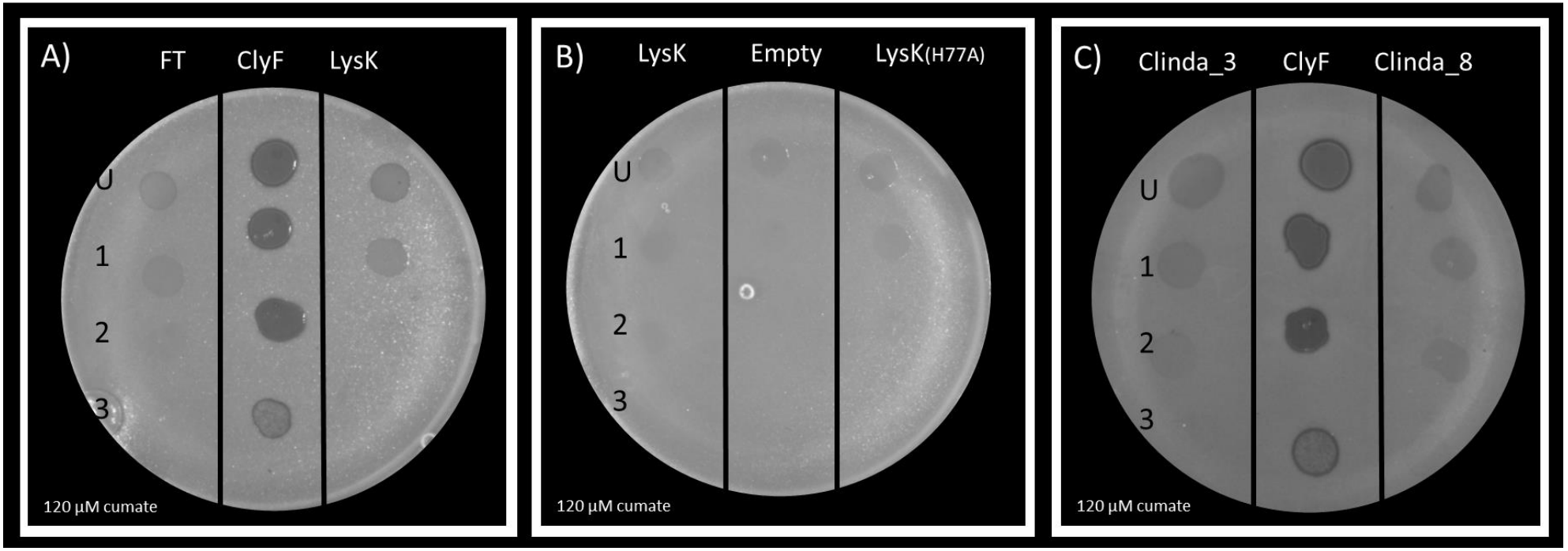

A)
FT
ClyF
LysK
U
1
2
3
120 µM cumate
B)
LysK
Empty
LysK(H77A)
U
1
2
3
120 µM cumate
C)
Clinda_3
ClyF
Clinda_8
U
1
2
3
120 µM cumate

Figure 63: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain C43. *Each plate contains the target strain MSSA S. aureus Strain C43, 120 µM cumate. Plates are divided into three section with either the empty control strain (HB302) or the ClyF expression strain (HB305), as labelled. Expression strains were grown to* OD_{600} *0.6 and a series of one in ten dilutions were made. Samples for each expression strain includes the undiluted sample (U), first (1/10) dilution (1) sample, second (1/100) dilution sample (2) and the final third (1/1000) dilution sample (3). A) Shows the plating results for the First Translation (FT) (HB301) expression strain and LysK (HB296) expression strain compared against the ClyF expression strain (HB305). LysK and FT shows identical results, with faint killing in the undiluted (U) and the first dilution (1/10) sample (1). ClyF expression samples produced killing up until the third (1/1000) dilution (3). B) Shows the plating results for the LysK(H77A) (HB306) mutant compared against another biological replicate of LysK (HB296). Both the LysK and LysK(H77A) results show similar results to the previous LysK results (A). The empty control strain shows the lawn disruption effect in the undiluted sample, which does not occur in any dilution samples. The LysK and LysK(H77A) both show faint killing in the undiluted samples (U), which also occurs in the first dilution (1/10) (1), although it is fainter than the previous results. C) Shows the plating results for the Clinda_3 (HB298) and Clinda_8 (HB294) expression strains compared to the ClyF expression strain. Both Clinda_3 and Clinda_8 produces killing in the undiluted samples (U) up to the second dilution (1/100) samples (2).*

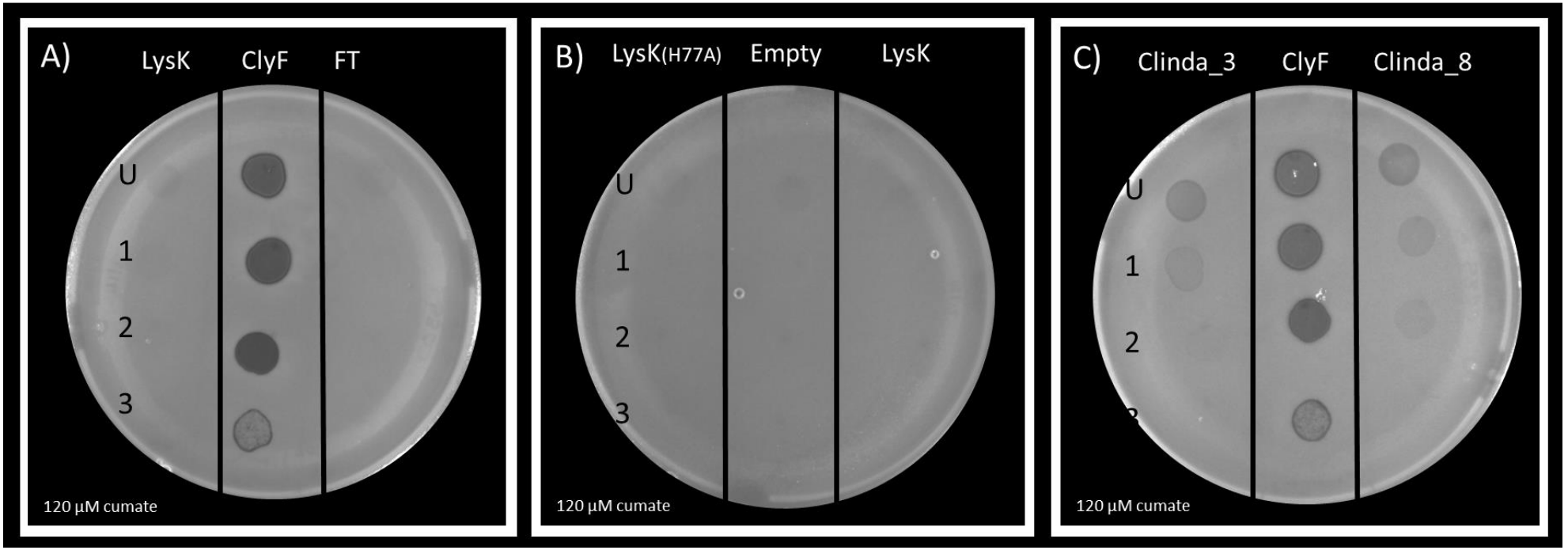

A)
LysK
ClyF
FT
U
1
2
3
120 µM cumate
B)
LysK(H77A)
Empty
LysK
U
1
2
3
120 µM cumate
C)
Clinda_3
ClyF
Clinda_8
U
1
2
120 µM cumate

Figure 64: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain 259. *Each plate contains the target strain MRSA S. aureus Strain 259, 120 µM cumate. Plates are divided into three section with either the empty control strain (HB302) or the ClyF expression strain (HB305), as labelled. Expression strains were grown to* OD_{600} *0.6 and a series of one in ten dilutions were made. Samples for each expression strain includes the undiluted sample (U), first (1/10) dilution (1) sample, second (1/100) dilution sample (2) and the final third (1/1000) dilution sample (3). A) Shows the plating results for the First Translation (FT) (HB301) expression strain and LysK (HB296) expression strain compared against the ClyF expression strain (HB305). LysK is the only expression strain on the plate that shows any lawn disruption in the undiluted sample (U). LysK shows no further killing in any of the dilution samples. FT shows no killing in any sample and only a small amount of lawn disruption in the undiluted sample. B) Shows the plating results for the LysK(H77A) (HB306) mutant compared against another biological replicate of LysK (HB296). Both the LysK results show the same results to the previous LysK results(A). The empty control strain shows a very faint lawn disruption effect in the undiluted sample, which does not occur in any dilution samples. The LysK and LysK(H77A) both show no killing in any sample, and both show no lawn disruption effect in the undiluted samples. C) Shows the plating results for the Clinda_3 (HB298) and Clinda_8 (HB294) expression strains compared to the ClyF expression strain. Both Clinda_3 and Clinda_8 produces killing in the undiluted samples (U) up to the second dilution (1/100) samples (2).*

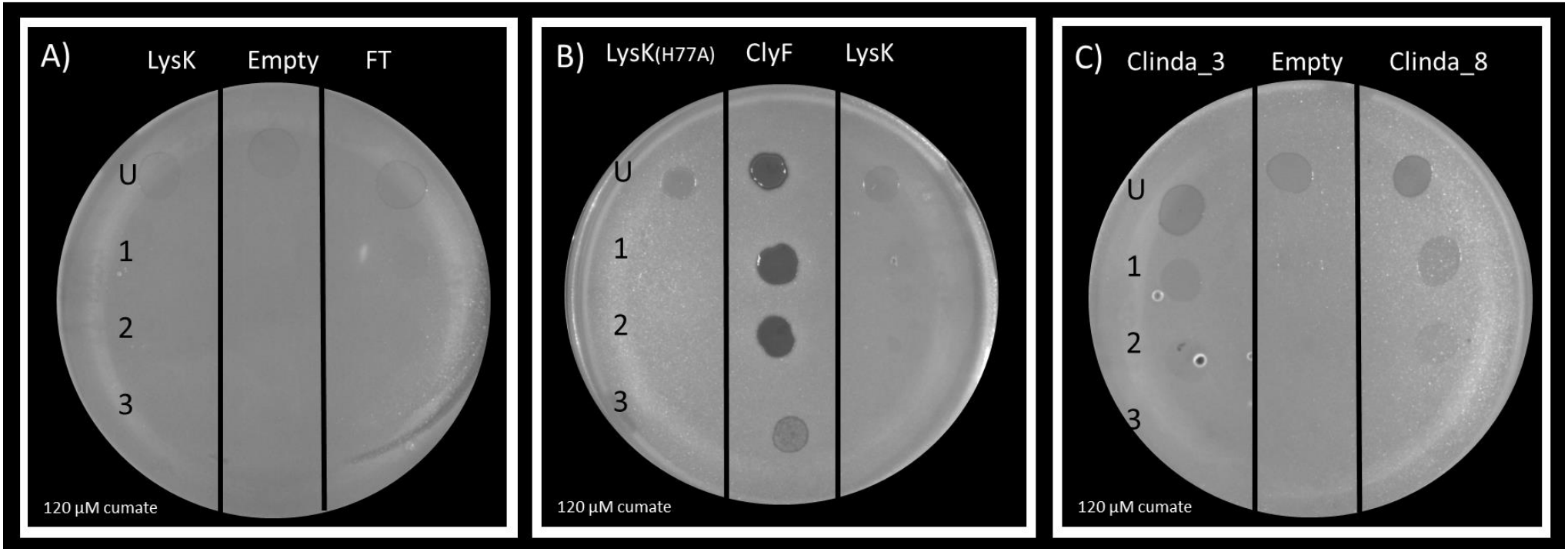

A)
LysK
Empty
FT
U
1
2
3
120 µM cumate
B)
LysK(H77A)
ClyF
LysK
U
1
2
3
120 µM cumate
C)
Clinda_3
Empty
Clinda_8
U
1
2
3
120 µM cumate

Figure 65: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain C330. *Each plate contains the target strain MRSA S. aureus Strain C330, 120 µM cumate. Plates are divided into three section with either the empty control strain (HB302) or the ClyF expression strain (HB305), as labelled. Expression strains were grown to* OD_{600} *0.6 and a series of one in ten dilutions were made. Samples for each expression strain includes the undiluted sample (U), first (1/10) dilution (1) sample, second (1/100) dilution sample (2) and the final third (1/1000) dilution sample (3). A) Shows the plating results for the First Translation (FT) (HB301) expression strain and LysK (HB296) expression strain compared against the empty (Empty) expression strain (HB302). Plating results for the First Translation (FT) (HB301) expression strain and LysK (HB296) expression strain compared against the ClyF expression strain (HB305). LysK shows lawn disruption in the undiluted sample (U) with no further killing in any of the dilution samples. FT shows no killing in any sample and only a small amount of lawn disruption in the undiluted sample. B) Shows the plating results for the LysK(H77A) (HB306) mutant compared against another biological replicate of LysK (HB296). The LysK results are the same as the previous LysK results(A). The LysK and LysK(H77A) both have the lawn disruption effect in the undiluted samples, but no killing is present in any dilution samples. C) Shows the plating results for the Clinda_3 (HB298) and Clinda_8 (HB294) expression strains compared to the ClyF expression strain. Both Clinda_3 and Clinda_8 produces killing in the undiluted samples (U) up to the second dilution (1/100) samples (2).*

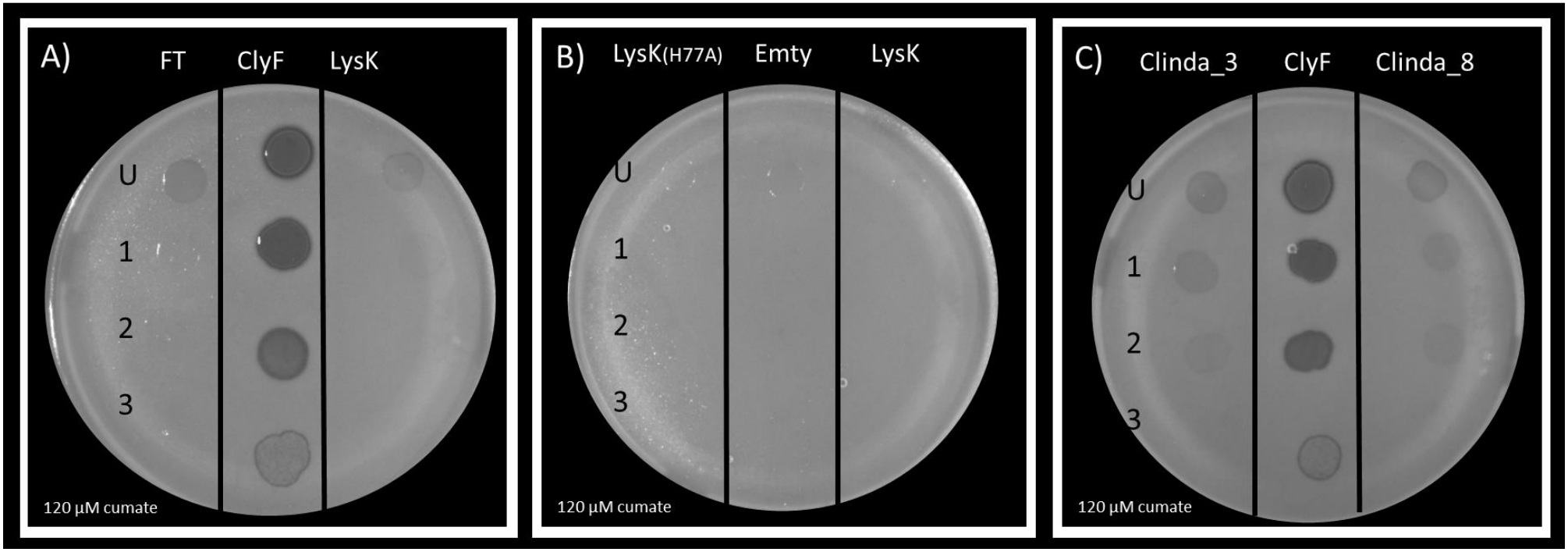

A)
FT
ClyF
LysK
U
1
2
3
120 µM cumate
B)
LysK(H77A)
Emty
LysK
U
1
2
3
120 µM cumate
C)
Clinda_3
ClyF
Clinda_8
U
1
2
3
120 µM cumate

Figure 66: PHEARLESS Assay for Antimicrobial Activity of Selected Proteins Against S. aureus Strain C244*. Each plate contains the target strain MRSA S. aureus Strain C244, 120 µM cumate. Plates are divided into three section with either the empty control strain (HB302) or the ClyF expression strain (HB305), as labelled. Expression strains were grown to* OD_{600} *0.6 and a series of one in ten dilutions were made. Samples for each expression strain includes the undiluted sample (U), first (1/10) dilution (1) sample, second (1/100) dilution sample (2) and the final third (1/1000) dilution sample (3). A) Shows the plating results for the First Translation (FT) (HB301) expression strain and LysK (HB296) expression strain compared against the empty (Empty) expression strain (HB302). Plating results for the First Translation (FT) (HB301) expression strain and LysK (HB296) expression strain compared against the ClyF expression strain (HB305). LysK shows lawn disruption in the undiluted sample (U) with no further killing in any of the dilution samples. FT shows no killing in any sample and only a small amount of lawn disruption in the undiluted sample. B) Shows the plating results for the LysK(H77A) (HB306) mutant compared against another biological replicate of LysK (HB296). Both the LysK results show the same results to the previous LysK results(A). The LysK and LysK(H77A) both show no killing in any sample, and both show no lawn disruption effect in the undiluted samples. C) Shows the plating results for the Clinda_3 (HB298) and Clinda_8 (HB294) expression strains compared to the ClyF expression strain. Both Clinda_3 and Clinda_8 produces killing in the undiluted samples (U) up to the second dilution (1/100) samples (2).*

6.c Testing of chimeric endolysins

6.c.i Construction of chimeric endolysin genes comprising the catalytic domains from putative endolysin P5 and the cell wall binding domain of ClyF

The previous chapter (Chapter 6.a.ii) described the results of assays testing for antimicrobial activity of putative endolysin P5 (PE_GDT_P5). Subsequent literature searches found a study describing the antimicrobial activity of a protein identical in sequence to P5 against a methicillin-resistant *S. aureus* strain (MRSA) (250). While disappointing that P5 was not a novel antimicrobial protein, we wondered whether its relatively weak activity could be improved through protein engineering. P5 was tested in our hands against a non-clinical hosts (RN4220, HER1049), suggesting that while activity is reduced compared to activity against clinical isolates, P5 may have a broad host range. Additionally, with lytic activity beginning to appear only after thirty days, it is likely that P5 has a long half-life. Both these features make P5 a good candidate for optimisation of muralytic activity.

Following P5 expression in the PHEARLESS assay format, target strain clearing became apparent over a long time period (>30 days), and the size of the bacteriolytic halo extended well beyond the location of the applied P5 droplet, suggesting that P5 is able to disperse faster than the positive control ClyF. We posited that P5 disperses faster than ClyF due to the P5 CWB domain (SH3) being unable to bind strongly to the bacterial cell wall. Therefore, we hypothesised that swapping the P5 SH3 CWB domain for the equivalent domain from ClyF (SH3_5) would improve P5 binding to the bacteria cell wall, thereby increasing lytic activity.

To test this, three versions of a chimeric endolysin containing the two catalytic domains of P5 and the CWB domain of ClyF were designed and synthesised (Figure 67). The SH3 domain used to construct the chimeric ClyF protein originated from lysin PlySs2, also referred to as CF-301 (130). The wild type PlySs2 lysin has been characterised to have weak lytic activity, but strong cell wall binding affinity for both streptococci and certain staphylococci (130).

As more artificial chimeric endolysins are generated and different protein linkers are used, it appears that the linkers many have an important role in the interaction between the domains (111,259), with some linkers shown to either contribute to the function of the domains or be critical for activity (111,260). Taking this into consideration, three chimeric endolysins were created using the P5 catalytic domains and the ClyF CWB (SH3) domain. Each chimera was created using a different chimeric junction within the linker region; Chimeric endolysin version 1 (CV1), P5(1-401)-P5(linker)-pLysS2(SH3), used the full P5 linker between the amidase and the ClyF CWB domain (Figure 67, A). Chimeric endolysin version 2 (CV2), P5(1-389)-P5/ClyF(linker)-pLysS2(SH3), used a combination of both P5 and ClyF linker (Figure 67, B) and Chimeric endolysin version 3 (CV3), P5(134)-ClyF(linker)- pLysS2(SH3) was a variant that used the majority of the ClyF linker between the CHAP domain and the SH3_5 domain (Figure 67, C). A fourth control version (CV4), was created which was truncated from the full length P5 construct, comprising the two P5 catalytic domains alone. This control construct was used to determine if P5 catalytic domains could retain their activity without requiring a CWB domain.

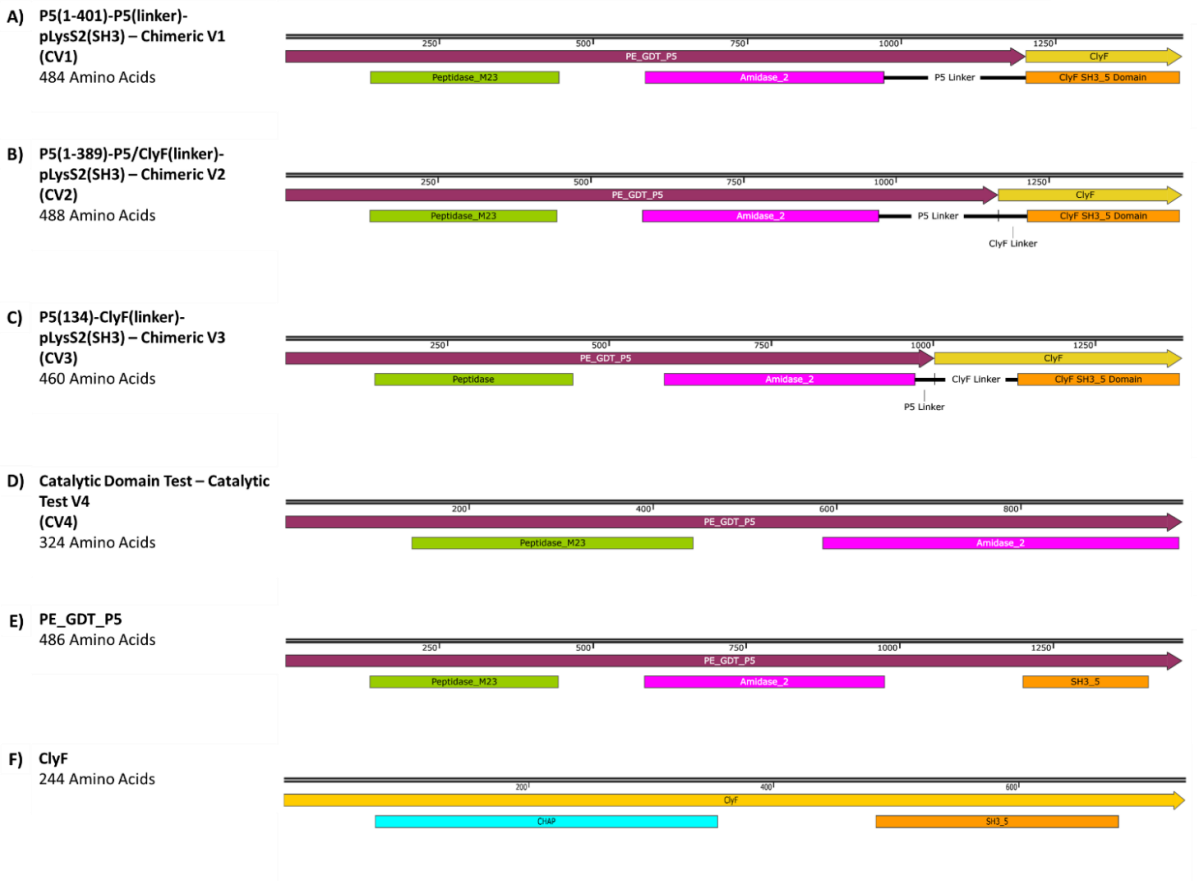


Figure 67: Chimeric Endolysin Protein Maps. Shows the general architecture of the three chimeric endolysin genes and the truncated P5 protein control which only contains the catalytic domains. *Figure also includes the two parent protein architecture. A) P5(1-401)-P5(linker)-pLysS2(SH3) – Chimeric V1 (CV1) is 484 amino acids long, containing the catalytic domains from P5 and the entire P5 linker sequence connecting it directly to the predicted SH3 domain from ClyF. B) P5(1-389)-P5/ClyF(linker)-pLysS2(SH3) – Chimeric V2 (CV2) is 488 amino acids long, containing the catalytic domains from P5 and a section of the P7 linker connected to a section of the linker from the SH3 domain on ClyF. C) P5(134)-ClyF(linker)- pLysS2(SH3) – Chimeric V3(CV3) is 460 amino acids, it contains a small portion of the P5 linker and the entire ClyF linker. Catalytic Domain Test – Catalytic Test V4 (CV4) contains only the catalytic domains of the P5 protein to test if the retain catalytic activity missing a CWB domain. E) PE_GDT_P5 is 486 amino acids long, containing the two catalytic domains, N terminalPeptidase_M23 and Aimidase_2 domains and the CWB domain for P5, SH3_5. F) ClyF is only 244 amino acids long, it contains a sing catalytic domain, a N terminal CHAP domain and a C terminal SH3_5 domain. Figure was generated using SnapGene DNA maps.*

Structures of individual P5 domains have been determined from the study of an identical endolysin protein from *S. aureus* phage 2638A (261) (Figure 68). Both the N-terminal domain, Peptidase_M23, and the C-terminal domain, SH3, have been crystallized (separately) and structures were published on UniProtKB in 2020 (262,263). The Peptidase_M23 N-terminal catalytic domain, which is uncommon in *S.*

aureus phage endolysins has a characteristic Peptidase_M23 family zinc ligand binding site, consisting of His49, Sap53 and His135, located within a nine beta-sheet structure (Figure 68, A). The SH3 (SH3_5) domain the typical structure of six beta-sheets, represented as green arrows (Figure 68, B).

The P5 linker (75 amino acids) (Figure 69, A) is twice as long as the linker in ClyF (43 amino acids) (Figure 69, B), and is predicted to form different secondary structure elements, suggesting they do not act as traditional unstructured linkers. Secondary structure predictions were performed using the online prediction tool PSIPRED (242) on the both parent proteins and chimeric endolysins, using the PSIPRED 4.0 (Predict Secondary Structure) analysis option. PSIPRED predicted that the P5 linker forms two alpha-helices, while the ClyF linker has two beta-strands. P5 domain boundaries were predicted using HMMSCAN, with ten amino acids added to the amidase domain boundary in case the domain boundary exceeded what HMMSCAN predicted.

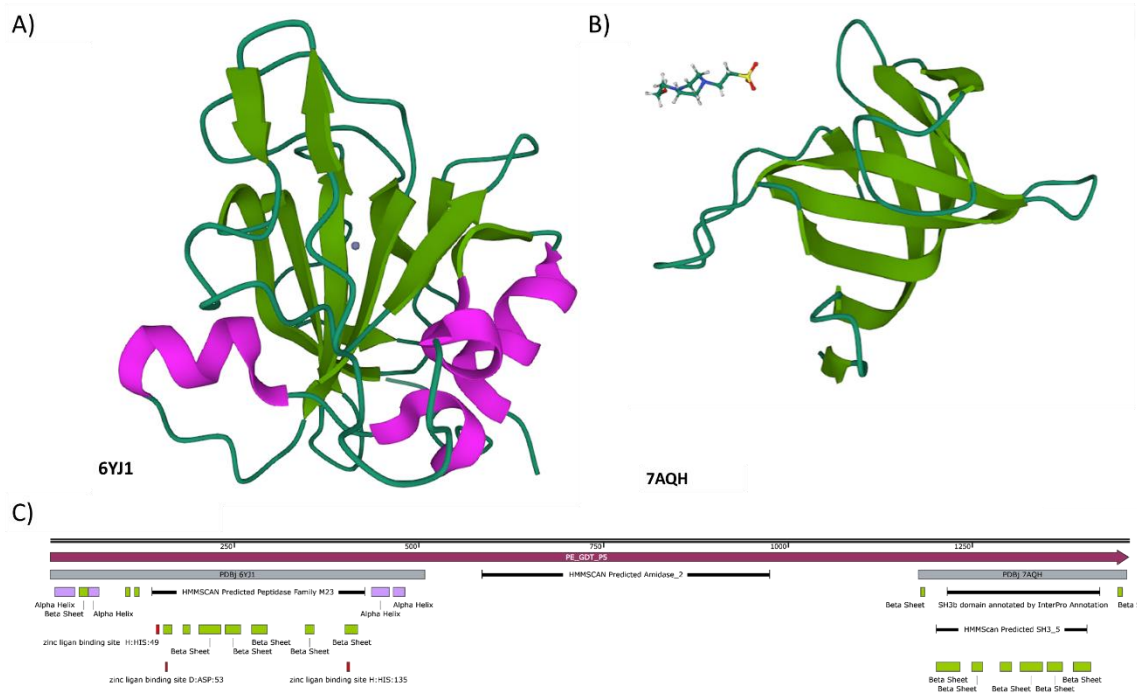


Figure 68: Crystal Structures and Secondary Structure Predictions of Staphylococcal phage 2638A endolysin Peptidase_M23 and SH3 domains. *A) Figure reprinted and adapted from RCSB Protein Data Bank (RCSB PDB), Protein ID 6YJ1 (264). The crystal structure of the Peptidase_M23 domain located at the N-terminus of the P5 endolysin protein (RCSB PDB: 6YJ1), Alpha helices identified in purple and Beta sheets identified as light green arrows. B) Figure reprinted and adapted from RCSB Protein Data Bank (RCSB PDB), Protein ID 7AQH (265). The crystal structure of the SH3 cell wall binding domain located at the C-terminus of the protein, Beta sheets identified as light green arrows. Crystal structure also includes 4-(2-Hydrocyethyl)-1-Piperazine Ethane Sulfonic Acid, located top left*

of the picture. Both crystal structure links can be found on UniProt: Q4ZD58 · Q4ZD58_9CAUD. C) Shows the HMMSCAN predicted domain boundaries (black) and the crystal structure regions (grey) and metal binds sites (Zn) (red). Secondary structure predictions are shown below, colour coded to match the crystal structure images. Figure was generated using SnapGene DNA maps.

CV1 (Figure 67, A) was designed with a fusion point based on an alignment generated with Clustal Omega (186), using the majority of the P5 linker with the ClyF CWB domain at its C-terminus. The predicted secondary structure of the P5 linker incorporated into CV1 comprises three predicted alpha helices and one potential beta-strand (Figure 69, C). This fusion point was chosen based on the HMM boundaries, with the fusion point starting at ClyF SH3 region S-YR-ETGTMTV which shares a similar region of amino acid type and charge with P5 (YKAEHASFTV) located just before P5 SH3 domain boundary. CV2 (Figure 67, B) was designed at an intermediate point, incorporating an N-terminal portion of the P5 linker and a portion of the ClyF linker at the C-terminus. The ClyF linker region of CV2 was chosen as this originated from the PlySs2 lysin from which the ClyF SH3 domain was derived. The third chimeric protein, CV3 (Figure 67, C), included ten amino acids at the C-terminus of the P5 catalytic domains and the entire linker region of ClyF. The predicted secondary structure of both CV2 (Figure 69, D) and CV3 (Figure 69, E) shows three alpha helices.

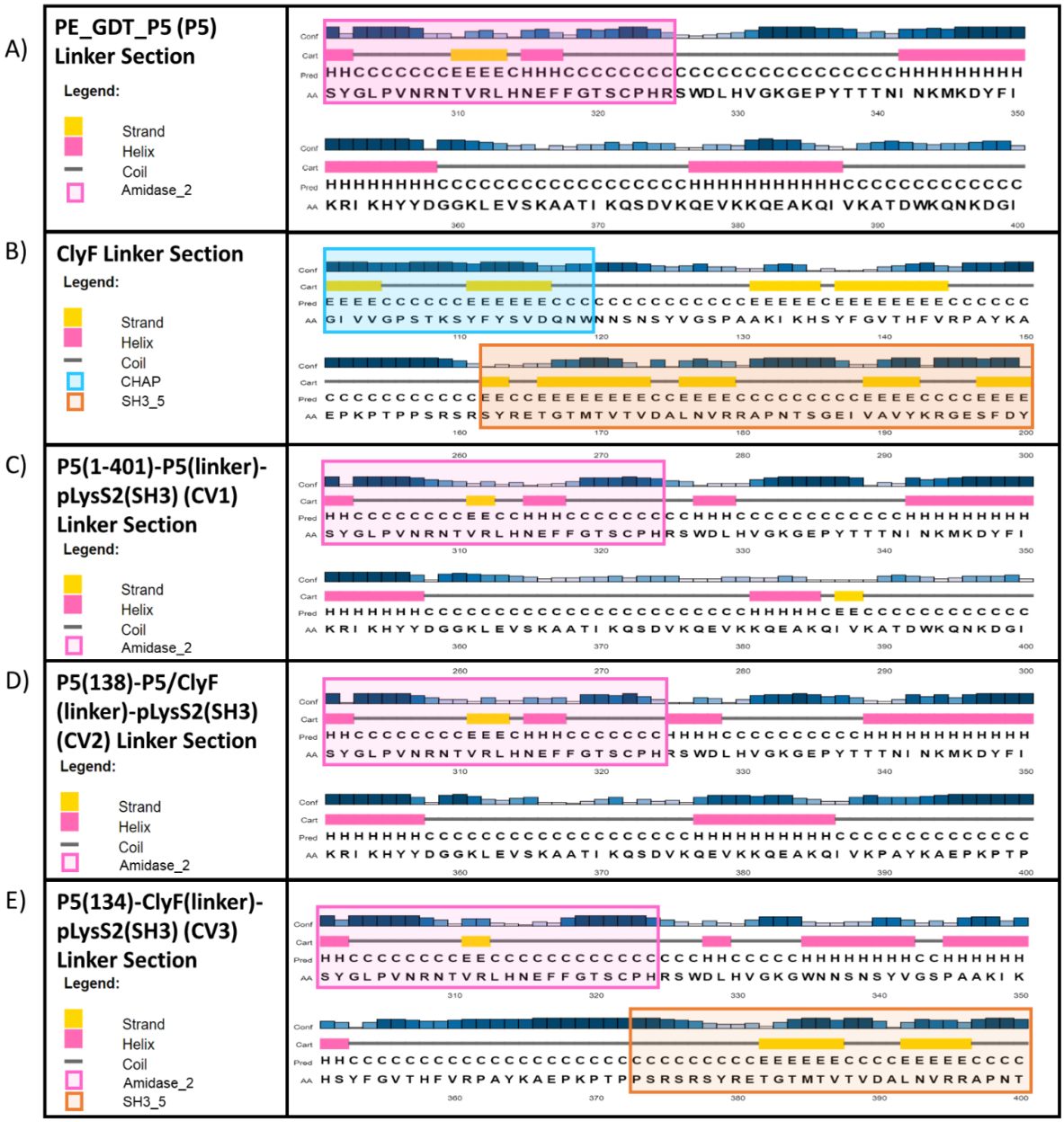


Figure 69: Secondary Structural predictions for the Linker Sequences of the Parent Proteins and Chimeric Protein. *Shows figure adapted from PSIPRED protein structure prediction online tool (242). A) Linker region for one of the parent proteins, P5, containing two predicted α helix. B) Linker region for one of the parent proteins, ClyF, containing two predicted β sheets. C) CV1 which contains the entire P5 linker connected immediately to the SH3 domain from ClyF contains three smaller α helix followed by a short β sheets. D) CV2 which contains a combination of the P5 and ClyF linker shows three predicted β sheets. E) CV3 which contains a small portion of the P5 linker and the entire ClyF linker shows three predicted β sheets in different positions relative to CV2.*

6.c.ii Chimeric Endolysin genes PHEARLESS Results

The PHEARLESS V2.2 system was deployed to test the chimeric endolysins (Figure 67, A, B, C) and the catalytic domain (Figure 67, D). PHEARLESS V2.2 expression

strains were used to ensure protein expression was maximised. Protocols follow the same methods previously used; all strains were grown to OD_{600} 0.6 in LB media, 100 µL and 2 µL of the target (RN4220) and propagation (QC024) strains, were plated on LB agar plates containing 240 µM cumate, then left to dry for a minimum of 1 hour.

New cultures were inoculated from the overnight expression strain cultures. Since the protein is being expressed from a plasmid integrated into the 186 prophage genome, cultures do not require antibiotic selection. This in turn removes the need to pellet and re-suspend the expression strain in antibiotic free media after reaching OD_{600} 0.6. Strains used were the empty strain (HB253), P5 (HB241), CV1 (HB278), CV2 (HB281), CV3 (HB279) and CV4 (HB280). Since P5 was demonstrated to have low endolytic activity, only showing bacterial lysis after about thirty days, these protocols used a method designed to increase the amount of enzyme expressed in an effort to show lytic activity more rapidly, in addition to the advantage gained via phage replication and propagation. After growing the expression strains to OD_{600} 0.6, 10 mL of the cultures were pelleted and resuspended in 1 mL of media, concentrating the undiluted sample tenfold; a one in ten dilution series were created using this concentrated culture and 10 µL spotted onto the lawns. Plates were left to dry at room temperature, then placed at 37°C overnight, and monitored for antimicrobial activity the day after, and sporadically thereafter. Each expression strain was tested on three different cumate concentration plates; 40 µM, 120 µM and 240 µM. Following incubation overnight, the chimeric endolysins and the isolated catalytic domains expression strains all showed identical results to the empty control (Figure 70), indicating they had no antimicrobial activity against the target strain.

Since the “undiluted” sample was 10x the usual concentration of cells, there is a somewhat irregular pattern of growth of the lawn where the undiluted samples were spotted. This irregular growth pattern was observed for each strain on each plate (Figure 70, A, B, C, D). for the undiluted samples.

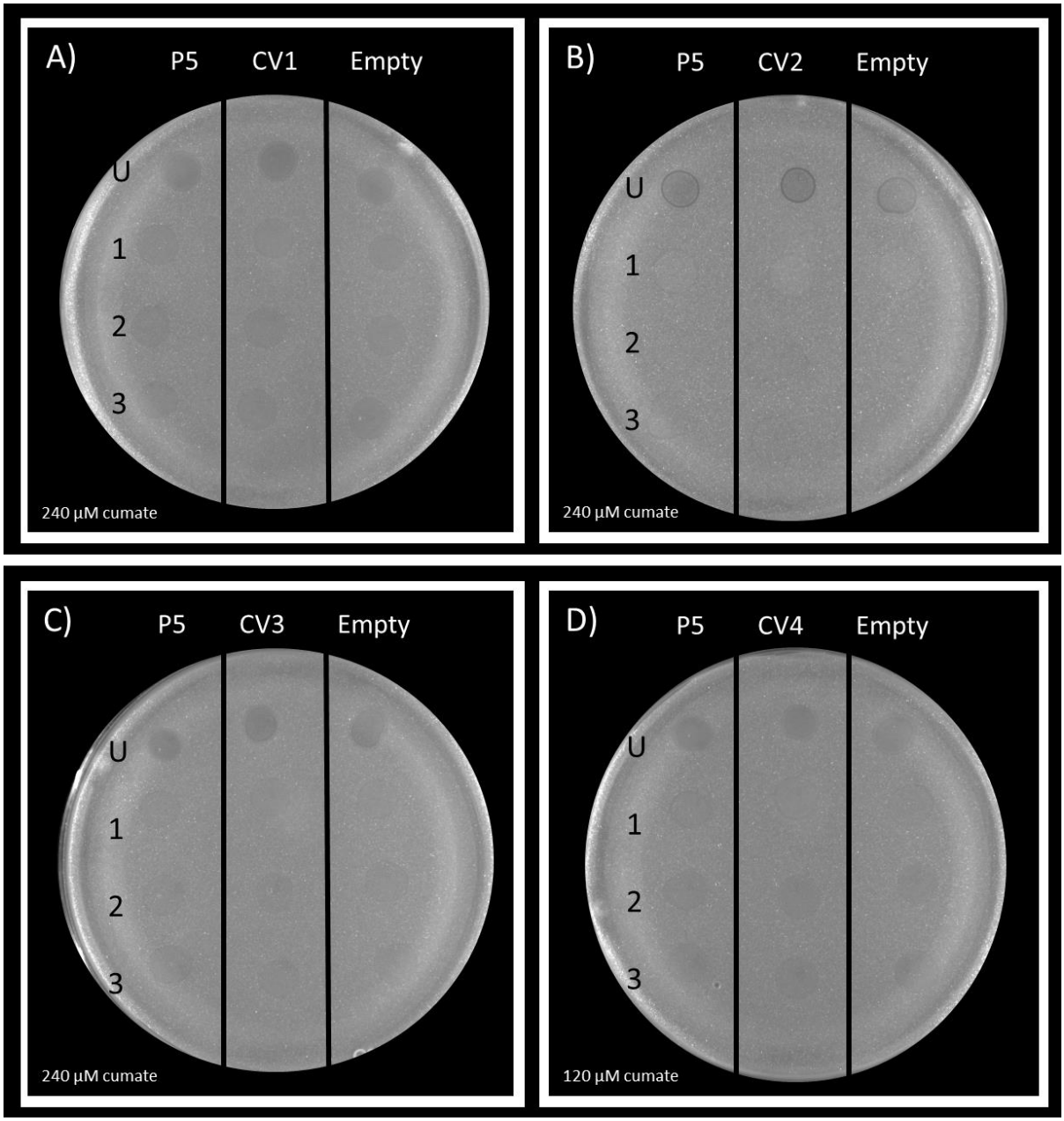


Figure 70: PHEARLESS Protein Screening Assay using the PHEARLESS V2.2 Expression Strain, Testing the Chimeric Endolysins and Catalytic Domains for Antimicrobial Activity Against S. aureus Strain RN4220. *Each plate contains the target strain (RN4220) and a propagation strain (QC024) at a 50:1 ratio respectively. Controls include the negative (Empty) expression strain (HB253), with the P5 expression strain used as the positive control (HB241). A) Contains the initial plating results for CV1 (HB278), which show no clearing. B) Contains the initial plating results for CV2 (HB281), which show no clearing. The undiluted (U) samples for the each of the plates show a slightly different appearance, likely due to the large quantity of expression cells plate. This appearance disappeared in the first (1/10) dilution (1) which would be the undiluted sample if concentrated protocols were not used. Sample after the undiluted sample (U) show little indication of where they were plated. This lawn disruption can be slightly observed in each of the other plates undiluted samples, but it appears the strongest in B). C) Contains the initial plating results for CV3 (HB279), which show no initial plaque*

formation. D) Contains the initial plating results for and CV4 (HB280). All the samples on these plates show no obvious killing after overnight incubation.

Since no initial activity was observed following overnight incubation of all strains (Figure 70) plates were stored at 4°C and reimaged later. Plates treated with 40 µM cumate were chosen for the time lapse photos since this concentration was observed to produce the highest antimicrobial activity following long incubation periods in past assays. Photos were taken of the plates over the period of one month, since muralytic activity of chimeras appeared at this timescale for wild type P5.

The 40 µM cumate plates (Figure 71, A, C, E, F) show identical results to the 240 µM and 120 µM cumate plates (Figure 70) after overnight incubation. The empty controls for the 40 µM cumate plates (Figure 71, A, C, E, F) also show identical results, lacking target strain RN4220 cell lysis.

Photos taken on the 48th day after preparation of the 40 µM cumate plates show that after incubation at 4°C for this period, RN4220 cell lysis is observed for some of the expression strains. The positive control used in this set of plates was wild type P5 where P5 activity appeared after approximately 30 days (6.a.ii Testing selected proteins for activity using PHEARLESS version 2.2). Here the plates were left for an extra 18 days (48 days total) to produce more obvious RN4220 cell lysis. Wild type P5 samples on each plate show similar killing and halo formation as had been seen before, with the strongest clearing occurring in the undiluted samples (Figure 71, B, D, F, H). For plates containing P5 CV1 (Figure 71, B), P5 CV3 (Figure 71, F) and P5 CV4 (Figure 71, H), the undiluted and first dilution (1) samples do appear to have a distinct clearing on day 48 compared to the day 1, however the second (2) and third (3) dilutions do not demonstrate any noticeable clearing. CV2 (Figure 71, B) gives the largest and clearest killing out of all the plates, similar to wild type P5.

CV4, which expresses the two P5 catalytic domains and lacks a CWB domain, was tested to determine if the catalytic domains would be active in isolation. CV4 plate images (Figure 71, G, H) show that after 48 days the dilution series appears identical to the empty control samples, suggesting the loss of the CWBD eliminates catalytic activity (Figure 71, G, H). None of the chimeric endolysins, nor the isolated catalytic domains showed improved activity compared to the P5 samples, clearly showing that adding the CWB domain of ClyF to P5 did not increase activity

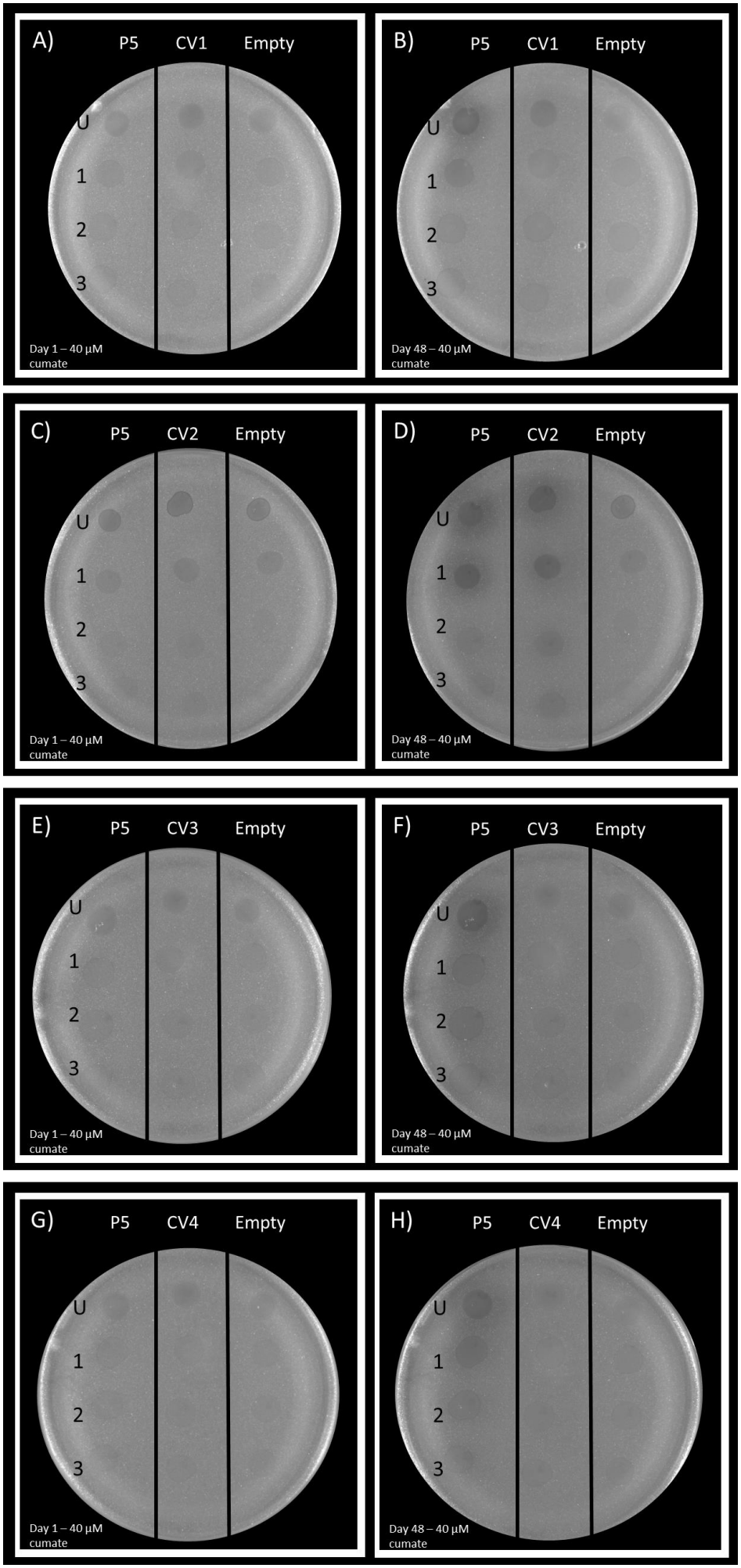

A)
P5
CV1
Empty
U
1
2
3
Day 1 – 40 µM cumate
B)
P5
CV1
Empty
U
1
2
3
Day 48 – 40 µM cumate
C)
P5
CV2
Empty
U
1
2
3
Day 1 – 40 µM cumate
D)
P5
CV2
Empty
U
1
2
3
Day 48 – 40 µM cumate
E)
P5
CV3
Empty
U
1
2
3
Day 1 – 40 µM cumate
F)
P5
CV3
Empty
U
1
2
3
Day 48 – 40 µM cumate
G)
P5
CV4
Empty
U
1
2
3
Day 1 – 40 µM cumate
H)
P5
CV4
Empty
U
1
2
3
Day 48 – 40 µM cumate

Figure 71: PHEARLESS Protein Screening Assay using the PHEARLESS V2.2 Expression Strain, Testing the Chimeric Endolysins and Catalytic Domains for Antimicrobial Activity Against S. aureus Strain RN4220. *Photos of the same plates were taken after overnight A), C), E), and G) or 48 days B), D), F) and H) later. Plates were kept at 4°C wrapped in cling wrap in the interim. Plate images A), C), E), and G) are This set of plates used 40 uM cumate and were prepared using the same cultures that were used to create the 240 µM cumate plates in Figure 70. Each plates contain the target strain (RN4220) and a propagation strain (QC024) at a 50:1 ratio respectively. PHEARLESS Protein Screening Assay protocols used were the concentrated expression culture version. Controls include the negative (Empty) expression strain (HB253), with the P5 expression strain used as the positive control (HB241). The Empty control strain shows no plaque formation after the forty-eight-day incubation, lawn have further recovered from the expression strain being added. Strongest recovery effect appears to have happen in the undiluted sample (U) on plate A)/B) and G)/H). P5 (HB241) on each plate shows strong plaque formation after the forty-eight-day incubation with plaques forming in the undiluted samples (U) and each dilution sample plated. A)/B) shows the long term incubation of CV1 (HB278), CV1 does not show clear indication the plaques have formed in the undiluted sample. The undiluted samples appear almost identical to photo taken of the sample at day one. There appear to be a slight lawn thinning in the undiluted sample at forty-eight-day incubation, this is not observed in any of the dilutions. There is also no lawn recovery effect in any of the CV1 samples. C)/D) shows the long term incubation of CV2 (HB281), CV2 produces identical results to P5 after the forty-eight-day incubation. With plaques appearing in the undiluted sample (U) and each of the dilutions. E)/F) shows the long term incubation of CV3 (HB279), CV3 produced identical results to the Empty control strain. no plaques form in any of the CV3 samples plate, there also appears to be a slight recovery effect for the CV3 samples. G)/H) shows the long term incubation of CV4 (HB280), CV4 was the version that only contained catalytic domains. CV4 produced identical results to the empty control strain with a recovery effect on the lawn also occurring.*

Summarising Chapter 6, seven proteins were chosen from bioinformatics searches, four from the bioinformatic searches performed in Chapter 5 and 3 from a separate bioinformatics study. Only one protein (P5) (Figure 59) showed any lytic activity after one month of incubation at 4°C degrees. P5's two catalytic domains were combined with the CWB domain of ClyF to engineer three chimeric endolysins, aiming to improve lytic activity. Of the three chimeric endolysins, only one chimeric endolysin (CV2) showed lytic activity (Figure 71), but the activity was not increased compared to the P5 parent endolysin. Five protein sequences were tested in collaboration with the BHI. Visible lytic activity was overserved for two proteins (Clinda_3 and Clinda_8) (Figure 60, Figure 61, Figure 62, Figure 63, Figure 64, Figure 65, Figure 66) against

multiple bacterial strains, and results were inconclusive for two proteins (LysK and LysK(H77A)) (Figure 62, Figure 63).

Chapter 7: Experimental Summary and Conclusions

7.a Discussion and Conclusion of the generation of PHEARLESS V1 expression strain and the Generation PHEARLESS Protein Screening Assay

In Chapter 3, each PHEARLESS V1 system component was tested and optimized, aiming for endolysin protein expression from a plasmid. Two interconnected cell lysis modules were successfully developed - a cumate-controlled 186*tum* induction module and an engineered 186 prophage capable of efficient cell lysis, both integrated into the *E. coli* genome (Figure 8).

By removing the *tum* gene from the 186 prophage, 186 cannot enter the lytic life cycle without the expression of Tum from the cumate-controlled Tum expression module (Figure 10, Figure 11, Figure 12). A positive control endolysin protein (ClyF) was expressed from the protein expression plasmid (pZS(^)45) (Figure 13) under the control of a 186 late promoter, which enabled endolysin expression to be controlled by lytic cycle activation upon addition of cumate. Two components of the expression plasmid were tested for optimal protein expression (Table 4); the plasmid origin of replication (high (~70/cell) and low (~3/cell)), and the 186 promoter used (186*pR*, 186*pJ*). In addition to the expression plasmid variants, two Tum switches were tested in this system: wild type Tum and the more active Tum72 variant.

Testing the different combinations showed that for the PHEARLESS V1 expression system, the Tum wild type cumate switch with the high copy expression plasmid carrying the 186*pJ* promoter produced the best results (Table 4, Figure 16). In comparison, while Tum72 activated cell lysis more efficiently than the Tum wild type (Figure 11, Figure 12). Tum72 mediated lysis was perhaps too efficient (Figure 15, Figure 17), resulting in lysis before sufficient protein could be produced.

Initially, 186*pR* appeared to be better than 186*pJ* in the low copy number plasmid, but after increasing the plasmid copy number from ~3 to ~70, 186*pR* proved problematic. The high copy 186*pR* expression plasmid was challenging to assemble, with most of the isolates found to contain mutations. From the range of cumate concentrations tested for antimicrobial activity, 120 µM was found to be the best concentration for observing strong lytic activity using overnight incubation. A lower cumate concentration of 40 µM used for observing weaker endolysin activity, which requires more time to produce viable target cell killing. 40 µM cumate concentration

was also the best concentration for long-term assays that are kept at 4°C for a long period of time.

After testing with ClyF expression and activity in the expression strain variations, the expression strains were tested for phage production. Previous work performed in Shearwin Laboratory indicated that 186 prophages can produce functioning phage particles even if the native *tum* gene had been removed but supplied from an alternative source in the cell (145).

Production of phage particles was shown to be dependent on the presence of an intact cos sequence within the prophage module. Expression strains with a *cos*$^{+}$ prophage and carrying the Tum72 (HB106) and Tum wild type (HB104) produced plaques following addition of cumate, while strains with a Δ*cos* 186 prophage produced no active phage. Whether active phage are produced or not, the system design is able to bring about coordinated protein-of-interest expression and prophage-mediated cell lysis, in response to the simple addition of a small molecule (cumate) to the culture (Figure 7).

7.b Discussion and Conclusion – Chapter 4. Generation of PHEARLESS Version 2.2 Expression Strain and its Application in Library Screening

While the plasmid based protein expression system was successful in being capable of demonstrating endolysin activity, there are difficulties associated with retrieving the gene of interest when the bacteria carrying that plasmid have been killed. A second approach, optimised for screening of libraries of endolysin mutants, was desirable. This a modified version of the PHEARLESS system was designed, whereby the gene of interest is protected by incorporation into the genome of a bacteriophage, allowing the phage to be propagated, retrieved and analysed.

The first aim of Chapter 4 was to confirm whether integration on the endolysin expression plasmid into the 186 prophage genome would still (1) be capable of expression strain cell lysis, (2) produce functional phage particles and (3) produce sufficient ClyF expression. After integrating the pIT4 integration plasmid carrying either 186*pJ* or 186*pR* into either the Tum wild type lysis strain or the Tum72 lysis strain, phage production was tested. Induced and un-induced phage samples were tested against a host strain, which confirmed that the redesigned strain could lyse

upon addition for the chemical inducer, and that the phage particles produced were functional (Figure 25).

After confirming (1) and (2), the same expression strains were tested using ClyF expression (Figure 25, Figure 26) using the same protocol used to test the plasmid-based version in Chapter 3. The 186*pJ* promoter was again the optimal promoter for ClyF expression; expression between the Tum72 and Tum wild type results were similar but in reverse

After confirming aims (1), (2), and (3), the final and most important activity of the system was to test if the progeny phage released could infect a new host, added as a separate *E. coli* propagation strain, to continue the cycle of expression. Five different ratios of target strain to propagation strain were tested to find the correct balance between the two strains (Figure 24). where there was a sufficient amount of propagation strain for good 186 propagation but not enough *E. coli* to out-compete the target strain. The ratio of 50:1, target strain to propagation strain, was chosen as the final ratio concentration.

The 186*pJ* promoter driving expression of the endolysin, in the Tum72 mediated lysis strain was determined to be the best design. After confirming that the new lysis and expression components were functioning as expected, protocols for screening of a mutant library were developed (PHEARLESS Protein Library Screening). A proof of principle assay was developed using a ClyF mutant which had been engineered for loss of activity. Error-prone PCR was used to generate a library likely to include a ClyF revertant mutation, and the system was able to select these from the library.

The sensitivity and optimization of the library screening system for the PHEARLESS V2 system showed several potential limitations of the system that required optimization. An efficient screening system needs to be able to screen as many clones as possible while identifying target cell lysis against a large background of inactive clones. Sensitivity was tested by counting the number of visible plaques on plates with increasing numbers of cells carrying inactive clones. Combining both results (Figure 32, Figure 33), the optimal cell concentration per plate for positive colony observation and maximum number of clones per plate is ~50,000 to ~75,000 cells per plate. Plates that increased above 100,000 negative clones showed an apparent decrease in the number of visible, active clones. A second limitation to

investigate was the transformation efficiency and efficiency of pIT4 plasmid integration into the *att2* site in the 186 prophage.

Initially, an antibiotic selection step was used after transforming the pIT4 Gibson assembly reaction into the expression strain, to select for pIT4 integrants. The individual colonies were collected the following day, diluted, and plated with the target and propagation strains, in the presence of cumate, for visualisation of antimicrobial activity. A design optimization was developed to remove the requirement for this time consuming and inefficient antibiotic selection step.

The 186 *cos* sequence was included on the pIT4 integration plasmid, and a (Δ*cos)* prophage used in the expression strain. Thus, only prophage which carry an integrated pIT4 (*cos*$^{+}$) plasmid should be packaged into phage particles (Figure 34. C). This expectation was confirmed experimentally (Figure 35). While cells which did not carry a non-pIT4 integrant could lyse upon addition of cumate, they could not produce functional phage. This optimization removes the need for a selection step, allowing Gibson assembly mixture to be directly added to the target cell/propagation strain screening plates

The final limitation of the system is the 186 genome size which in turn limits the size of the gene of interest. Increasing the 186 genome size by more than 10% (3,000bp) impairs genome packaging, resulting in unstable or non-functional phage particles. In the final strain, the expression strain plus the pIT4 plasmid without an expression gene, the 186 genome is net + 1,388 bp. With the addition of the ClyF gene, the genome size increases by + 2,120 bp. ClyF only contains a single catalytic domain, unlike common *S. aureus* endolysins, which carry two. Adding a second catalytic domain and linkers would increase the genome size by another 500 to 700bp.

A solution for testing proteins or gene fragments that increase the 186 genome size past the +3.1 Kb threshold would be to remove more non-essential sections of the 186 genome. An analysis of genes known to be non-essential for 186 prophage maintenance or phage production shows that 1219 bp could be removed (266,267). Alternatively, the system could be reconfigured to be based on a larger temperate phage, such as lambda (52kbp), though a different induction system would be required.

A Proof of Principle test for the PHEARLESS Protein Library Screening assay was to recover a loss of function ClyF gene using error-prone PCR. The library of error-prone fragments were assembled into the pIT4 plasmid using Gibson assembly and integrated into the PHEARLESS V2.2 expression strain (E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ*tum, att2,* Δ*Cm*)]) (HB186). The transformation sample was divided and plated onto a mixed ratio lawn of target strain and propagation strain.

The first screen isolated a revertant mutation (G36C), while a second screen also isolated an active phage carrying the revertant mutation (G36C), in addition to a silent mutation (A175A) and a missense mutation (K153R).

In the conclusion of Chapter 4, proof of principle mutant library assays demonstrate that the PHEARLESS Protein Library Screening assay system functions correctly. The protein of interest can be integrated into the 186 prophage, which can then be used as a vector for continuous expression of the protein of interest using a host strain. The final expression strain chosen for this assay is the HB186 expression strain, E4643 [pIT4-KT-cymR-pCym-*Tum72*]$^{\phi 21}$ [186(Δ*tum, att2,* Δ*Cm*)]. HB186 was generated for the HB61 lineage, which cannot produce functional phage particles due to loss of native *cos*. Functional phage production was rescued by re-integrating the *cos* sequence using the new pIT4 backbone (pIT4_*attP*2_LoxP_S_186*pJ*_ClyF_cos+, pIT4_*attP*2_LoxP_S_186*pJ*_cos+). This optimization was key in removing a selection step from the protocols.

Many of the different components of the system were evaluated and optimized, including the lawn strain, pIT4 integration and ClyF expression, assay sensitivity and 186 genome size limitations. While the PHEARLESS V2.2 assay can potentially be used to look at improved activity, depending on the wild type activity of the protein being investigated, it may be challenging to determine if a particular mutant has modest improvements in activity. The size of the killing zones size could be potentially be measured using imaging software and used as a proxy for improvement in activity.

7.c Discussion and Conclusion Chapter 5 - Bioinformatics Search and Screening of Putative Lytic Proteins in the PHEARLESS Assay System

7.c.i Bioinformatics Search

In Chapter 5: Bioinformatic Pipeline for Putative Antimicrobial Phage Protein Discovery, two large data sets were generated from a bioinformatics search aiming to identify phage proteins predicted to be present in several antibiotic-resistant *S. aureus* strains. One set was provided by the ENT surgery group at the Basil Hetzel Institute of *S. aureus* genomes collected from patients with chronic rhinosinusitis (189). The second data set consisted of *S. aureus* genomes isolated from patients with cystic fibrosis (BioProject accession number PRJNA480016). These bacterial genomes were searched for prophage sequences using an online tool, PHASTER. The resulting hits were analysed for domains of interest, discovered using an online tool, HMMSCAN, which searches using profile hidden Markov Models. From both lists, almost 14,000 prophage related proteins were identified, and the number of proteins predicted to have a potential antimicrobial function was filtered down to 1,058 individual proteins (Table 6, Table 7). At the conclusion of Chapter 5, two data sets were generated that contained a list of predicted phage proteins that could contain antimicrobial activity—providing a database of potential proteins of interest that can be investigated for antimicrobial activity.

7.c.ii Discussion and Conclusion Chapter 6 PHEARLESS Assay of Proteins Selected from Bioinformatics Search

In Chapter 6, four proteins discovered in the bioinformatics searchs (Chapter 5) were investigated for lytic activity, in addition to three other proteins from a published bioinformatics study (232), which investigate *Staphylococcal* phage evolutionary relationships and genomic diversity. Proteins were reviewed for previous functional and characterization studies performed on identical proteins or proteins with high homology. After performing the PHEARLESS Protein Screening assay, one of these proteins was discovered to have been previously tested for lytic activity (P5). Out of the seven proteins tested, P1 to P7, P5 was the only protein that showed lytic activity against one (RN4220) of the two laboratory *S. aureus* strains, RN4220 and HER0149 tested (Figure 57, Figure 58). While P5 showed lytic activity against RN4220, the activity was very weak, only observed on the plate after 30 days (Figure 59), with the most evident activity observed on plates that had been stored and sealed at 4°C for 48 days (Figure 71).

Once P5 activity appeared, it did not clear the target bacteria completely like the positive control ClyF, but did noticeably reduce the density of target cells in that

location, compared to the rest of the lawn. A second noticeable difference between ClyF and P5 was that P5 appeared to diffuse further through the plate (Figure 59, Figure 71).

The most likely reason that none of the other tested proteins showed activity is strain specificity. Most of the phage these proteins originated from were clinical, antibiotic-resistant *S. aureus* strains. Therefore, it is most likely that the two *S. aureus* strains used in the assays are outside these protein's target range.

Between P1 and P3, P1 is the only predicted protein with the architecture of an endolysin protein: N terminal CHAP, Amidase_3 and a C terminal SH3_5 domains. Five of the seven proteins contain a typical endolysin architecture (N-terminal CHAP, Amidase_3 and a C-terminal SH3_5 domains): P1, P4, P5, P6 and P7. Nevertheless, given that these proteins are from predicted prophage genomes, it is possible that they are not functional and have lost the ability to lysis the host cell. P2 and P3 do not present typical endolysin architecture and are more likely to function as structural proteins, such as tail-associated cell wall hydrolases and could therefore require additional phage genes to aid assembly and activate the lytic function.

The last protein, P7, did not show any activity against the laboratory strain, RN4220, either initially or after thirty days. However, previous studies demonstrated that a homolog of this protein originates from a phage (SA97) capable of infecting RN4220 (252). There are a few possible reasons why our results do not match those observed in the previous studies. One possibility is that the RN4220 strain used in this study has been misidentified. Some laboratories may receive their bacterial strain from another laboratory that got there through another laboratory, and so on. Receiving a new RN4220 clone or the RN4220 clone used in the 2015 study could be performed to re-test our protein. Sequencing could be performed on our RN4220 strain, but that would be costly and time-consuming. It could also be possible that our RN4220 stain has evolved into a strain of RN4220 that is resistant to the endolysin now.

7.c.iii PHEARLESS Assay Results and Discussion for Proteins Tested in Collaboration with the Basil Hetzel Institute

The PHEARLESS Protein Screening system was also used in a collaborative project with the ENT group at the Basil Hetzel Institute (BHI), using the PHEARLESS V1

expression system. The main proteins of interest being tested were discovered from a phage species that was unable to infect a clinical *S. aureus* strain, but was found to be capable of infection if the host cell was treated with antibiotics. The progeny (exit) phage were then shown to be capable of infection of the original host strain without the use of antibiotics. Genome sequencing of the exit phage showed a single mutation in the endolysin gene, compared to the parent (entry phage) genome. This mutation indicated removal a section from the endolysin gene, the entire amidase domain located in the middle. The detailed mechanism of this phage-antibiotic synergy in under investigation, but my interest was in assessing the activity of the endolysin variants.

Five pZS(^)45 expression plasmids were constructed; one contained a truncated version of the parental LysK endolysin (FT), translated up until the first premature stop codon in the amidase domain (Figure 60, B). Two full length LysK proteins were tested, one identical to the initially identified phage parental endolysin, containing an HNH endonuclease mutation (Figure 60, A, F). In this study, the phage parent endolysin containing an endonuclease domain assembled with a mutation in the HNH endonuclease (H77A), because the active HNH endolysin variant could not be cloned. The second LysK (Figure 60, E) was another wild type endolysin containing the final protein architecture lacking the endonuclease that also exists naturally in nature (258). The final two expression plasmids created were the progeny phage endolysins, which both have had the HNH endonuclease domain and NUMOD4 motif and entire amidase domain removed to regenerate a CHAP-SH3 architecture. Both proteins are identical except for six nucleotides (2 amino acids) where the domain fragments are fused (Figure 60, C, D) (Clinda_3 and Clinda_8).

All five proteins were tested against six clinical isolates *S. aureus* strains. An empty control strain acted as the negative control, while ClyF was a positive control which showed strong activity against all target strains (Figure 61, Figure 62, Figure 63, Figure 64, Figure 65, Figure 66). The exit phage endolysin proteins, Clinda_3 and Clinda_8, each showed consistent activity against all strains tested, whereas the entry phage variants of LysK had no activity.

Modification of the assay to use 10-fold higher density of expression cells did not indicate any endolysin activity for the entry phage variants. There is the possibility that even this concentrated version of the assay is not sensitive enough to detect

weak activity over a short (overnight) time frame. One possible way to increase sensitivity would be to re-develop the PHEARLESS assay system for liquid cultures, optimizing the assay for that setting. The advantage of this adaptation would be the ability to monitor OD_{600} over time, and hence follow the kinetics of endolysin activity. This will be a project for another member of the Shearwin lab in the future, and initial results look promising.

7.c.iv PHEARLESS Assay Results and Discussion for Proteins for Chimeric Endolysins

In Chapter 6.c, the aim was improving P5 antimicrobial activity against RN4220 by protein engineering. Chimeric endolysins were generated using an SH3 domain from ClyF, which has strong activity against RN4220. The hypothesis was that the P5 CWB (SH3) domain does not bind to the RN4220 cell wall efficiently, and that replacing the P5 CWB domain with the SH3 from ClyF would increase lytic activity. Given the very weak activity of P5, improved activity should have been straightforward to identify.

Controls showed that the P5 catalytic domain alone is non-functional (Figure 70, D). A small number of variants were generated, varying in the position of the linker between the domains. Only one of the variants showed lytic activity against RN4220, producing the same lytic activity as P5 wild type, taking the same time to appear on the plate. Thus, the aim of improving the lytic activity of P5 by adding a different CWB domain was not achieved. In the future, a more extensive array of chimeras could be generated, perhaps based on the VERSATILE approach to variant production (144), and tested in a higher throughput manner using a liquid culture based PHEARLESS assay.

In conclusion of this thesis, the PHEARLESS assay system was successfully constructed as a novel assay system for both antimicrobial protein screening and library screening. It is an innovative system that utilises the properties of a prophage in a significantly novel way, contributing to the current knowledge of how phage can be used in synthetic biology. The system's advantages include; low cost in setting-up and running, fast testing, parallel protein testing, protein solubility is not a limiting factor, high throughput library screening and positive clone genotypes can be easily retrieved and re-tested. The limitations of the system, is its capability to screen protein with medium to weak lytic activity. After generating and validating both assay

systems, the second area of study for this thesis was addressed. The bioinformatics pipeline, was completed and able to filter 13,737 putative prophage proteins down to 1,058 proteins that were predicted to have antimicrobial activity. Four of the proteins discovered in the bioinformatics search were tested in the PHEARLESS assay system, as well as twelve other proteins. Antimicrobial activity was confirmed for five proteins, four of which were novel proteins with no previous functional studies performed on them. For The one protein that was found later not to be novel (P5), new avenues of research were started, including aiming to improve the antimicrobial activity of P5 by using it to construct a chimeric endolysin, which were screened for improved activity in the PHEARLESS assay. The work performed in this thesis has contributed to several areas of current scientific knowledge in molecular biology, microbiology biology, synthetic biology, assay development, bioinformatics and antimicrobial resistance mitigation.

Chapter 8: Future Directions

8.a.i Adapting the PHEARLESS Assay System for Liquid Culture Screening.

In the development of the PHEARLESS assay, it was observed that although it was sufficient for showing antimicrobial activity for strong proteins, proteins with medium to weak strength were more challenging to detect. This weak activity can be inconsistent and appear similar to the batch-to-batch variation observed in the negative controls due to growing and lysing the expression strain in the same location as the target species.

All PHEARLESS experiments have been performed on solid media (agar plates) except for growth curve assays performed in Chapter 3. Solid media was chosen to allow simple visualisation of differences between the expression strain spot and background where no expression strain is added.

An optimised version of the assay using liquid media is proposed to improve the sensitivity of the PHEALESS assay for proteins with weak activity. Using a liquid media, with shaking in a temperature controlled environment in microtiter plates (e.g. 96 wells), could improve the lysis and dispersion of the protein of interest, and allow kinetic monitoring of lysis in a relatively high throughput manner.

Certain modifications would be required to move the assay to a liquid-based platform, depending on whether it was based on PHEARLESS version 1 or 2. Version 2 has the advantage of being able to harvest phage carrying the gene of interest from the wells at the conclusion of the assay. The final liquid cultures testing protein activity would contain a mixture of expression cells and target bacterial cells, plus phage propagation strain, if based on version 2.

Given the reduced volumes (~150 µL) and different availability of oxygen in the microtiter format, the choice of Tum- or Tum72-based lysis would need to be tested, as would the most appropriate concentration of cumate required to initiate lysis. The plate assay would be set up to use normalisation of growth curves based on negative control samples, which would contain the expression strain alone, or target strain alone. Controls with and without cumate induction would also be included. With an appropriate plate reader, it should be possible to separately follow growth of target

and expression cells, if strains were each labelled with a fluorophore, such as expressing a different fluorescent protein.

8.a.ii Adapting the PHEARLESS Assay System for Environmental Library Screening

Microorganisms are the most abundant source of genetic diversity on the planet, making them a rich source of novel proteins and enzymes with various metabolic functions, binding structures and physiological properties. Therefore, they are a significant resource for potential pharmaceutical molecules—several challenges affect our ability to find these unique and desirable proteins. One is their abundance. Identifying and testing them one by one would be impossible.

A second major challenge when searching for these novel proteins is whether the bacteria they originated from can even be cultured. It is predicted that for the current *in vitro* growth of bacteria, only to 1-2% of the bacteria species found on Earth can be cultured (268,269).

An issue that can increase the difficulty of identifying desired proteins is that no other proteins with high homology have been functionally investigated. Metagenomics library screening is a tool to accelerate the discovery of novel proteins for different catabolic processes using high throughput screening. Metagenomics libraries are genomic libraries generated from all microorganisms in an environmental sample. The environmental samples are taken directly from the environmental source, including all the microorganism species (bacteria, viruses, fungi). These samples are taken directly from the source and do not undergo culturing, producing a library that includes genomes from microorganism species that are currently ‘uncultivable’ using current culturing methods.

Metagenomics library screening is accomplished using three key steps (270). The first step involves isolating and purifying DNA found in the environmental samples. DNA collected needs to be purified, removing any co-purified polyphenolic compounds during DNA purification. The second step follows the construction of the DNA library into a cloning vector and host strain. Methods for this step differ depending on the size of the final DNA fragments. The third step for metagenomics library screening is either a sequenced-based analysis or a functional analysis.

The functional analysis pathway can be difficult to perform since the active genome needs to be retrieved from the expression host. This can restrict the type of genes being screened. For example, antibiotic resistance genes are more readily screened for, with successful clones being capable of growing on antibiotic medium. Single colonies of a successful clone could be retrieved from the media and further cultured. Functional tests that require expression and release would require individual copies of the colonies kept and stored so the genomes can be retrieved after testing. This significantly reduces the high throughput testing, requiring individual selection and growth of each clone. This can be achieved with the use of colony or plaque picking using automated robotics. The PHEARLESS assay, particularly version 2, where the gene of interest is packaged and protected within a functional phage particle, is designed to be adaptable to high throughput library screening.

While phage-derived proteins have been shown to be highly effective against specific bacterial species, they have advantages and disadvantages. One advantage is being capable of targeting a specific disease-causing bacterial species while leaving the bacterial fauna unaffected. The disadvantage is that the target of disease-causing bacteria would need to be determined, and there could be more than one species or strain which causes the infection. Determining the pathogenic bacteria would increase the waiting time before treatment, causing risk to the patient's life. A solution to this problem would be using cocktails of protein based drugs, with multiple proteins that target different species, aiming for species overlap. This could also reduce the rate of resistance development by bacteria against the treatment. To achieve this, the discovery and optimization of many antibacterial phage proteins to cover a broad host range is needed, and ideally the PHEARLESS system described here may play a role.

Chapter 9: Appendix Data

Appendix Fig. 1: G-Block version 1, 186 recombineering fragment.

CCATATCATGTACATACAGTGTATTTAACTGTGATTTTTTTCGGCGTGGCGGGGAAAGCATT

186 Sequence Overlap — 186 cos End

GCGTTTGTAAAGGAGACTGATAATGGCATGTACAACTATACTCGTCGGTAAAAAGGCATCTT

attB3 Sequence

ATATTCCCCGAATCCGGTTTAAACCGGAGGACGGCGCAGAAGGGGAGTAGCTCTTCGCCGGA

attB2 Sequence

CCGTCGACATACTGCTCAGCTCGTCCGGATATAAGTAATTCCTCGTAAGTTACCGTTCGTAT

LoxP RE

AATGTATGCTATACGAAGTTATGATATCTGGCGAAAATGAGACGTTGATCGGCACGTAAGAG

Chloramphenicol Resistance Gene

GTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGA

TTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATTACTGGGTATACAACCGTAGACAT

TTCGCAGTGGCACCGCAAGGAACATTTTGAAGCGTTCCAGTCGGTGGCGCAATGCACATATA

ACCAAACTGTCCAATTGGACATTACTGCGTTTTTGAAGACTGTCAAGAAAAATAAACACAAG

TTCTACCCTGCGTTTATTCATATCCTGGCTCGTCTGATGAATGCACACCCTGAGTTCCGCAT

GGCCATGAAGGACGGCGAATTGGTTATCTGGGACAGCGTTCATCCCTGCTACACTGTGTTTC

ACGAGCAGACTGAAACATTCTCCTCACTTTGGTCAGAGTACCATGACGACTTTCGCCAATTC

TTACATATTTATTCGCAGGACGTAGCCTGTTATGGAGAGAATCTTGCGTATTTTCCGAAGGG

GTTCATCGAGAACATGTTTTTTGTGAGTGCGAACCCTTGGGTGTCATTTACGTCGTTTGATC

TGAACGTGGCCAATATGGATAACTTCTTTGCCCCTGTGTTCACTATGGGCAAGTACTACACT

CAAGGAGACAAAGTTCTGATGCCTCTTGCTATCCAAGTACACCACGCAGTCTGTGATGGCTT

CCACGTTGGGCGTATGTTGAATGAACTGCAACAGTACTGCGATGAATGGCAGGGCGGGGCTT

AATAAGAATTAGGAAGATAACTTCGTATAATGTATGCTATACGAACGGTACGTGTGGCGCGT

LoxP LE

CTGAGACGTGATGGTGGCGGGGTATGAA

186 Overlap Sequence

First G-Block designed to delete p95-Tum-orf97-t97, insert two Voigt att sites (att2B and att3B) (peach) and loxP sites (pink) flanked chloramphenicol resistance gene cassette (yellow). Codon optimises the chloramphenicol resistance gene to remove commonly used restriction sites. The start and end of the G-Block have an overlap section (underline), which matches the 186 genomes. The section directly after the first 186 overlap sequence matches the end of the 186 cos region (green). All strains originating from the HB61 lineage contain this G-Block.

Appendix Fig. 2: G-Block version 2, 186 recombineering fragment.

CCATATCATGTACATACAGTGTATTTAACTGTGATTTTTTTcGTTTGTAAAGGAGACTGATAATGGCA

186 Sequence Overlap

TGTACAACTATACTCGTCGGTAAAAAGGCATCTTATattccccgaatccggtttAaAccggaGGACGG

attB3 Sequence

CGCAGAAGGGGAGTAGCTCTTCGCCGGACCGTCGACATACTGCTCAGCTCGTCcggatataagtaatt

attB2 Sequence

cctcgtaagttaccgTTCGTATAATGTATGCTATACGAAGTTATGATATCTGGCGAAAATGAGACGTT

LoxP RE Chloramphenicol Resistance Gene

GATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGA

GTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATTACTGGGTATACAACCGTAG

ACATTTCGCAGTGGCACCGCAAGGAACATTTTGAAGCGTTCCAGTCGGTGGCGCAATGCACATATAAC

CAAACTGTCCAATTGGACATTACTGCGTTTTTGAAGACTGTCAAGAAAAATAAACACAAGTTCTACCC

TGCGTTTATTCATATCCTGGCTCGTCTGATGAATGCACACCCTGAGTTCCGCATGGCCATGAAGGACG

GCGAATTGGTTATCTGGGACAGCGTTCATCCCTGCTACACTGTGTTTCACGAGCAGACTGAAACATTC

TCCTCACTTTGGTCAGAGTACCATGACGACTTTCGCCAATTCTTACATATTTATTCGCAGGACGTAGC

CTGTTATGGAGAGAATCTTGCGTATTTTCCGAAGGGGTTCATCGAGAACATGTTTTTTGTGAGTGCGA

ACCCTTGGGTGTCATTTACGTCGTTTGATCTGAACGTGGCCAATATGGATAACTTCTTTGCCCCTGTG

TTCACTATGGGCAAGTACTACACTCAAGGAGACAAAGTTCTGATGCCTCTTGCTATCCAAGTACACCA

CGCAGTCTGTGATGGCTTCCACGTTGGGCGTATGTTGAATGAACTGCAACAGTACTGCGATGAATGGC

AGGGCGGGGCTTAATAAGAATTAGGAAGATAACTTCGTATAATGTATGCTATACGAAcggtaATGCCA

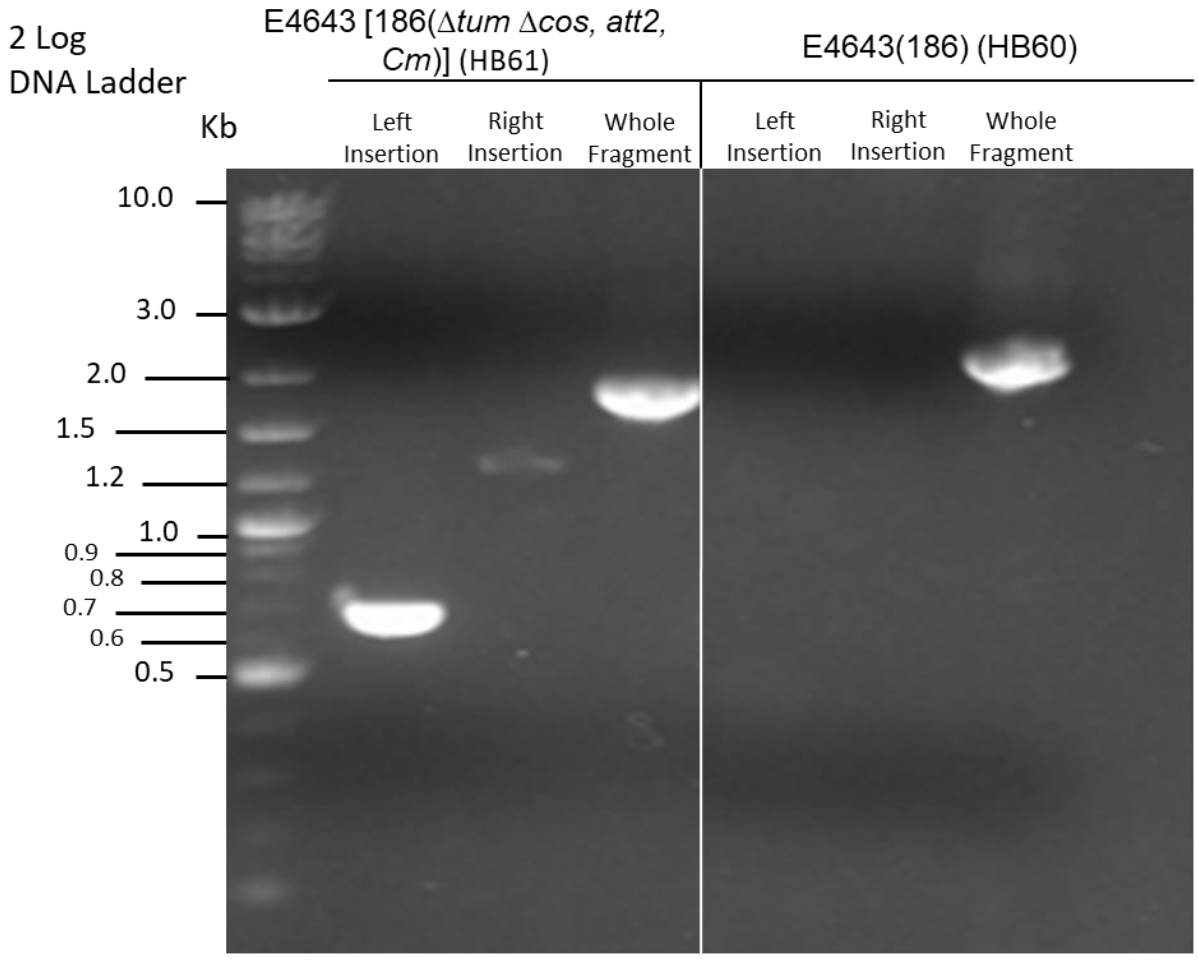
LoxP LE

CTTTTCCACGAATTTCCCGTTTTTTTAGCCGTGC

186 Overlap Sequence

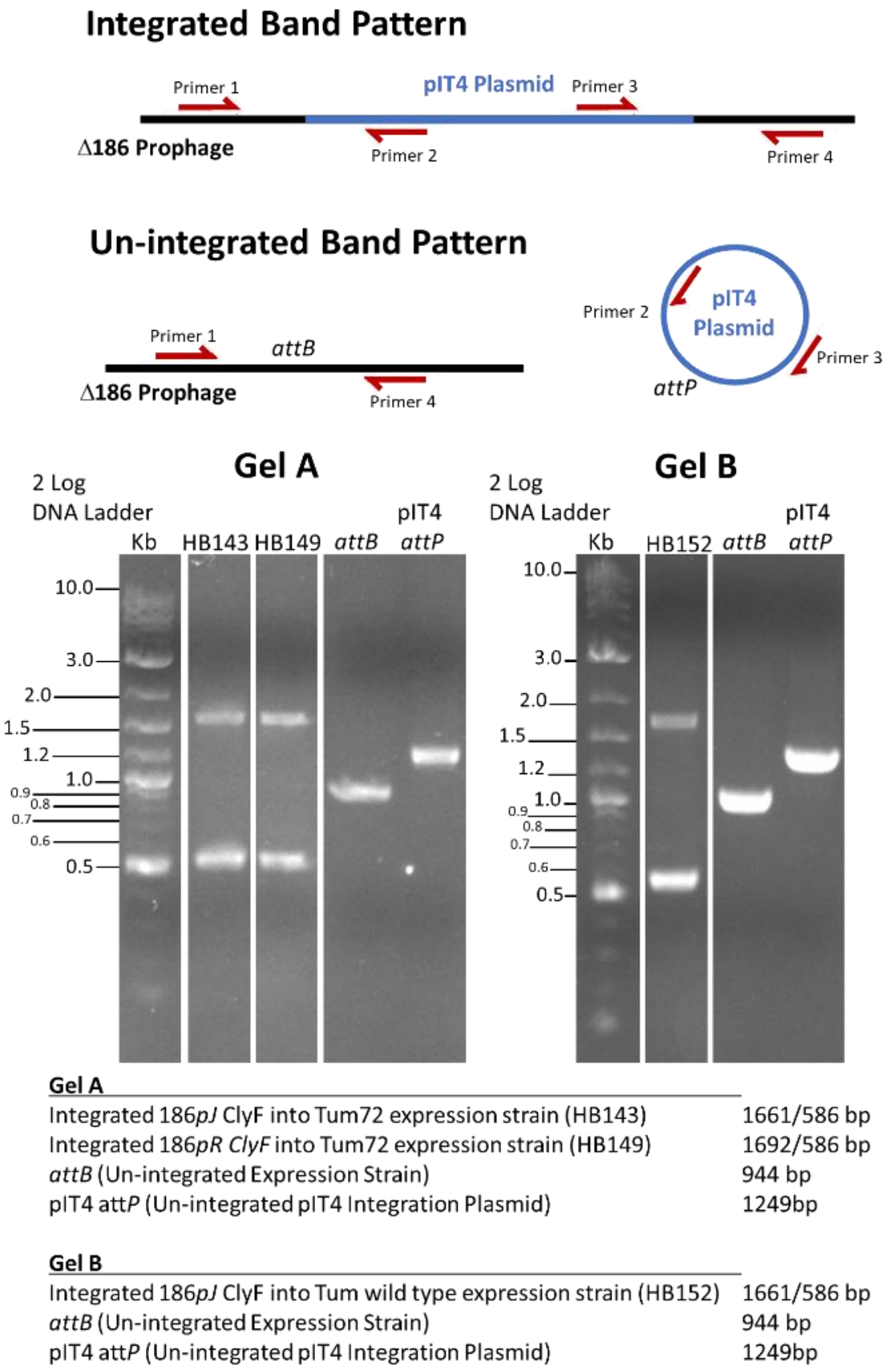
Second G-Block designed to delete p95-Tum-orf97-t97, insert two Voigt att sites (attB2 and attB3) (peach) and loxP sites (pink) flanked chloramphenicol resistance gene cassette (yellow). Codon optimises the chloramphenicol resistance gene to remove commonly used restriction sites. The start and end of the G-Block have an overlap section (underline), which matches the 186 genomes. The second version, G-Block, is located on the left side of the 186 cos sequence, leaving the entire native cos sequence. All strains originating from the HB93 lineage contain this G-Block.

Appendix Fig. 3: Results for the PCR amplification of correct recombineering fragment present in the 186 genome, strain HB61.



After recombineering, individual clones were isolated and tested by PCR to amplify the left and right insertion sites (i.e. using primers 186 genome-Recombineering Fragment, Recombineering Fragment-186 genome) as well as the whole recombineering fragment (186 genome-Recombineering Fragment-186 genome). E4643 [186(Δ*tum* Δ*cos, att2, Cm*)] (HB61) produced the correct band sizes for the left insertion site (0.6 Kb) and for the right insertion site (1.3 Kb). HB61 also produced the correct band size for the whole fragment (1.9 Kb). The whole fragment sample was sequenced (AGRF) to confirm the correct nucleotide sequence. E4643(186) (HB60) was used as the wild type control. This strain should not produce any amplification for the left or right insertion site primers. For the whole fragment amplification, it amplified the native *tum* and *cp97* genes, producing the expected band size of 2.0 Kb. The white line between the HB61 and HB60 PCR samples shows where the gel was cut to remove additional samples from other PCR results.

Appendix Fig. 4: Results for the PCR amplification of correct pIT4 plasmid integration into the 186 genome, for expression strains HB143, HB149 and HB152.



Gel A

Integrated 186*pJ* ClyF into Tum72 expression strain (HB143)	1661/586 bp
Integrated 186*pR ClyF* into Tum72 expression strain (HB149)	1692/586 bp
attB (Un-integrated Expression Strain)	944 bp
pIT4 att*P* (Un-integrated pIT4 Integration Plasmid)	1249bp

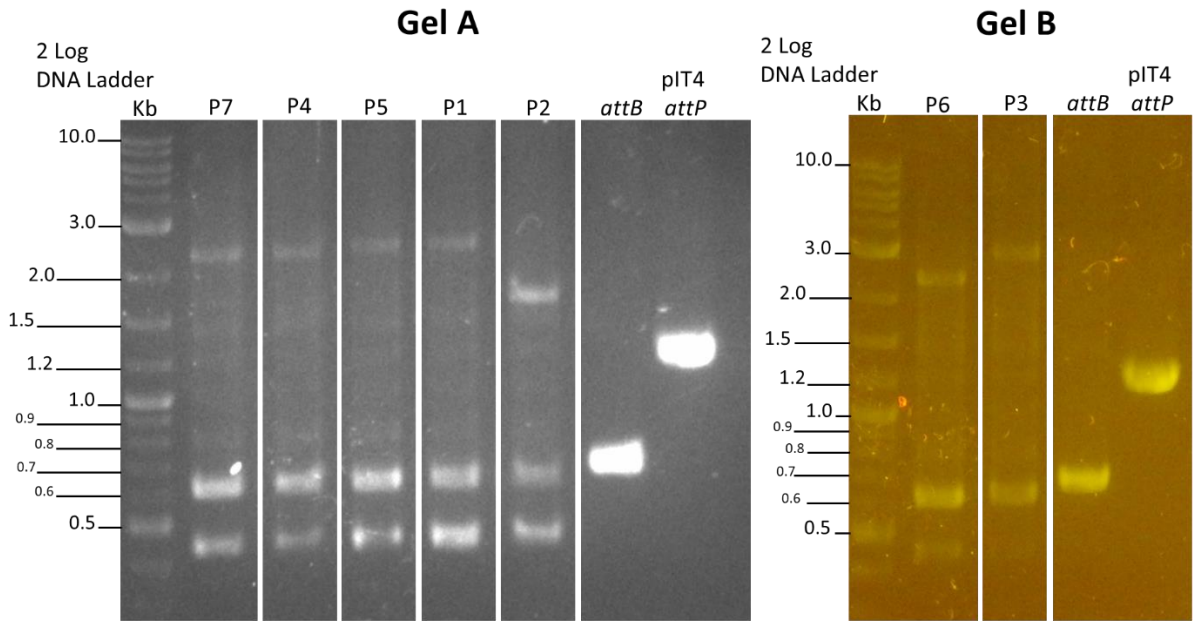
Gel B

Integrated 186*pJ* ClyF into Tum wild type expression strain (HB152)	1661/586 bp
attB (Un-integrated Expression Strain)	944 bp
pIT4 att*P* (Un-integrated pIT4 Integration Plasmid)	1249bp

To confirm the correct integration of the pIT4 integration plasmid into the *att*2 site of the

modified 186 lysogen, multiplex PCR, using four primers per PCR reaction, was used. One pair of primers (Primer 1 and Primer 4) bind to the ∆186 prophage. The second pair of primers (Primer 2 and Primer 3) will only anneal with the pIT4 integration plasmid. The top diagram shows the band pattern produced for integration of the pIT4 plasmid; primers 1 and 2 amplify the *attL* site, and primers 3 and 4 amplify the *attR* site, producing two bands. If there is no pIT4 integration in the ∆186 prophage, primers 1 and 4 will amplify a single band at the *attB* site. If unintegrated pIT4 plasmid is present, primers 2 and 3 will amplify a single band at the *attP* site. From the gel results (Gel 1 and Gel 2), it was concluded that all of the expression strains (HB143, HB149 and HB152) have an integrated pIT4 plasmid. All produced the expected band sizes 1,692 bp (186*pR*) / 1,661 bp (186*pJ*) for the *attL* site and 586 bp for the *attR* site. Each gel also shows the fragments produced from two control PCRs - one control for the *attB* site using an expression strain control and a second control for *attP* using pIT4 plasmid. Each control produced the expected band sizes; *attB* produced a 944 bp band, and pIT4 *attP* control produced a 1,249 bp band. The white line in between PCR samples represent where PCR results were removed from the gel since they were unrelated PCR samples. The HB141 strain was also PCR screened for integration; however, the image could not be shown due to an issue with the gel. HB141 has been PCR screened for integration using four primer amplification and PCR screening for the entire length of the pIT4 integration plasmid present in the expression strain.

Appendix Fig. 5: Results of the PCR amplification of pIT4 plasmid integration into the 186 genome, for each of the putative proteins investigated in Chapter 6.a.



Gel A	
Integrated P7	2359/621 bp
Integrated P4	2317/621 bp
Integrated P5	2407/621 bp
Integrated P1	2401/621 bp
Integrated P2	1702/621 bp
attB (Un-integrated Expression Strain)	680 bp
pIT4 attP (Un-integrated pIT4 Integration Plasmid)	1624bp
Gel B	
Integrated P6	2389/621 bp
Integrated P3	2845/621 bp
attB (Un-integrated Expression Strain)	680 bp
pIT4 attP (Un-integrated pIT4 Integration Plasmid)	1624bp

To confirm the correct integration of the pIT4 integration plasmid into the *att*2 site of the modified 186 lysogen, multiplex PCR, using four primers per PCR reaction, was used. The left and right attachment sites that are created after integration (*attL* and *attR*) are amplified in the same PCR tube using four primers.

The *attL* bands differ in length between proteins since the reaction covers the entire protein sequence. Each protein (P1 to P7) produced the expected band pattern, with the exception of an additional 450 bp band present in all the integration samples at various intensities. This band was likely caused by non-specific amplification due to the cycling settings. Integration sample results were observed over two gels, each also containing two control PCR reactions. Each gel contained two control PCR samples for the un-integrated expression strain control (*attB*) and a pIT4 plasmid control (pIT4 *attP*). These controls produced the expected band sizes; *attB* produced

a 680 bp band, and pIT4 *attP* control produced a 1,624 bp band. Fragments covering the protein sequences were purified and sent for Sanger sequencing to confirm the absence of mutations. The white line in between PCR samples represent where PCR results were removed from the gel since they were additional PCR samples.

References:

1. Centers for Disease Control and Prevention (U.S.). Antibiotic resistance threats in the United States, 2019 [Internet]. Centers for Disease Control and Prevention (U.S.); 2019 Nov [cited 2021 Sep 23]. Available from: https://stacks.cdc.gov/view/cdc/82532

2. Dadgostar P. Antimicrobial Resistance: Implications and Costs. IDR. 2019 Dec;Volume 12:3903–10.

3. World Health Organization, editor. Antimicrobial resistance: global report on surveillance. Geneva, Switzerland: World Health Organization; 2014. 232 p.

4. Lestner JM, Hill LF, Heath PT, Sharland M. Vancomycin toxicity in neonates: a review of the evidence. Current Opinion in Infectious Diseases. 2016 Jun;29(3):237–47.

5. Zamoner W, Prado IRS, Balbi AL, Ponce D. Vancomycin dosing, monitoring and toxicity: Critical review of the clinical practice. Clin Exp Pharmacol Physiol. 2019 Apr;46(4):292–301.

6. Llor C, Bjerrum L. Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. Therapeutic Advances in Drug Safety. 2014 Dec;5(6):229–41.

7. Centers for Disease Control and Prevention (CDC). National Action Plan for Combating Antibiotic-Resistant Bacteria. 2015;

8. Church NA, McKillip JL. Antibiotic resistance crisis: challenges and imperatives. Biologia. 2021 May;76(5):1535–50.

9. Jim O'Neill. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. Review on Antimicrobial Resistance; 2014.

10. Laxminarayan R, Duse A, Wattal C, Zaidi AKM, Wertheim HFL, Sumpradit N, et al. Antibiotic resistance—the need for global solutions. The Lancet Infectious Diseases. 2013 Dec 1;13(12):1057–98.

11. World Health Organization. Regional Office for Europe, Policies EO on HS and, Anderson M, Clift C, Schulze K, Sagan A, et al. Averting the AMR crisis: what are the avenues for policy action for countries in Europe? [Internet]. World Health Organization. Regional Office for Europe; 2019 [cited 2022 Aug 31]. 34 p. Available from: https://apps.who.int/iris/handle/10665/331973

12. Santoro-Lopes G. Multidrug-resistant bacterial infections after liver transplantation: An ever-growing challenge. WJG. 2014;20(20):6201.

13. Finley RL, Collignon P, Larsson DGJ, McEwen SA, Li XZ, Gaze WH, et al. The Scourge of Antibiotic Resistance: The Important Role of the Environment. Clinical Infectious Diseases. 2013 Sep 1;57(5):704–10.

14. Sørensen SJ, Bailey M, Hansen LH, Kroer N, Wuertz S. Studying plasmid horizontal transfer in situ: a critical review. Nat Rev Microbiol. 2005 Sep;3(9):700–10.

15. Ashbolt NJ, Amézquita A, Backhaus T, Borriello P, Brandt KK, Collignon P, et al. Human Health Risk Assessment (HHRA) for Environmental Development and Transfer of Antibiotic Resistance. Environmental Health Perspectives. 2013 Sep;121(9):993–1001.

16. Mann A. Antibiotic resistance in agriculture: Perspectives on upcoming strategies to overcome upsurge in resistance. 2021;14.

17. American Chemical Society [Internet]. 1999 [cited 2022 Aug 25]. Alexander Fleming Discovery and Development of Penicillin - Landmark. Available from: https://www.acs.org/content/acs/en/education/whatischemistry/landmarks/flemingpenicillin.html

18. Davies J, Davies D. Origins and Evolution of Antibiotic Resistance. Microbiol Mol Biol Rev. 2010 Sep;74(3):417–33.

19. Pandey N, Cascella M. Beta Lactam Antibiotics. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 [cited 2022 Sep 1]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK545311/

20. Lima LM, Silva BNM da, Barbosa G, Barreiro EJ. β-lactam antibiotics: An overview from a medicinal chemistry perspective. European Journal of Medicinal Chemistry. 2020 Dec;208:112829.

21. Ventola CL. The Antibiotic Resistance Crisis. Pharmacy and Therapeutics. 2015;40(4):277.

22. Silver LL. Challenges of Antibacterial Discovery. Clin Microbiol Rev. 2011 Jan;24(1):71–109.

23. Stennett HL, Back CR, Race PR. Derivation of a Precise and Consistent Timeline for Antibiotic Development. Antibiotics. 2022 Sep 12;11(9):1237.

24. BioNTech SE. A PHASE 1/2/3, PLACEBO-CONTROLLED, RANDOMIZED, OBSERVER-BLIND, DOSE-FINDING STUDY TO EVALUATE THE SAFETY, TOLERABILITY, IMMUNOGENICITY, AND EFFICACY OF SARS-COV-2 RNA VACCINE CANDIDATES AGAINST COVID-19 IN HEALTHY INDIVIDUALS [Internet]. clinicaltrials.gov; 2022 Nov [cited 2023 Jan 10]. Report No.: NCT04368728. Available from: https://clinicaltrials.gov/ct2/show/NCT04368728

25. Hutchings MI, Truman AW, Wilkinson B. Antibiotics: past, present and future. Current Opinion in Microbiology. 2019 Oct;51:72–80.

26. Chatterjee A, Modarai M, Naylor NR, Boyd SE, Atun R, Barlow J, et al. Quantifying drivers of antibiotic resistance in humans: a systematic review. The Lancet Infectious Diseases. 2018 Dec;18(12):e368–78.

27. Low CX, Tan LTH, Ab Mutalib NS, Pusparajah P, Goh BH, Chan KG, et al. Unveiling the Impact of Antibiotics and Alternative Methods for Animal Husbandry: A Review. Antibiotics. 2021 May 13;10(5):578.

28. U.S. Food & Drug Administration, Center For Veterinary Medicine. 2019 Summary Report On Antimicrobials Sold or Distributed for Use in Food-Producing Animals. FDA. 2020 Dec;49.

29. Centers for Disease Control and Prevention (U.S.);, National Center for Emerging Zoonotic and Infectious Diseases (U.S.);, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (U.S.);, National Center for Immunization and Respiratory Diseases (U.S.); Antibiotic resistance threats in the United States, 2013 [Internet]. 2013 [cited 2022 Sep 9]. Available from: https://stacks.cdc.gov/view/cdc/20705

30. Taylor P, Reeder R. Antibiotic use on crops in low and middle-income countries based on recommendations made by agricultural advisors. CABI Agric Biosci. 2020 Dec;1(1):1.

31. Jacobs A, Adno M. Citrus Farmers Facing Deadly Bacteria Turn to Antibiotics, Alarming Health Officials. The New York Times [Internet]. 2019 May 18 [cited 2022 Sep 14]; Available from: https://www.nytimes.com/2019/05/17/health/antibiotics-oranges-florida.html

32. Karkman A, Do TT, Walsh F, Virta MPJ. Antibiotic-Resistance Genes in Waste Water. Trends in Microbiology. 2018 Mar;26(3):220–8.

33. Chi T, Zhang A, Zhang X, Li AD, Zhang H, Zhao Z. Characteristics of the antibiotic resistance genes in the soil of medical waste disposal sites. Science of The Total Environment. 2020 Aug;730:139042.

34. Edwards SE, Morel CM. Learning from our mistakes: using key opportunities to remove the perverse incentives that help drive antibiotic resistance. Expert Review of Pharmacoeconomics & Outcomes Research. 2019 Nov 2;19(6):685–92.

35. Randall JR, Davies BW. Mining for novel antibiotics. Current Opinion in Microbiology. 2021 Oct;63:66–9.

36. Committee on Ensuring Patient Access to Affordable Drug Therapies, Board on Health Care Services, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. Making Medicines Affordable: A National Imperative [Internet]. Augustine NR, Madhavan G, Nass SJ, editors. Washington, D.C.: National Academies Press; 2018 [cited 2022 Sep 15]. Available from: https://www.nap.edu/catalog/24946

37. Topic: Antibiotics [Internet]. [cited 2022 Sep 15]. Antibiotics. Available from: http://pew.org/MIIugN

38. Simpkin VL, Renwick MJ, Kelly R, Mossialos E. Incentivising innovation in antibiotic drug discovery and development: progress, challenges and next steps. J Antibiot. 2017 Dec;70(12):1087–96.

39. Wooldridge M. Evidence for the circulation of antimicrobial-resistant strains and genes in nature and especially between humans and animals. Revue scientifique et technique (International Office of Epizootics). 2012;31(1):231–47.

40. Collignon P. Antibiotic resistance: are we all doomed? 2015;8.

41. Hacker J, Blum-Oehler G, Muhldorfer I, Tschape H. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol Microbiol. 1997 Mar;23(6):1089–97.

42. Lerminiaux NA, Cameron ADS. Horizontal transfer of antibiotic resistance genes in clinical environments. Can J Microbiol. 2019 Jan;65(1):34–44.

43. Johnston C, Martin B, Fichant G, Polard P, Claverys JP. Bacterial transformation: distribution, shared mechanisms and divergent control. Nat Rev Microbiol. 2014 Mar;12(3):181–96.

44. Sun D. Pull in and Push Out: Mechanisms of Horizontal Gene Transfer in Bacteria. Front Microbiol. 2018 Sep 6;9:2154.

45. Massoudieh A, Mathew A, Lambertini E, Nelson KE, Ginn TR. Horizontal Gene Transfer on Surfaces in Natural Porous Media: Conjugation and Kinetics. Vadose zone j. 2007 May;6(2):306–15.

46. Cabezón E, Ripoll-Rozada J, Peña A, de la Cruz F, Arechaga I. Towards an integrated model of bacterial conjugation. FEMS Microbiol Rev. 2014 Sep;n/a-n/a.

47. Bennett PM. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria: Plasmid-encoded antibiotic resistance. British Journal of Pharmacology. 2008 Mar;153(S1):S347–57.

48. Rakhuba DV, Kolomiets EI, Szwajcer Dey E, Novik GI. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. Polish Journal of Microbiology. 2010;59(3):145–55.

49. Hendrix RW. Bacteriophages: Evolution of the Majority. Theoretical Population Biology. 2002 Jun;61(4):471–80.

50. Cisek AA, Dąbrowska I, Gregorczyk KP, Wyżewski Z. Phage Therapy in Bacterial Infections Treatment: One Hundred Years After the Discovery of Bacteriophages. Curr Microbiol. 2017 Feb;74(2):277–83.

51. Harrison E, Brockhurst MA. Ecological and Evolutionary Benefits of Temperate Phage: What Does or Doesn't Kill You Makes You Stronger. BioEssays. 2017 Dec;39(12):1700112.

52. Chlebowicz MA, Mašlaňová I, Kuntová L, Grundmann H, Pantůček R, Doškař J, et al. The Staphylococcal Cassette Chromosome mec type V from Staphylococcus aureus ST398 is packaged into bacteriophage capsids. International Journal of Medical Microbiology. 2014 Jul;304(5–6):764–74.

53. Czaplewski L, Bax R, Clokie M, Dawson M, Fairhead H, Fischetti VA, et al. Alternatives to antibiotics-a pipeline portfolio review. The Lancet Infectious Diseases. 2016;16(2):239–51.

54. Sulakvelidze A, Alavidze Z, Morris JG. Bacteriophage Therapy. ANTIMICROB AGENTS CHEMOTHER. 2001;45:11.

55. Górski A, Międzybrodzki R, Węgrzyn G, Jończyk-Matysiak E, Borysowski J, Weber-Dąbrowska B. Phage therapy: Current status and perspectives. Med Res Rev. 2020 Jan;40(1):459–63.

56. Międzybrodzki R, Hoyle N, Zhvaniya F, Łusiak-Szelachowska M, Weber-Dąbrowska B, Łobocka M, et al. Current Updates from the Long-Standing Phage Research Centers in Georgia, Poland, and Russia. In: Harper DR, Abedon ST, Burrowes BH, McConville ML, editors. Bacteriophages [Internet]. Cham: Springer International Publishing; 2021 [cited 2023 Jan 31]. p. 921–51. Available from: https://link.springer.com/10.1007/978-3-319-41986-2_31

57. Petrovski S, Dyson ZA, Seviour RJ, Tillett D. Small but Sufficient: the Rhodococcus Phage RRH1 Has the Smallest Known Siphoviridae Genome at 14.2 Kilobases. J Virol. 2012 Jan;86(1):358–63.

58. González B, Monroe L, Li K, Yan R, Wright E, Walter T, et al. Phage G Structure at 6.1 Å Resolution, Condensed DNA, and Host Identity Revision to a Lysinibacillus. Journal of Molecular Biology. 2020 Jun;432(14):4139–53.

59. Malik DJ. Formulation, stabilisation and encapsulation of bacteriophage for phage therapy. Advances in Colloid and Interface Science. 2017;34.

60. Wittebole X, De Roock S, Opal SM. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. Virulence. 2014 Jan;5(1):226–35.

61. King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, editors. Order - Caudovirales. In: Virus Taxonomy [Internet]. San Diego: Elsevier; 2012 [cited 2022 Sep 7]. p. 39–45. Available from: https://www.sciencedirect.com/science/article/pii/B978012384684600001X

62. Turner D, Shkoporov AN, Lood C, Millard AD, Dutilh BE, Alfenas-Zerbini P, et al. Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. Arch Virol. 2023 Feb;168(2):74.

63. McGowan S, Buckle AM, Mitchell MS, Hoopes JT, Gallagher DT, Heselpoth RD, et al. X-ray crystal structure of the streptococcal specific phage lysin PlyC. Proceedings of the National Academy of Sciences. 2012;109(31):12752–7.

64. Barbirz S, Müller JJ, Uetrecht C, Clark AJ, Heinemann U, Seckler R. Crystal structure of *Escherichia coli* phage HK620 tailspike: podoviral tailspike endoglycosidase modules are evolutionarily related. Molecular Microbiology. 2008 Jul;69(2):303–16.

65. Ackermann HW. Bacteriophage Electron Microscopy. In: Advances in Virus Research [Internet]. Elsevier; 2012 [cited 2023 Feb 1]. p. 1–32. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780123946218000170

66. Lee DY. Predicting the capsid architecture of phages from metagenomic data. Computational and Structural Biotechnology Journal. 2022;12.

67. Maniloff J, Ackermann HW, Jarvis A. PHAGE TAXONOMY AND CLASSIFICATION. In: Granoff A, Webster RG, editors. Encyclopedia of Virology (Second Edition) [Internet]. Oxford: Elsevier; 1999 [cited 2022 Sep 7]. p. 1221–8. Available from: https://www.sciencedirect.com/science/article/pii/B0122270304000248

68. Aksyuk AA, Rossmann MG. Bacteriophage Assembly. Viruses. 2011 Feb 25;3(3):172–203.

69. Scholl D. Phage Tail–Like Bacteriocins. 2017;18.

70. Zinke M, Schröder GF, Lange A. Major tail proteins of bacteriophages of the order Caudovirales. Journal of Biological Chemistry. 2022;298(1).

71. Yang C, Wang H, Ma H, Bao R, Liu H, Yang L, et al. Characterization and Genomic Analysis of SFPH2, a Novel T7virus Infecting Shigella. Front Microbiol. 2018 Dec 14;9:3027.

72. Padilla-Sanchez V. Bacteriophage Lambda Structural Model at Atomic Resolution [Internet]. 2021 [cited 2023 Apr 26]. Available from: https://zenodo.org/record/5134493

73. John Dennehy. What has phage lambda ever done for us? [Internet]. [cited 2023 Apr 26]. Available from: http://evilutionarybiologist.blogspot.com/2007/04/what-has-phage-lambda-ever-done-for-us.html

74. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. Nat Rev Microbiol. 2020 Mar;18(3):125–38.

75. Ongenae V, Briegel A, Claessen D. Cell wall deficiency as an escape mechanism from phage infection. Open Biol. 2021 Sep;11(9):210199.

76. Hay ID, Lithgow T. Filamentous phages: masters of a microbial sharing economy. EMBO Rep [Internet]. 2019 Jun [cited 2023 Feb 1];20(6). Available from: https://onlinelibrary.wiley.com/doi/10.15252/embr.201847427

77. Groth AC, Calos MP. Phage Integrases: Biology and Applications. Journal of Molecular Biology. 2004 Jan;335(3):667–78.

78. Hao N, Agnew D, Krishna S, Dodd IB, Shearwin KE. Analysis of Infection Time Courses Shows CII Levels Determine the Frequency of Lysogeny in Phage 186. Pharmaceuticals. 2021 Sep 29;14(10):998.

79. Shotland Y, Shifrin A, Ziv T, Teff D, Koby S, Kobiler O, et al. Proteolysis of Bacteriophage λ CII by *Escherichia coli* FtsH (HflB). J Bacteriol. 2000 Jun;182(11):3111–6.

80. Rodríguez-Rubio L, Martínez B, Donovan DM, Rodríguez A, García P. Bacteriophage virion-associated peptidoglycan hydrolases: potential new enzybiotics. Critical Reviews in Microbiology. 2013 Nov;39(4):427–34.

81. Drulis-Kawa Z, Majkowska-Skrobek G, Maciejewska B. Bacteriophages and phage-derived proteins–application approaches. Current medicinal chemistry. 2015;22(14):1757–73.

82. Oliveira H, São-José C, Azeredo J. Phage-Derived Peptidoglycan Degrading Enzymes: Challenges and Future Prospects for In Vivo Therapy. Viruses. 2018 May 29;10(6):292.

83. Roach DR, Donovan DM. Antimicrobial bacteriophage-derived proteins and therapeutic applications. Bacteriophage. 2015 Jul 3;5(3):e1062590.

84. Suresh Kumar A, Mody K, Jha B. Bacterial exopolysaccharides – a perception. Journal of Basic Microbiology. 2007;47(2):103–17.

85. Pires DP, Oliveira H, Melo LDR, Sillankorva S, Azeredo J. Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. Appl Microbiol Biotechnol. 2016 Mar;100(5):2141–51.

86. Topka-Bielecka G, Dydecka A, Necel A, Bloch S, Nejman-Faleńczyk B, Węgrzyn G, et al. Bacteriophage-Derived Depolymerases against Bacterial Biofilm. Antibiotics. 2021 Feb 10;10(2):175.

87. Catalão MJ, Gil F, Moniz-Pereira J, São-José C, Pimentel M. Diversity in bacterial lysis systems: bacteriophages show the way. FEMS Microbiol Rev. 2013 Jul;37(4):554–71.

88. Yang H, Yu J, Wei H. Engineered bacteriophage lysins as novel anti-infectives. Frontiers in Microbiology [Internet]. 2014 Oct 16 [cited 2018 Sep 17];5. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2014.00542/abstract

89. Nelson D, Schuch R, Chahales P, Zhu S, Fischetti VA. PlyC: A multimeric bacteriophage lysin. Proceedings of the National Academy of Sciences. 2006;103(28):10765–70.

90. Chang Y, Yoon H, Kang DH, Chang PS, Ryu S. Endolysin LysSA97 is synergistic with carvacrol in controlling Staphylococcus aureus in foods. International Journal of Food Microbiology. 2017;244:19–26.

91. Guo M, Feng C, Ren J, Zhuang X, Zhang Y, Zhu Y, et al. A novel antimicrobial endolysin, LysPA26, against Pseudomonas aeruginosa. Frontiers in Microbiology. 2017;8(FEB):1–9.

92. Maliničová L, Piknová M, Pristaš P, Javorský P. Peptidoglycan hydrolases as novel tool for anti-enterococcal therapy. Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology The Formatex Microbiology Book Series. 2010;1:11.

93. Jun SY, Jung GM, Yoon SJ, Youm SY, Han HY, Lee JH, et al. Pharmacokinetics of the phage endolysin-based candidate drug SAL200 in monkeys and its appropriate intravenous dosing period. Clin Exp Pharmacol Physiol. 2016 Oct;43(10):1013–6.

94. Zhang L, Li D, Li X, Hu L, Cheng M, Xia F, et al. LysGH15 kills Staphylococcus aureus without being affected by the humoral immune response or inducing inflammation. Scientific Reports. 2016;6(July):1–9.

95. Totté JEE, van Doorn MB, Pasmans SGMA. Successful Treatment of Chronic Staphylococcus aureus-Related Dermatoses with the Topical Endolysin Staphefekt SA.100: A Report of 3 Cases. Case Reports in Dermatology. 2017;19–25.

96. Vollmer W, Blanot D, De Pedro MA. Peptidoglycan structure and architecture. FEMS Microbiology Reviews. 2008 Mar;32(2):149–67.

97. Govindarajan S, Amster-Choder O. The bacterial Sec system is required for the organization and function of the MreB cytoskeleton. Casadesús J, editor. PLoS Genet. 2017 Sep 25;13(9):e1007017.

98. Wang IN, Smith DL, Young R. Holins: The Protein Clocks of Bacteriophage Infections. Annu Rev Microbiol. 2000 Oct;54(1):799–825.

99. Broendum SS, Buckle AM, McGowan S. Catalytic diversity and cell wall binding repeats in the phage encoded endolysins. Molecular Microbiology [Internet]. 2018 Sep 19 [cited 2018 Sep 24]; Available from: http://doi.wiley.com/10.1111/mmi.14134

100. São-José C. Engineering of Phage-Derived Lytic Enzymes: Improving Their Potential as Antimicrobials. Antibiotics. 2018;7(2):29.

101. Gerstmans H, Rodriguez-Rubio L, Lavigne R, Briers Y. From endolysins to Artilysin(R)s: novel enzyme-based approaches to kill drug-resistant bacteria. Biochemical Society Transactions. 2016 Feb 15;44(1):123–8.

102. Bissaro B, Monsan P, Fauré R, O'Donohue MJ. Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. Biochemical Journal. 2015 Mar 20;467(1):17–35.

103. Dik DA, Marous DR, Fisher JF, Mobashery S. Lytic transglycosylases: concinnity in concision of the bacterial cell wall. Critical Reviews in Biochemistry and Molecular Biology. 2017 Sep 3;52(5):503–42.

104. Low LY, Yang C, Perego M, Osterman A, Liddington R. Role of Net Charge on Catalytic Domain and Influence of Cell Wall Binding Domain on Bactericidal Activity, Specificity, and Host Range of Phage Lysins. Journal of Biological Chemistry. 2011 Sep;286(39):34391–403.

105. Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. Structure. 1995 Sep;3(9):853–9.

106. Rigden DJ, Jedrzejas MJ, Galperin MY. Amidase domains from bacterial and phage autolysins define a family of g-D,L-glutamate-specific amidohydrolasesq. 2003;5.

107. Foster SJ. 257 - N-Acetylmuramoyl-l-alanine amidase. In: Barrett AJ, Rawlings ND, Woessner JF, editors. Handbook of Proteolytic Enzymes (Second Edition) [Internet]. London: Academic Press; 2004. p. 866–8. Available from: https://www.sciencedirect.com/science/article/pii/B9780120796113502652

108. Love MJ, Abeysekera GS, Muscroft-Taylor AC, Billington C, Dobson RCJ. On the catalytic mechanism of bacteriophage endolysins: Opportunities for engineering. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2020 Jan;1868(1):140302.

109. Layec S, Decaris B, Leblond-Bourget N. Characterization of Proteins Belonging to the CHAP-Related Superfamily within the Firmicutes. J Mol Microbiol Biotechnol. 2008;14(1–3):31–40.

110. Linden SB, Zhang H, Heselpoth RD, Shen Y, Schmelcher M, Eichenseher F, et al. Biochemical and biophysical characterization of PlyGRCS, a bacteriophage endolysin active against methicillin-resistant Staphylococcus aureus. Appl Microbiol Biotechnol. 2015 Jan;99(2):741–52.

111. Pohane AA, Patidar ND, Jain V. Modulation of domain-domain interaction and protein function by a charged linker: A case study of mycobacteriophage D29 endolysin. FEBS Letters. 2015 Mar 12;589(6):695–701.

112. Becker SC, Swift S, Korobova O, Schischkova N, Kopylov P, Donovan DM, et al. Lytic activity of the staphylolytic Twort phage endolysin CHAP domain is enhanced by the SH3b cell wall binding domain. FEMS Microbiol Lett. 2015 Jan;362(1):1–8.

113. Benešík M, Nováček J, Janda L, Dopitová R, Pernisová M, Melková K, et al. Role of SH3b binding domain in a natural deletion mutant of Kayvirus endolysin LysF1 with a broad range of lytic activity. Virus Genes. 2018 Feb;54(1):130–9.

114. Bustamante N, Iglesias-Bexiga M, Bernardo-García N, Silva-Martín N, García G, Campanero-Rhodes MA, et al. Deciphering how Cpl-7 cell wall-binding repeats recognize the bacterial peptidoglycan. Scientific reports. 2017;7(1):1–17.

115. Maciejewska B, Źrubek K, Espaillat A, Wiśniewska M, Rembacz KP, Cava F, et al. Modular endolysin of Burkholderia AP3 phage has the largest lysozyme-like catalytic subunit discovered to date and no catalytic aspartate residue. Sci Rep. 2017 Dec;7(1):14501.

116. Dunne M, Leicht S, Krichel B, Mertens HDT, Thompson A, Krijgsveld J, et al. Crystal structure of the CTP1L endolysin reveals how its activity is regulated by a secondary translation product. Journal of Biological Chemistry. 2016;291(10):4882–93.

117. Fernández-Tornero C, López R, García E, Giménez-Gallego G, Romero A. A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA. Nat Struct Biol. 2001 Dec 1;8(12):1020–4.

118. Perez-Dorado I, Campillo NE, Monterroso B, Hesek D, Lee M, Paez JA, et al. Elucidation of the molecular recognition of bacterial cell wall by modular pneumococcal phage endolysin CPL-1. Journal of Biological Chemistry. 2007;282(34):24990–9.

119. Gründling A, Schneewind O. Cross-Linked Peptidoglycan Mediates Lysostaphin Binding to the Cell Wall Envelope of *Staphylococcus aureus*. J Bacteriol. 2006 Apr;188(7):2463–72.

120. Kurochkina N, editor. SH Domains [Internet]. Cham: Springer International Publishing; 2015 [cited 2021 Jul 5]. Available from: http://link.springer.com/10.1007/978-3-319-20098-9

121. Dong H, Zhu C, Chen J, Ye X, Huang YP. Antibacterial activity of Stenotrophomonas maltophilia endolysin P28 against both gram-positive and gram-negative bacteria. Frontiers in Microbiology. 2015;6(NOV):1–8.

122. Chen CY, Richardson JP. Sequence elements essential for rho-dependent transcription termination at lambda tR1. Journal of Biological Chemistry. 1987 Aug;262(23):11292–9.

123. Gu J, Feng Y, Feng X, Sun C, Lei L, Ding W, et al. Structural and Biochemical Characterization Reveals LysGH15 as an Unprecedented “EF-Hand-Like” Calcium-Binding Phage Lysin. PLoS Pathogens. 2014;10(5).

124. Oliveira H, Melo LDR, Santos SB, Nobrega FL, Ferreira EC, Cerca N, et al. Molecular Aspects and Comparative Genomics of Bacteriophage Endolysins. Journal of Virology. 2013;87(8):4558–70.

125. Fernandes S, Proença D, Cantante C, Silva FA, Leandro C, Lourenço S, et al. Novel Chimerical Endolysins with Broad Antimicrobial Activity Against Methicillin-Resistant *Staphylococcus aureus*. Microbial Drug Resistance. 2012 Jun;18(3):333–43.

126. Gerstmans H, Criel B, Briers Y. Synthetic biology of modular endolysins. Biotechnology Advances. 2018 May;36(3):624–40.

127. Liu A, Wang Y, Cai X, Jiang S, Cai X, Shen L, et al. Characterization of endolysins from bacteriophage LPST10 and evaluation of their potential for controlling Salmonella Typhimurium on lettuce. LWT. 2019 Nov;114:108372.

128. García P, Martínez B, Rodríguez L, Rodríguez A. Synergy between the phage endolysin LysH5 and nisin to kill Staphylococcus aureus in pasteurized milk. International Journal of Food Microbiology. 2010 Jul;141(3):151–5.

129. O'Flaherty S, Coffey A, Meaney W, Fitzgerald GF, Ross RP. The Recombinant Phage Lysin LysK Has a Broad Spectrum of Lytic Activity against Clinically Relevant Staphylococci, Including Methicillin-Resistant *Staphylococcus aureus*. J Bacteriol. 2005 Oct 15;187(20):7161–4.

130. Yang H, Zhang H, Wang J, Yu J, Wei H. A novel chimeric lysin with robust antibacterial activity against planktonic and biofilm methicillinresistant Staphylococcus aureus. Scientific Reports. 2017;7(December 2016):1–13.

131. Li X, Wang S, Nyaruaba R, Liu H, Yang H, Wei H. A Highly Active Chimeric Lysin with a Calcium-Enhanced Bactericidal Activity against Staphylococcus aureus In Vitro and In Vivo. Antibiotics. 2021 Apr 19;10(4):461.

132. Dong Q, Wang J, Yang H, Wei C, Yu J, Zhang Y, et al. Construction of a chimeric lysin Ply187N-V12C with extended lytic activity against staphylococci and streptococci. Microbial Biotechnology. 2015;8(2):210–20.

133. Gilmer DB, Schmitz JE, Thandar M, Euler CW, Fischetti VA. The phage lysin PlySs2 Decolonizes Streptococcus suis from murine intranasal mucosa. PLoS ONE. 2017;12(1):1–13.

134. Gonzalez-Delgado LS, Walters-Morgan H, Salamaga B, Robertson AJ, Hounslow AM, Jagielska E, et al. Two-site recognition of Staphylococcus aureus peptidoglycan by lysostaphin SH3b. Nat Chem Biol [Internet]. 2019 Nov 4 [cited 2019 Nov 26]; Available from: http://www.nature.com/articles/s41589-019-0393-4

135. Farhadian S, Asoodeh A, Lagzian M. Purification, biochemical characterization and structural modeling of a potential htrA-like serine protease from Bacillus subtilis DR8806. Journal of Molecular Catalysis B: Enzymatic. 2015 May;115:51–8.

136. Xu X, Zhang D, Zhou B, Zhen X, Ouyang S. Structural and biochemical analyses of the tetrameric cell binding domain of Lys170 from enterococcal phage F170/08. Eur Biophys J [Internet]. 2021 Feb 20 [cited 2021 Mar 2]; Available from: http://link.springer.com/10.1007/s00249-021-01511-x

137. Gong P, Cheng M, Li X, Jiang H, Yu C, Kahaer N, et al. Characterization of Enterococcus faecium bacteriophage IME-EFm5 and its endolysin LysEFm5. Virology. 2016;492:11–20.

138. Gilmer DB, Schmitz JE, Euler CW, Fischetti VA. Novel Bacteriophage Lysin with Broad Lytic Activity Protects against Mixed Infection by Streptococcus pyogenes and Methicillin-Resistant Staphylococcus aureus. Antimicrob Agents Chemother. 2013 Jun;57(6):2743–50.

139. Swift S, Seal B, Garrish J, Oakley B, Hiett K, Yeh HY, et al. A Thermophilic Phage Endolysin Fusion to a Clostridium perfringens-Specific Cell Wall Binding Domain Creates an Anti-Clostridium Antimicrobial with Improved Thermostability. Viruses. 2015 Jun 12;7(6):3019–34.

140. Xiang Y, Leiman PG, Li L, Grimes S, Anderson DL, Rossmann MG. Crystallographic Insights into the Autocatalytic Assembly Mechanism of a Bacteriophage Tail Spike. Molecular Cell. 2009 May;34(3):375–86.

141. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. Journal of Pharmacological and Toxicological Methods. 2000 Jul;44(1):235–49.

142. Trevino SR, Scholtz JM, Pace CN. Measuring and Increasing Protein Solubility. Journal of Pharmaceutical Sciences. 2008 Oct;97(10):4155–66.

143. Loessner MJ, Wendlinger G, Scherer S. Heterogeneous endolysins in Listeria monocytogenes bacteriophages: a new class of enzymes and evidence for conserved holin genes within the siphoviral lysis cassettes. Molecular Microbiology. 1995;16(6):1231–41.

144. Gerstmans H, Grimon D, Gutiérrez D, Lood C, Rodríguez A, van Noort V, et al. A VersaTile-driven platform for rapid hit-to-lead development of engineered lysins. Sci Adv. 2020 Jun;6(23):eaaz1136.

145. Isabel A. Bacteriophage 186-Investigating the role of transcriptional regulators CI, Apl, CII and Tum at the lytic/lysogenic switch during 186 prophage induction. 2020.

146. PubChem. 4-Isopropylbenzoic acid [Internet]. [cited 2022 Oct 31]. Available from: https://pubchem.ncbi.nlm.nih.gov/compound/10820

147. Yang L, Nielsen AAK, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK, et al. Permanent genetic memory with >1-byte capacity. Nature Methods. 2014;11(12):1261–6.

148. Egan M, Ramirez J, Xander C, Upreti C, Bhatt S. Lambda Red-mediated Recombineering in the Attaching and Effacing Pathogen Escherichia albertii. Biol Proced Online. 2016 Dec;18(1):3.

149. Hao N, Sullivan AE, Shearwin KE, Dodd IB. The loopometer: a quantitative *in vivo* assay for DNA-looping proteins. Nucleic Acids Research. 2021 Apr 19;49(7):e39–e39.

150. Pinkett HW, Shearwin KE, Stayrook S, Dodd IB, Burr T, Hochschild A, et al. The Structural Basis of Cooperative Regulation at an Alternate Genetic Switch. Molecular Cell. 2006 Mar;21(5):605–15.

151. Crameri A, Whitehorn EA, Tate E, Stemmer WPC. l111proved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling. 1996;14.

152. Segall-Shapiro TH, Sontag ED, Voigt CA. Engineered promoters enable constant gene expression at any copy number in bacteria. Nature Biotechnology. 2018;1–41.

153. Hao N, Whitelaw ML, Shearwin KE, Dodd IB, Chapman-Smith A. Identification of residues in the N-terminal PAS domains important for dimerization of Arnt and AhR. Nucleic Acids Research. 2011 May;39(9):3695–709.

154. Murchland I, Ahlgren-Berg A, Priest DG, Dodd IB, Shearwin KE. Promoter activation by CII, a potent transcriptional activator from bacteriophage 186. Journal of Biological Chemistry. 2014;289(46):32094–108.

155. PickMutant™ Error Prone PCR Kit [Internet]. Canvax Reagents SL. [cited 2024 Feb 7]. Available from: https://www.canvaxbiotech.com/mutagenesis/random-mutagenesis/pickmutant-error-prone-pcr-kit/

156. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009 May;6(5):343–5.

157. Chung CT, Niemela SL, Miller RH. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. Proc Natl Acad Sci USA. 1989 Apr;86(7):2172–5.

158. Hao N, Chen Q, Dodd IB, Shearwin KE. The pIT5 Plasmid Series, an Improved Toolkit for Repeated Genome Integration in *E. coli*. ACS Synth Biol. 2021 Jul 16;10(7):1633–9.

159. St-Pierre F, Cui L, Priest DG, Endy D, Dodd IB, Shearwin KE. One-Step Cloning and Chromosomal Integration of DNA. ACS Synthetic Biology. 2013 Sep 20;2(9):537–41.

160. Sharan SK, Thomason LC, Kuznetsov SG, Court DL. Recombineering: a homologous recombination-based method of genetic engineering. Nat Protoc. 2009 Feb;4(2):206–23.

161. Hao N, Shearwin KE, Dodd IB. Programmable DNA looping using engineered bivalent dCas9 complexes. Nature Communications [Internet]. 2017 Dec [cited 2019 Mar 22];8(1). Available from: http://www.nature.com/articles/s41467-017-01873-x

162. Janion C. Inducible SOS Response System of DNA Repair and Mutagenesis in Escherichia coli. Int J Biol Sci. 2008;338–44.

163. Shearwin KE, Brumby AM, Egan JB. The Tum Protein of Coliphage 186 Is an Antirepressor. Journal of Biological Chemistry. 1998 Mar 6;273(10):5708–15.

164. Mullick A, Xu Y, Warren R, Koutroumanis M, Guilbault C, Broussau S, et al. The cumate gene-switch: a system for regulated expression in mammalian cells. BMC Biotechnol. 2006;6(1):43.

165. Brumby AM, Lamont I, Dodd IB, Egan JB. Defining the SOS Operon of Coliphage 186. Virology. 1996 May;219(1):105–14.

166. Dibbens JA, Gregory SL, Egan JB. Control of gene expression in the temperate coliphage 186. X. The *cI* repressor directly represses transcription of the late control gene *B*. Molecular Microbiology. 1992 Sep;6(18):2643–50.

167. Schubert R. Regulation of expression and activity of the late gene activator, B, of bacteriophage 186 / Rachel Ann Schubert. [Internet]. University of Adelaide; 2005. Available from: https://digital.library.adelaide.edu.au/dspace/handle/2440/22207

168. Cui L, Shearwin KE. Clonetegration Using OSIP Plasmids: One-Step DNA Assembly and Site-Specifi c Genomic Integration in Bacteria. In: Synthetic DNA [Internet]. New York: Humana Press; 2017. p. 139–55. Available from: http://link.springer.com/10.1007/978-1-4939-6343-0

169. Carter Z, Delneri D. New generation of loxP-mutated deletion cassettes for the genetic manipulation of yeast natural isolates. Yeast. 2010 Sep;27(9):765–75.

170. Hooper I, Woods WH, Egan B. Coliphage 186 Replication is delayed when the host cell is UV irradiated before infection. J Virol. 1981 Nov;40(2):341–9.

171. Dellis S, Feng J, Filutowicz M. Replication of Plasmid R6K g Origin in Vivo and in Vitro: Dependence on IHF Binding to the ihf1 Site. Journal of Molecular Biology. 1996 Apr 5;11.

172. Hao N, Shearwin KE, Dodd IB. Positive and Negative Control of Enhancer-Promoter Interactions by Other DNA Loops Generates Specificity and Tunability. Cell Reports. 2019 Feb;26(9):2419-2433.e3.

173. Dodd IB, Egan JB. Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. Molecular Microbiology. 2002 Aug;45(3):697–710.

174. Waldor MK, Friedman DI. Phage regulatory circuits and virulence gene expression. Current Opinion in Microbiology. 2005 Aug;8(4):459–65.

175. Morrison KL, Weiss GA. Combinatorial alanine-scanning. Current Opinion in Chemical Biology. 2001 Jun;5(3):302–7.

176. Patnaik BK. Textbook of Biotechnology. 1st ed. New Delhi: Tata McGraw Hill Education; 2012.

177. Fujisawa H, Morita M. Phage DNA packaging. Genes to Cells. 2003 Nov 14;2(9):537–45.

178. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic acids research. 2019 Jul;47(W1):W636–41.

179. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic acids research. 2016;44(W1):W16–21.

180. Arndt D, Marcu A, Liang Y, Wishart DS. PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. Briefings in Bioinformatics [Internet]. 2017 Sep 25 [cited 2018 Aug 13]; Available from: http://academic.oup.com/bib/article/doi/10.1093/bib/bbx121/4222653/PHAST-PHASTER-and-PHASTEST-Tools-for-finding

181. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990;215(3):403–10.

182. Srividhya KV, Rao GV, Raghavenderan L, Mehta P, Prilusky J, Manicka S, et al. Database and comparative identification of prophages. In: Intelligent Control and Automation. Springer; 2006. p. 863–8.

183. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics. 2007;23(6):673–9.

184. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic acids research. 1997;25(5):955–64.

185. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic acids research. 2004;32(1):11–6.

186. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd. 1996. p. 226–31.

187. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A Fast Phage Search Tool. Nucleic Acids Research. 2011;39(SUPPL. 2):347–52.

188. Eddy S. HMMER user's guide. Department of Genetics, Washington University School of Medicine. 1992;2(1):13.

189. Bardy JJ, Sarovich DS, Price EP, Steinig E, Tong S, Drilling A, et al. Staphylococcus aureus from patients with chronic rhinosinusitis show minimal genetic association between polyp and non-polyp phenotypes. BMC Ear, Nose and Throat Disorders [Internet]. 2018 Dec [cited 2019 Feb 20];18(1). Available from: https://bmcearnosethroatdisord.biomedcentral.com/articles/10.1186/s12901-018-0064-1

190. Bernardy EE, Petit RA, Moller AG, Blumenthal JA, McAdam AJ, Priebe GP, et al. Whole-Genome Sequences of Staphylococcus aureus\ Isolates from Cystic Fibrosis Lung Infections. Putonti C, editor. Microbiol Resour Announc. 2019 Jan 17;8(3):e01564-18, /mra/8/3/MRA.01564-18.atom.

191. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer EL, et al. Pfam: The protein families database in 2021. Nucleic Acids Research. 2021;49(D1):D412–9.

192. Rosenfeld RM, Piccirillo JF, Chandrasekhar SS, Brook I, Ashok Kumar K, Kramper M, et al. Clinical Practice Guideline (Update): Adult Sinusitis Executive Summary. Otolaryngol Head Neck Surg. 2015 Apr;152(4):598–609.

193. Rudmik L. Economics of Chronic Rhinosinusitis. Curr Allergy Asthma Rep. 2017 Apr;17(4):20.

194. Strausbaugh SD, Davis PB. Cystic Fibrosis: A Review of Epidemiology and Pathobiology. Clinics in Chest Medicine. 2007 Jun;28(2):279–88.

195. Besier S, Smaczny C, von Mallinckrodt C, Krahl A, Ackermann H, Brade V, et al. Prevalence and Clinical Significance of Staphylococcus aureus Small-Colony Variants in Cystic Fibrosis Lung Disease. Journal of Clinical Microbiology. 2007 Jan 1;45(1):168–72.

196. Bateman A, Rawlings ND. The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolasesq. 2003;4.

197. Zou Y, Hou C. Systematic analysis of an amidase domain CHAP in 12 Staphylococcus aureus genomes and 44 staphylococcal phage genomes. Computational Biology and Chemistry. 2010;34(4):251–7.

198. Xin W, Wilkinson BJ, Jayaswal RK. Sequence analysis of a Staphylococcus aureus gene encoding a peptidoglycan hydrolase activity. Gene. 1991 Jun;102(1):105–9.

199. Love M, Bhandari D, Dobson R, Billington C. Potential for Bacteriophage Endolysins to Supplement or Replace Antibiotics in Food Production and Clinical Care. Antibiotics. 2018 Feb 27;7(1):17.

200. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Research. 2018 Jul 2;46(W1):W200–4.

201. Hatfield MJ, Umans RA, Hyatt JL, Edwards CC, Wierdl M, Tsurkan L, et al. Carboxylesterases: General detoxifying enzymes. Chemico-Biological Interactions. 2016 Nov;259:327–31.

202. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021 Jan;49(D1):D344–54.

203. Li HF, Wang XF, Tang H. Predicting Bacteriophage Enzymes and Hydrolases by Using Combined Features. Front Bioeng Biotechnol. 2020 Mar 24;8:183.

204. Rawlings ND, Bateman A. Origins of peptidases. Biochimie. 2019 Nov;166:4–18.

205. Ramadurai L, Lockwood KJ, Lockwood J, Nadakavukaren MJ, Jayaswal RK. Characterization of a chromosomally encoded glycylglycine endopeptidase of Staphylococcus aureus. Microbiology. 1999 Apr 1;145(4):801–8.

206. Rawlings ND, Barrett AJ. [2] Families of serine peptidases. In: Methods in Enzymology [Internet]. Elsevier; 1994 [cited 2022 Nov 7]. p. 19–61. Available from: https://linkinghub.elsevier.com/retrieve/pii/0076687994440042

207. Rawlings ND, Barrett AJ. [13] Evolutionary families of metallopeptidases. In: Proteolytic Enzymes: Aspartic and Metallo Peptidases [Internet]. Academic Press; 1995. p. 183–228. (Methods in Enzymology; vol. 248). Available from: https://www.sciencedirect.com/science/article/pii/0076687995480153

208. Odintsov SG, Sabala I, Marcyjaniak M, Bochtler M. Latent LytM at 1.3Å Resolution. Journal of Molecular Biology. 2004 Jan;335(3):775–85.

209. Xiang Y, Morais MC, Cohen DN, Bowman VD, Anderson DL, Rossmann MG. Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage ϕ29 tail. Proc Natl Acad Sci USA. 2008 Jul 15;105(28):9552–7.

210. Stockdale SR, Mahony J, Courtin P, Chapot-Chartier MP, Van Pijkeren JP, Britton RA, et al. The Lactococcal Phages Tuc2009 and TP901-1 Incorporate Two Alternate Forms of Their Tail Fiber into Their Virions for Infection Specialization*. Journal of Biological Chemistry. 2013 Feb;288(8):5581–90.

211. Wang F, Xiong Y, Xiao Y, Han J, Deng X, Lin L. MMPphg from the thermophilic Meiothermus bacteriophage MMP17 as a potential antimicrobial agent against both Gram-negative and Gram-positive bacteria. Virol J. 2020 Dec;17(1):130.

212. Sudiarta IP, Fukushima T, Sekiguchi J. Bacillus subtilis CwlP of the SP-β Prophage Has Two Novel Peptidoglycan Hydrolase Domains, Muramidase and Cross-linkage Digesting dd-Endopeptidase. Journal of Biological Chemistry. 2010 Dec;285(53):41232–43.

213. Ellermeier CD, Hobbs EC, Gonzalez-Pastor JE, Losick R. A three-protein signaling pathway governing immunity to a bacterial cannibalism toxin. Cell. 2006;124(3):549–59.

214. Bose B, Auchtung JM, Lee CA, Grossman AD. A conserved anti-repressor controls horizontal gene transfer by proteolysis. Molecular Microbiology. 2008 Nov;70(3):570–82.

215. InterPro Classification of Protein Families [Internet]. [cited 2024 Feb 1]. Peptidase S24-like (PF00717) - Pfam entry - InterPro. Available from: https://www.ebi.ac.uk/interpro/entry/pfam/PF00717/#PUB00014015

216. Henriques BS, Gomes B, Costa SGD, Moraes CDS, Mesquita RD, Dillon VM, et al. Genome Wide Mapping of Peptidases in Rhodnius prolixus: Identification of Protease Gene Duplications, Horizontally Transferred Proteases and Analysis of Peptidase A1 Structures, with Considerations on Their Role in the Evolution of Hematophagy in Triatominae. Front Physiol. 2017 Dec 12;8:1051.

217. Fogh RH, Ottleben G, Rüterjans H, Schnarr M, Boelens R, Kaptein R. Solution structure of the LexA repressor DNA binding domain determined by 1H NMR spectroscopy. The EMBO Journal. 1994 Sep;13(17):3936–44.

218. Tišáková L, Vidová B, Farkašovská J, Godány A. Bacteriophage endolysin Lyt μ1/6: characterization of the C-terminal binding domain. FEMS Microbiol Lett. 2014 Jan;350(2):199–208.

219. Liu J, Mushegian A. Displacements of Prohead Protease Genes in the Late Operons of Double-Stranded-DNA Bacteriophages. J Bacteriol. 2004 Jul;186(13):4369–75.

220. Hedstrom L. Serine Protease Mechanism and Specificity. Chem Rev. 2002 Dec;102(12):4501–24.

221. Yamaguchi T, Hayashi T, Takami H, Nakasone K, Ohnishi M, Nakayama K, et al. Phage conversion of exfoliative toxin A production in Staphylococcus aureus. Mol Microbiol. 2000 Nov;38(4):694–705.

222. Ladhani S, Joannou CL, Lochrie DP, Evans RW, Poston SM. Clinical, Microbial, and Biochemical Aspects of the Exfoliative Toxins Causing Staphylococcal Scalded-Skin Syndrome. Clin Microbiol Rev. 1999 Apr;12(2):224–42.

223. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research. 2021;49(D1):D480–9.

224. Wang J, Hartling JA, Flanagan JM. The Structure of ClpP at 2.3 Å Resolution Suggests a Model for ATP-Dependent Proteolysis. Cell. 1997 Nov;91(4):447–56.

225. Maurizi Mr. Proteases and protein degradation in Escherichia coli. Experientia. 1992;48(2):178–201.

226. Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP, Davies G. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proceedings of the National Academy of Sciences. 1995 Jul 18;92(15):7090–4.

227. Steen A, van Schalkwijk S, Buist G, Twigt M, Szeliga M, Meijer W, et al. Lytr, a phage-derived amidase is most effective in induced lysis of Lactococcus lactis compared with other lactococcal amidases and glucosaminidases. International Dairy Journal. 2007 Aug;17(8):926–36.

228. Estrella LA, Quinones J, Henry M, Hannah RM, Pope RK, Hamilton T, et al. Characterization of novel *Staphylococcus aureus* lytic phage and defining their combinatorial virulence using the OmniLog® system. Bacteriophage. 2016 Jul 2;6(3):e1219440.

229. Peschel A, Vuong C, Otto M, Götz F. The D-Alanine Residues of Staphylococcus aureus Teichoic Acids Alter the Susceptibility to Vancomycin and the Activity of Autolytic Enzymes. Antimicrob Agents Chemother. 2000 Oct;44(10):2845–7.

230. Becker SC, Foster-Frey J, Stodola AJ, Anacker D, Donovan DM. Differentially conserved staphylococcal SH3b_5 cell wall binding domains confer increased staphylolytic and streptolytic activity to a streptococcal prophage endolysin domain. Gene. 2009 Aug;443(1–2):32–41.

231. Sugai M, Fujiwara T, Akiyama T, Ohara M, Komatsuzawa H, Inoue S, et al. Purification and molecular characterization of glycylglycine endopeptidase produced by Staphylococcus capitis EPK1. Journal of Bacteriology. 1997;179(4):1193–202.

232. Oliveira H, Sampaio M, Melo LDR, Dias O, Pope WH, Hatfull GF, et al. Staphylococci phages display vast genomic diversity and evolutionary relationships. BMC Genomics. 2019 Dec;20(1):357.

233. Lee DS, Kim BK, Kwon SJ, Jin HC, Park OK. Arabidopsis GDSL lipase 2 plays a role in pathogen defense via negative regulation of auxin signaling. Biochemical and Biophysical Research Communications. 2009 Feb;379(4):1038–42.

234. Oh IS, Park AR, Bae MS, Kwon SJ, Kim YS, Lee JE, et al. Secretome Analysis Reveals an *Arabidopsis* Lipase Involved in Defense against *Alternaria brassicicola*. The Plant Cell. 2005 Oct 3;17(10):2832–47.

235. Monzingo AF, Marcotte EM, Hart PJ, Robertas JD. Chitinases, chitosanases, and lysozymes can be divided into procaryotic and eucaryotic families sharing a conserved core. Nat Struct Mol Biol. 1996 Feb;3(2):133–40.

236. Taylor NMI, Prokhorov NS, Guerrero-Ferreira RC, Shneider MM, Browning C, Goldie KN, et al. Structure of the T4 baseplate and its function in triggering sheath contraction. Nature. 2016 May;533(7603):346–52.

237. EMBL-EBI Pfam 34.0 (March 2021, 19179 entries) [Internet]. [cited 2021 Jul 7]. Pfam: Structure: 4HPE. Available from: http://pfam.xfam.org/structure/4HPE

238. Anantharaman V, Aravind L. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biology. 2003;12.

239. EMBL-EBI Pfam 34.0 (March 2021, 19179 entries) [Internet]. [cited 2021 Jul 7]. Pfam: Family: HTH_19 (PF12844). Available from: http://pfam.xfam.org/family/HTH_19

240. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000 May;25(1):25–9.

241. The Gene Ontology Consortium, Carbon S, Douglass E, Good BM, Unni DR, Harris NL, et al. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Research. 2021 Jan 8;49(D1):D325–34.

242. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000 Apr 1;16(4):404–5.

243. Osipovitch DC, Therrien S, Griswold KE. Discovery of novel S. aureus autolysins and molecular engineering to enhance bacteriolytic activity. Appl Microbiol Biotechnol. 2015 Aug;99(15):6315–26.

244. Jones DT. Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices. Journal of molecular biology. 1999;292(2):195–202.

245. Daniel A, Euler C, Collin M, Chahales P, Gorelick KJ, Fischetti VA. Synergism between a novel chimeric lysin and oxacillin protects against infection by methicillin-resistant Staphylococcus aureus. Antimicrobial Agents and Chemotherapy. 2010;54(4):1603–12.

246. Fernandes S, São-José C. Enzymes and Mechanisms Employed by Tailed Bacteriophages to Breach the Bacterial Cell Barriers. Viruses. 2018 Jul 27;10(8):396.

247. Hoshiba H, Uchiyama J, Kato S ichiro, Ujihara T, Muraoka A, Daibata M, et al. Isolation and characterization of a novel Staphylococcus aureus bacteriophage, φMR25, and its therapeutic potential. Arch Virol. 2010 Apr;155(4):545–52.

248. Rashel M, Uchiyama J, Ujihara T, Uehara Y, Kuramoto S, Sugihara S, et al. Efficient Elimination of Multidrug-Resistant Staphylococcus aureus by Cloned Lysin Derived from Bacteriophage ϕMR11. The Journal of Infectious Diseases. 2007;196(8):1237–47.

249. Zhang Q, Xing S, Sun Q, Pei G, Cheng S, Liu Y, et al. Characterization and complete genome sequence analysis of a novel virulent Siphoviridae phage against Staphylococcus aureus isolated from bovine mastitis in Xinjiang, China. Virus Genes. 2017 Jun;53(3):464–76.

250. Abaev I, Juli Foster-Frey, Korobova O, Shishkova N, Kiseleva N, Kopylov P, et al. Staphylococcal Phage 2638A endolysin is lytic for Staphylococcus aureus and harbors an inter-

lytic-domain secondary translational start site. Appl Microbiol Biotechnol. 2013 Apr;97(8):3449–56.

251. Lewis K. Programmed Death in Bacteria. Microbiol Mol Biol Rev. 2000 Sep;64(3):503–14.

252. Chang Y, Shin H, Lee JH, Park C, Paik SY, Ryu S. Isolation and Genome Characterization of the Virulent Staphylococcus aureus Bacteriophage SA97. Viruses. 2015 Oct 1;7(10):5225–42.

253. Liu S, Zhao Y, Hayes A, Hon K, Zhang G, Bennett C, et al. Overcoming bacteriophage insensitivity in *Staphylococcus aureus* using clindamycin and azithromycinat subinhibitory concentrations. Allergy. 2021 May 16;all.14883.

254. Keary R, McAuliffe O, Ross RP, Hill C, O'Mahony J, Coffey A. Genome analysis of the staphylococcal temperate phage DW2 and functional studies on the endolysin and tail hydrolase. Bacteriophage. 2014;4(October 2015):e28451.

255. Sitbon E, Pietrokovski S. New types of conserved sequence domains in DNA-binding regions of homing endonucleases. Trends in Biochemical Sciences. 2003 Sep;28(9):473–7.

256. Hausner G, Hafez M, Edgell DR. Bacterial group I introns: mobile RNA catalysts. Mobile DNA. 2014 Dec;5(1):8.

257. Kala S, Cumby N, Sadowski PD, Hyder BZ, Kanelis V, Davidson AR, et al. HNH proteins are a widespread component of phage DNA packaging machines. Proceedings of the National Academy of Sciences. 2014 Apr 22;111(16):6022–7.

258. Fujiki J, Nakamura T, Furusawa T, Ohno H, Takahashi H, Kitana J, et al. Characterization of the Lytic Capability of a LysK-Like Endolysin, Lys-phiSA012, Derived from a Polyvalent Staphylococcus aureus Bacteriophage. Pharmaceuticals. 2018 Feb 24;11(1):25.

259. van Leeuwen HC, Strating MJ, Rensen M, de Laat W, van der Vliet PC. Linker length and composition influence the flexibility of Oct-1 DNA binding. The EMBO journal. 1997;16(8):2043–53.

260. Haddad Kashani H, Fahimi H, Dasteh Goli Y, Moniri R. A Novel Chimeric Endolysin with Antibacterial Activity against Methicillin-Resistant Staphylococcus aureus. Frontiers in Cellular and Infection Microbiology. 2017;7(June):1–12.

261. UniProt [Internet]. 2020 [cited 2022 Nov 3]. ORF007 - Staphylococcus phage 2638A | UniProtKB | UniProt. Available from: https://www.uniprot.org/uniprotkb/Q4ZD58/entry

262. Sobieraj A, Dunne M, Ernst P, Pluckthun A, Loessner MJ. (CASP target) Crystal structure of the M23 peptidase domain of Staphylococcal phage 2638A endolysin. [Internet]. UniProtKB unreviewed (TrEMBL); 2020 [cited 2022 Sep 11]. Available from: https://www.uniprot.org/citations/CI-EMCGRQ6KBRKDN

263. Sobieraj A, Dunne M, Ernst P, Pluckthun A, Loessner MJ. Cell wall binding domain of the Staphylococcal phage 2638A endolysin. [Internet]. UniProtKB unreviewed (TrEMBL); 2020 [cited 2022 Sep 11]. Available from: https://www.uniprot.org/citations/CI-PNOFG5UGAI4C

264. Dunne M, Ernst P, Sobieraj A, Pluckthun A, Loessner MJ. RCSB Protein Data Bank. 2020 [cited 2022 Nov 8]. RCSB PDB - 6YJ1: The M23 peptidase domain of the Staphylococcal phage 2638A endolysin. Available from: https://www.rcsb.org/structure/6YJ1

265. Dunne M, Sobieraj A, Ernst P, Mittl PRE, Pluckthun A, Loessner MJ. CSB Protein Data Bank. 2020 [cited 2022 Nov 3]. RCSB PDB - 7AQH: Cell wall binding domain of the Staphylococcal phage 2638A endolysin. Available from: https://www.rcsb.org/structure/7AQH

266. Sivaprasad AV, Jarvinen R, Puspurs A, Egan JB. DNA Replication Studies with Coliphage 186 III.¢ A Single Phage Gene is Required for Phage 186 Replication. Journal of Molecular Biology. 1990 May 6;213(3):15.

267. Richardson H, Egan JB. DNA replication studies with coliphage 186: II. Depression of host replication by a 186 gene. Journal of Molecular Biology. 1989 Mar 5;206(1):59–68.

268. Martinez A, Kolvek SJ, Yip CLT, Hopke J, Brown KA, MacNeil IA, et al. Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. Appl Environ Microbiol. 2004 Apr;70(4):2452–63.

269. Vartoukian SR, Palmer RM, Wade WG. Strategies for culture of 'unculturable' bacteria: Culturing the unculturable. FEMS Microbiology Letters. 2010 Apr 27;no-no.

270. Streit WR, Schmitz RA. Metagenomics – the key to the uncultured microbes. Current Opinion in Microbiology. 2004 Oct;7(5):492–8.