University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Therapeutic and Diagnostic Potential of Endolysins Targeting Gram-Positive Pathogens

Thesis presented by

Ellen Murray, B.Sc., M.Sc.

for the degree of

Doctor of Philosophy

School of Microbiology, University College Cork, Cork Ireland

APC Microbiome Ireland, University College Cork, Cork, Ireland

Supervisors: Prof. Colin Hill, Prof. Paul Ross

Head of School: Dr. David Clarke

2024

Table of Contents

Declaration 8

List of Publications 9

Abbreviations 10

Thesis Abstract 12

Chapter 1. The Advantages and Challenges of Using Endolysins in a Clinical Setting. 16

Abstract 17

Introduction 18

Phages. 21

Endolysins. 21

Classification of Phage Lysins. 24

Why Not Just Use Phages? 25

Research Success Stories. 27

Gram-Positive Targets. 27

Gram-Negative Targets. 29

Targeting Biofilms. 31

Synergy. 32

Safety. 32

Limitations of Endolysin Therapy. 33

Drug Delivery Methods. 33

Not A “One Size Fits All” Concept. 34

Regulatory Body Approval. 34

Looking to the Future – Next Generation Lysins 35

Virion Associated Lysins. 35

Chimeric Lysins. 36

Overcoming Barriers in Drug Delivery Methods. 40

Nanoparticles. 40
Lysocins. 41
Current Commercial & Medical Applications of Endolysins. 41
Clinical Trials. 42
Commercial Products. 43
Conclusions. 45
Supplementary Materials 46
References 47
Chapter 2. Assessing the Therapeutic Potential of an Endolysin Present in Six Novel Phages Targeting *Cutibacterium acne*s. 52
Abstract 53
Introduction 54
Results 58
Host Range. 58
Phage Visualisation and Morphology. 58
Genome Analysis. 59
Endolysins of *C. acnes* Phages. 67
Cloning and Recombinant Expression of LysCa5A2. 70
Challenges of insoluble proteins and assessment of lytic activity. 71
Discussion 75
Materials and Methods 79
Bacterial Strains and Culturing Conditions. 79
Sampling, Phage Screening and Purification. 79
Transmission Electron Microscopy. 80
DNA Extraction and Sequencing of C. acnes Phages. 80
Genome Analysis. 81
Morphological Analysis of the Phages. 82

In silico Characterisation of Endolysins Genes. 82
Cloning. 82
Recombinant Protein Expression and Purification. 83
Protein Solubilisation 84
Lytic Activity of Endolysin LysCa5A2. 84
Supplementary Materials 85
Author Contributions 90
References 91
Chapter 3. LysH1 is a Novel *Enterococcus faecium* Phage Endolysin with Activity Against *Ruminococcus gnavus* in a Simplified Human Gut Consortium. 98
Abstract 99
Introduction 100
Results 104
In Silico Analysis of LysH1. 104
Cloning and Expression of Whole Endolysin and its Putative Cell Wall Binding Domain. 105
Spectrum of Activity of LysH1. 106
Lytic Activity of LysH1 111
Biochemical Characterisation of LysH1. 112
Fluorescent Binding Assays of LysH1 CBD-GFP. 113
The Effect of LysH1 on a Simplified Human Gut Community. 117
Discussion 120
Materials and Methods 125
In silico Analysis of LysH1. 125
Bacterial Strains and Growth Conditions. 125
Cloning of Endolysin and Cell Wall Binding Domain. 126
Recombinant Expression and Purification 127

Lytic Activity of LysH1 127
Binding Assays and Fluorescent Microscopy. 128
Analysis of Endolysin Treatment in a Simplified Gut Consortium by Quantitative Real-Time PCR. 129
Supplementary Materials 131
Author Contributions 138
References 139
Chapter 4. Genomic and Phenotypic Analysis of Endolysin-Resistant Mutants of *Ruminococcus gnavus*. 146
Abstract 147
Introduction 148
Results 151
In Silico and Structural Analysis of Endolysin Genes. 151
Cloning, Expression, and Purification of Endolysins. 155
Lytic Activity of Endolysins. 155
Characterisation of *R. gnavus* Mutants. 160
Whole Genome Sequencing of *R. gnavus* ER Mutants. 163
Discussion 168
Materials and Methods 175
Bacterial Strains and Growth Conditions. 175
Bioinformatic Analysis. 175
Cloning. 176
Overexpression in Recombinant E. coli and Protein Purification. 177
Lytic Activity of Endolysins. 177
Generation of Endolysin-Resistant Mutants. 178
Sequencing of Mutants and Analysis of Genomes. 178
Fluorescence microscopy. 179
Supplementary Materials 180

Author Contributions 192
References 193
Chapter 5. Functional Characterisation of a Novel Receptor Binding Domain from a Tail-Associated Lysin in a *Clostridioides difficile* Bacteriophage. 201
Abstract 202
Introduction 203
Results 207
Bioinformatic Analysis. 207
Cloning and Purification of CdRBP and a CdRBP-GFP Hybrid. 209
Functional Analysis of CdRBP-GFP. 209
Potential for Diagnostics. 215
Analysis of Binding Efficiency in Complex Samples. 215
Discussion 219
Materials and Methods 223
Bioinformatic Analysis. 223
Bacterial Strains and Growth Conditions. 223
Cloning. 223
Recombinant Expression and Purification. 224
Binding Assays and Fluorescent Microscopy. 224
Temperature and pH Stability Tests. 225
Author Contributions 226
References 227
Chapter 6. Discussion 231
Concluding Remarks 232
Acknowledgements 235

This work was supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2273_P2, a Science Foundation Ireland's Spokes Programme, which is co-funded under the European Regional Development Fund under Grant Number SFI/14/SP APC/B3032, and a research grant from Janssen Biotech, Inc.

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

Student number: 114308696

Date: 30th June 2024

Signed: 

List of Publications

Reviews

Murray, E.; Draper, L.A.; Ross, R.P.; Hill, C. The Advantages and Challenges of Using Endolysins in a Clinical Setting. Viruses 2021, 13, 680. https://doi.org/10.3390/v13040680

Research Papers

Buttimer C, Sutton T, Colom J, **Murray E**, Bettio PH, Smith L, Bolocan AS, Shkoporov A, Oka A, Liu B, Herzog JW, Sartor RB, Draper LA, Ross RP and Hill C (2022) Impact of a phage cocktail targeting Escherichia coli and Enterococcus faecalis as members of a gut bacterial consortium in vitro and in vivo. Front. Microbiol. 13:936083. doi: 10.3389/fmicb.2022.936083

In Preparation

Ellen Murray, Ekaterina V. Khokhlova, David Hourigan, Lorraine A. Draper, Andrey Shkoporov, Paul R. Ross, Colin Hill. Genomic and Phenotypic Analysis of Endolysin-Resistant Mutants of *Ruminococcus gnavus*.

Ellen Murray, Natalia S. Ríos Colombo, David Hourigan, Lorraine A. Draper, Neda Nezam Abadi, Paul R. Ross, Colin Hill. LysH1 is a novel *Enterococcus faecium* phage endolysin with activity against *Ruminococcus gnavus* in a simplified human gut consortium.

Ellen Murray, Shakhinur Islam Mondal, Lorraine A. Draper, Neda Nezam Abadi, Paul R. Ross, Colin Hill. Functional characterisation of a novel receptor binding domain from a tail-associated lysin in a *Clostridioides difficile* bacteriophage.

Abbreviations

ABB	Anaerobe Basal Broth
AMR	Antimicrobial Resistance
BHI	Brain Heart Infusion
CBD	Cell Wall Binding Domain
CDAD	*C. difficile*-associated disease
CDI	*C. difficile* Infection
CDS	Coding Sequence
CV	Cyclic Voltammetry
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CWH	Cell Wall Hydrolase
DIAMOND	Double Index Alignment of Next-Generation Sequencing Data
EAD	Enzymatically Active Domain
EIS	Electrochemical Impedance Spectroscopy
FMT	Faecal Microbiota Transplant
GFP	Green Fluorescent Protein
IBD	Inflammatory Bowel Disease
IMP	Inosine-5′-monophosphate
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
LB	Lysogeny Broth
LSP	Lipopolysaccharide
MRSA	Methicillin-resistant *Staphylococcus aureus*
NAG	N-acetylglucosamine
NAM	N-acetylmuramic acid

NCBI	National Centre for Biotechnology Information
OD	Optical Density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pLDDT	Predicted Local Distance Difference Test
PG	Peptidoglycan
RBP	Receptor Binding Protein
SDS	Sodium dodecyl-sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEM	Scanning Electron Microscopy
SIHUMI	Simplified Human Intestinal Microbiota
SNPs	Single Nucleotide Polymorphisms
SPE	Screen-Printed Electrode
TAL	Tail-associated Lysin
TEM	Transmission Electron Microscopy
TNF-α	Tumour Necrosis Factor Alpha
TRA	Turbidity Reduction Assay
UC	Ulcerative Colitis
VAL	Virion-associated Lysin
VRE	Vancomycin Resistant Enterococci
WGS	Whole Genome sequencing
WHO	World Health Organisation

Thesis Abstract

Infection as a result of antibiotic resistant microbes (AMR) is a growing health concern, creating a need for the development of novel antimicrobial and diagnostic strategies. Bacteriophage-derived lytic enzymes, endolysins, could offer a promising additional weapon in the battle to overcome AMR. This thesis investigates the therapeutic and diagnostic potential of endolysins targeting gram-positive pathogens *Cutibacterium acnes*, *Ruminococcus gnavus*, *Enterococcus faecium*, and *Clostridioides difficile*. The research conducted focuses on several key objectives: the isolation of novel endolysins, the identification of their catalytic and cell wall binding domains, the characterisation of lytic activity, the assessment of resistance mechanisms and their potential for diagnostics.

Chapter 1 describes the advantages and challenges of using endolysins in a clinical setting. This review explores the potential of bacteriophage-derived endolysins as novel antimicrobials. The advantages discussed here include specificity, synergistic effects with other antimicrobials and the ability to engineer endolysins for a specific gain. Endolysins have shown efficacy against several topical pathogens, including Methicillin-resistant *Staphylococcus aureus* (MRSA), *Listeria monocytogenes* and *Pseudomonas aeruginosa*. Challenges facing the future of endolysin therapy include regulatory hurdles, production and stability, and delivery methods.

C. acnes is a commensal skin bacterium that is associated with the skin condition acne vulgaris. Acne is estimated to affect 9.4% of the world's population, making it the eighth most prevalent disease worldwide. Current treatment plans are inefficient and include the use of antibiotics and harsh topical treatments that are paired with undesirable side effects. In this thesis, six novel phages infecting *C. acnes* were isolated, and their genomes were sequenced. The endolysin genes of these phages were selected for their potential biotechnological applications. The six endolysin genes were determined to be identical. The endolysin LysCa5A2, derived from phage 005A2, was expressed. This study revealed that LysCa5A2 is a highly insoluble protein, and its expression in *E. coli* BL21(DE3) led to the formation of inclusion bodies. In Chapter 2, we describe the largely unsuccessful attempts made to recover a native bioactive protein.

The human microbiome plays an important role in human health and disease. While assessing the potential of a new therapeutic where selectivity is often postulated

as an important attribute, it is vital to ensure that the candidate drug is specifically targeting the microbe of choice. In Chapter 3, we describe the isolation of LysH1 – a novel endolysin derived from an *E. faecium* phage. Structural analysis of this endolysin revealed a two-domain structure composed of a catalytic domain and a previously uncharacterised domain. Binding assays with a green fluorescent protein (GFP) fusion protein allowed us to functionally characterise this unknown domain as the cell wall binding domain of LysH1. While initially investigated for its potential to manage vancomycin-resistant enterococci (VRE), host range analysis showed that LysH1 also has cross-species activity against *R. gnavus*. To assess the therapeutic potential of LysH1, a simplified human intestinal microbiota (SIHUMI) was established. Here, we show that LysH1 can selectively target *R. gnavus* without impacting the other members of the consortium.

Endolysin therapy is commended for many reasons, but one of particular note is the reported lack of resistance mechanisms. In Chapter 4, we isolated endolysins that could target *Ruminococcus gnavus*, a commensal gut bacterium associated with inflammatory bowel disease (IBD). Five novel endolysins were successfully cloned and recombinantly expressed, and one was selected for further testing (LysIBDN1). LysIBDN1 initially demonstrated effective lytic activity against *R. gnavus*, but further analysis revealed the emergence of endolysin resistant (ER) mutants of *R. gnavus*. Genomic analysis revealed a series of SNPs, deletions, and insertions in the genomes of mutant strains, largely associated with the stringent stress response. Phenotypic analysis demonstrated significant fitness costs of resistance. This study provides insights into the mechanisms of endolysin resistance.

The functional characterisation of a novel receptor binding domain from a *Clostridiodes difficile* bacteriophage tail-associated lysin is described in Chapter 5. *C. difficile* infection (CDI) is a significant nosocomial infection. To reduce the transmission of CDI in healthcare settings, it is important to be able to rapidly monitor *C. difficile* levels to inform treatment options and to control infection. In this study, we assess the potential of this receptor binding protein (CdRBP) as a means of identifying *C. difficile*. Functional analysis with a GFP-fusion protein revealed that the domain can effectively bind *C. difficile* strains. However, analysis with more complex samples revealed that CdRBP displays off-target binding. In the appendix of Chapter 5, we describe the use

of CdRBP in an electrochemical biosensor platform. With further optimisation, we propose that CdRBP could be a promising candidate for the detection of *C. difficile.*

This thesis contributes to the growing body of knowledge of endolysin therapy. Critical analysis of the endolysins described in this thesis indicates that, while some candidates show promise as therapeutics, we must be aware of challenges, including the emergence of resistance and off-target effects. Here, we offer insights into the specificity of binding domains, emphasise the need for microbiome-driven studies, and discuss potential resistance mechanisms and their consequences.

Chapter 1. The Advantages and Challenges of Using Endolysins in a Clinical Setting.

This chapter was published as a review in Viruses.

Ellen Murray, Lorraine A. Draper, R. Paul Ross, and Colin Hill. "The Advantages and Challenges of Endolysins in a Clinical Setting." Viruses 2021, 13(4), 680.

https://doi.org/10.3390/v13040680

Abstract

Antibiotic resistant pathogens are increasingly more prevalent and problematic. Traditional antibiotics are no longer a viable option for dealing with these multidrug resistant microbes and so new approaches are needed. Phage-derived proteins such as endolysins could offer one effective solution. Endolysins are bacteriophage-encoded peptidoglycan hydrolases that act to lyse bacterial cells by targeting their cell's wall, particularly in gram-positive bacteria, due to their naturally exposed peptidoglycan layer. These lytic enzymes have received much interest from the scientific community in recent years for their specificity, mode of action, potential for engineering, and lack of resistance mechanisms. Over the past decade, a renewed interest in endolysin therapy has led to a number of successful applications. Recombinant endolysins have been shown to be effective against prominent pathogens such as MRSA, *Listeria monocytogenes, Staphylococcus* strains in biofilm formations, and *Pseudomonas aeruginosa*. Endolysins have also been studied in combination with other antimicrobials, giving a synergistic effect. Although endolysin therapy comes with some regulatory and logistical hurdles, the future looks promising with the emergence of engineered "next generation" lysins. This review will focus on the likelihood that endolysins will become a viable new antimicrobial therapy and the challenges that may have to be overcome along the way.

Introduction

Antibiotic resistance presents a growing threat in terms of treating infections. It is a widely recognised problem, and the solution will depend at least in part on developing alternative antimicrobials. The World Health Organisation (WHO) has recently referred to antibiotic resistance as “one of the biggest threats to global health, food security and development today” [1]. The WHO has also stated that this ever-emerging resistance will inevitably lead to higher medical costs, longer hospital stays and higher rates of mortality. It is clear that this phenomenon will impact society, including economic, political and social aspects, and so it is of interest to all in the scientific community to identify and commercialise alternatives to standard antibiotic therapy. Bacteriophage-derived proteins such as endolysins provide one potential solution to this pressing issue. Bacteriophages (phages) have long been used to combat bacterial infections in some parts of the world, but the use of their endolysins is relatively new and is showing great promise. Unlike traditional broad-spectrum antibiotics, endolysins can target specific bacterial taxa with little collateral damage to the surrounding microbiome, and a lack of bacterial resistance mechanisms has also been noted as a positive aspect of phage lysin therapy. These features have resulted in lytic enzymes such as endolysins being assigned to a new class of antimicrobial agents termed “enzybiotics”.

Although phage therapy gained some popularity in the 1920’s, it was quickly overshadowed by the discovery and development of broad-spectrum antibiotics. The discovery of penicillin in 1928 placed antibiotics centre stage in terms of antibacterial therapies. For many decades it was believed that broad-spectrum antibiotics were more desirable than a more targeted approach. This idea was further endorsed by pharmaceutical companies as it was much more advantageous to create a product that would be efficacious against multiple infectious pathogens. The era dubbed “The Golden Age of Antibiotics” spanned the 1940’s to 1950’s and was responsible for the discovery and production of many of the antibiotics still in use today. At the beginning of this period, little was understood about the potential for the development of resistant mechanisms in bacteria, so widespread antibiotic use was not considered to be problematic. It wasn’t until the epidemic of penicillin-resistant *Staphylococcus aureus* in the 1950’s that concerns were raised. It was shown that the ability to hydrolyse the beta-lactam ring of penicillin had spread across many different bacteria. To address

this issue, methicillin, a synthetic version of penicillin, was introduced. This was followed by the emergence of methicillin-resistant *S. aureus* in the late 1950's. By the 1980's the discovery and development of new antibiotics had slowed dramatically. Since then, only a handful of novel antibiotics have made it through the commercial pipeline while resistance continues to spread.

While the Western world witnessed the rise of antibiotic resistance, phage therapy and research continued to be used, particularly in Eastern Europe. A 1983 study showed that bacteriophages could combat bone and joint infections caused by antibiotic resistant strains of *Staphylococcus* and *Pseudomonas* [2]. In more recent years, endolysin therapy has become a popular area of research. This involves applying the phage-derived protein exogenously to the bacterium, resulting in cell lysis. Controlled clinical trials of phage therapy have also become more widespread.

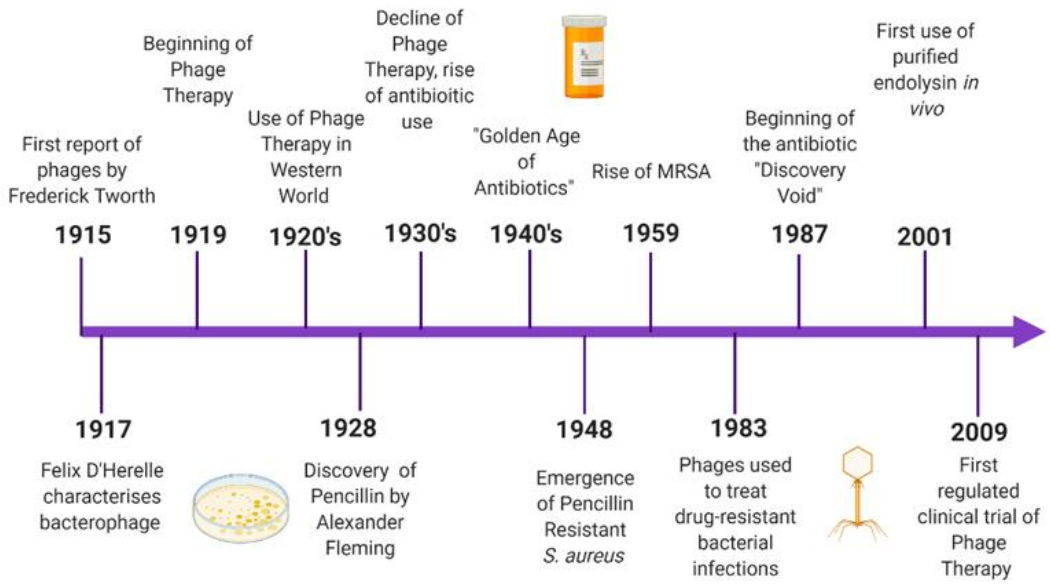


Figure 1. Timeline of Phage Therapy. Outlined is the history of phage therapy vs antibiotic therapy, from the discovery of bacteriophages to the present day.

Phages.

Phages are viruses that can specifically infect and kill bacteria. They were discovered independently by Henry Twort and Felix d'Herelle in 1915 and 1917, respectively. After the discovery of phages, they were quickly put to use against bacterial infections in animals, soon followed by use in humans. In 1919, a phage cocktail was administered to a child suffering from bacterial dysentery and proved to be successful with no side-effects [3]. This was the beginning of phage therapy. As a result of a collaboration between French researcher Felix d'Herelle and a team of Georgian scientists led by George Eliava, phage therapy became increasingly popular in the Soviet Union, while it was abandoned in the Western world with the commercialisation of antibiotics [4]. However, in recent years, the emergence of multi-drug resistant bacteria has left the world scrambling for alternative treatments. This worldwide issue could give phage therapy a new lease of life.

Phages are the most ubiquitous and most abundant biological entity on earth. They are obligate parasites that require a bacterial host in order to replicate and multiply. They can undergo two life cycles, lytic or lysogenic. The first step is attachment of the phage to the host bacterium. It achieves this by interacting with receptors on the host cell, followed by tight adsorption onto the cell and the subsequent injection of its genome. The lytic process involves the phage hijacking the host cell machinery in order to produce its viral progeny until lysis of the bacteria releases phage progeny into their immediate environment. Lysogenic or temperate phages form a more long-term relationship with the bacterial host. During lysogeny, the viral genome integrates into the bacterial chromosome as a prophage. At a given stage, usually when the bacteria is in a condition of stress, the prophage can exit lysogeny and begin its replication, lysis and release [5]. The lysis of the bacterial cell that occurs at the end of both replication cycles is made possible by an enzyme called an endolysin.

Endolysins.

Endolysins are phage-encoded peptidoglycan hydrolases. These enzymes accumulate within the host cell independently of the phage virion, together with an associated holin protein. As endolysins don't possess their own signal sequences, the purpose of the holin is to allow the endolysin access to the bacterial peptidoglycan by

creating pores in the cytoplasmic membrane. This is a highly regulated sequence of events that is only triggered when the holin concentration reaches a certain threshold. Endolysins can now access and degrade peptidoglycan, upsetting the osmotic balance in the cell resulting in lysis and ultimately cell death [3]. Endolysins and holins together are essential to a successful phage infection process (**Figure 2**).

Endolysins have been of interest as potential antimicrobials because this lytic activity can also occur when the lysin is applied exogenously. They are particularly active in the case of gram-positive bacteria that do not have a protective outer cell membrane. If administered in the form of recombinant proteins applied externally to the bacterial cells, they can cause rapid lysis and cell death, a desirable trait in an antimicrobial agent. In the case of gram-negative bacteria, the presence of an outer membrane limits but does not completely prevent the use of endolysin treatment.

Although endolysins have been tested in a number of studies to demonstrate their potential as antimicrobials, there are few cases of these phage-derived proteins actually being put to use in a clinical setting. How is it that such an apparently promising therapeutic has not progressed from the research laboratory to the clinic? Here we aim to assess the advantages and disadvantages of endolysin therapy, and the challenges of implementing endolysins as antimicrobial agents.

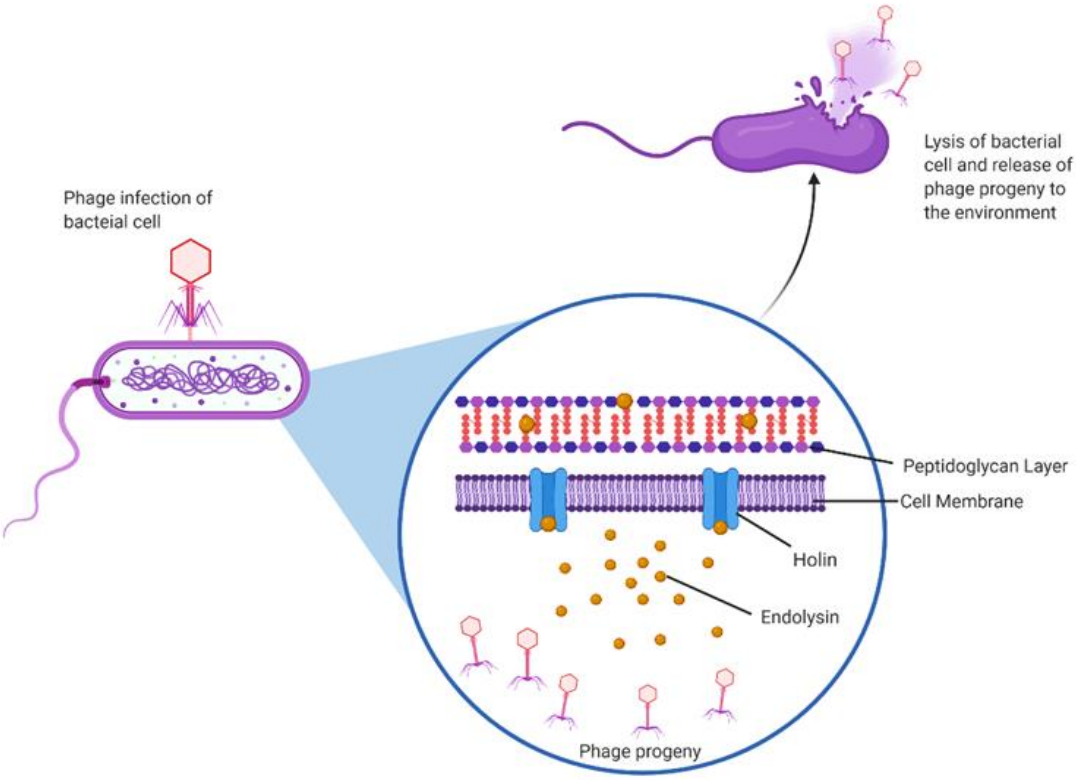


Figure 1. Phage-Endolysin Mechanism. The process of a basic phage infection of a gram-positive bacterial cell, and the role of endolysins in infection.

Classification of Phage Lysins.

Phage lysins have been reported to largely consist of a two-domain modular structure and are usually greater than 25 kDa in size. The two domains are made up of the enzymatically active catalytic domain and a domain which is responsible for cell wall binding. Lysins of gram-positive bacteria usually consist of one or more enzymatic domains and a cell wall binding domain. In contrast, gram-negative lysins are often composed of a single catalytic domain. The cell wall binding domain of gram-positive endolysins can be responsible for their specificity to bacterial targets due to their recognition and binding of ligands noncovalently to a particular substrate in a target cell wall. The binding domains of gram-positive targeted lysins tend to confer their narrow host range, whereas gram-negative lysins recognise the conserved peptidoglycan composition of their targets, which guides their specificity [6].

The bacterial cell wall possesses a mixture of components that are vital for its structure, the balancing of osmotic stress, and regulating what gets in and out of the cell. The cell wall is made up mostly of peptidoglycan (PG). These polysaccharide monomers consist of a disaccharide N-acetylglucosamine (NAG), linked by a short peptide bridge to N-acetylmuramic acid (NAM) residues. The cell wall of bacteria is constantly broken down and remade by a number of enzymes to facilitate processes such as cell division, sacculus expansion during growth and peptidoglycan remodelling [7].

Endolysins are generally separated into groups based on their cleavage sites. These groups are lysozymes (N-acetylmuramidases), glycosidases (N-acetyl-β-D-glucosamidases), N-acetylmuramoyl-L-alanine amidases (amidases) and L-alanoyl-D-glutamate endopeptidases [8]. These enzymes may be placed under the umbrella term of phage-encoded cell wall hydrolases. Endolysins are usually composed of one of these four N-terminals catalytic domains with a variation of a cell wall binding domain.

Lysozymes (N-acetylmuramidases) kill bacteria through targeted hydrolysis. Peptidoglycan polymers are held together by β-1,4 glycosidic links between the NAG and NAM monomers. Lysozyme acts by hydrolysing these bonds, thus disrupting the structural integrity of the peptidoglycan cell wall [9]. This results in an imbalance of turgor pressure, causing the bacterial cell to lyse. Glycosidases (N-acetyl-β-D-

glucosamidases) act to catalyse the hydrolysis of the glycosidic linkage. N-acetylmuramoyl-L-alanine amidases, more commonly referred to as peptidoglycan amidases, act by hydrolysing the amide bond that separates the glycan strand from the stem peptide that lies between the N-acetylmuramic acid and L-alanine residues. L-alanoyl-D-glutamate endopeptidases and interpeptide bridge-specific endopeptidases target the peptide that is composed of the L-Lysine and D-Alanine link [8].

Why Not Just Use Phages?

Although endolysins have been seen to have promising results in a research setting, one might wonder why a clinician would not just use the parental phage? If the endolysin already exists as a natural part of a phage it might seem counterproductive to clone the protein into a recombinant expression system. The use of genetic modification is a process that will not always find favour, particularly in the food industry, but lysins have several advantages over bacteriophages.

While the production of most bacteriophages in a laboratory setting is a well-established process with reliable protocols in place, the scaling-up of production for industry still poses many challenges. In a lab setting it is relatively easy to propagate well characterised phages for experimental use once a phage, host and selective media are identified. By nature, phages require a bacterial host in order to replicate. Given that the aim of the majority of commercial companies is usually to eradicate a particular pathogen, it is likely that the host strain required to propagate a phage of interest is pathogenic in nature. Having to grow such strains, many of which are potentially multi-drug resistant, in high volumes to generate a high yield of phage is not a very desirable concept. It introduces an array of safety issues, extra measures to be taken to ensure contamination does not occur, higher costs of production and a great deal of time. One way of avoiding this issue is using a “surrogate” host instead of the original pathogen [10]. This involves the identification of a non-pathogenic bacteria in the host-range of a given phage that would be suitable to facilitate propagation, however this is not always feasible.

Even in the case of a successful surrogate being available and where large-scale production is possible, a number of disadvantages are still associated with the use of phages in a clinical setting. Several roadblocks stand in the way of such therapy making it to market, and many of these involve regulatory issues. There is still some question around the use of natural phages and how they will behave in their human host. Although phages are abundant and exist naturally in the human body, the introduction of a phage in order to target a specific bacterium could potentially have long-term off-target effects. Unlike endolysins, the phage virion will interact with the bacterial host and use them for their own replication and potential evolution. It is assumed that the introduction of a phage for therapeutic use will only persist as long as its pathogenic host exists in the environment. In theory, once the infection is eradicated, the phage will not have a host on which to replicate and should pass through the human body. However, phages could potentially mutate and find another target within the microbiome.

A number of phages have also been identified that can facilitate horizontal gene transfer. One of the most well characterised phages that has played a role in this gene transfer is the temperate phage CTX, which targets *Vibrio cholerae*. CTX is a filamentous, single stranded DNA phage. It is widely reported that the CTX prophage is responsible for the presence of genes encoding the cholera toxin *ctxAB*. Strains that are capable of producing this toxin, such as *V. cholerae* O1 El Tor and *V. cholerae* O139, are more pathogenic than their predecessors as this toxin is the causative agent of diarrhoea which is a classical feature of cholera infection. In the late 1990's it was discovered that the genes for the notorious cholera toxin were encoded in the genome of ϕCTX [11]. The genome of ϕCTX is split into two parts – the RS2 sequence and the core region. The RS2 region is responsible for prophage replication and integration. The core region consists of a number of genes associated with phage assembly and structural formation, but also includes elements *zot* and *ace* that code for cholera toxin subunits. Once the phage carrying these elements infects susceptible *V. cholerae* cells they integrate their DNA into the chromosome of *V. cholerae* at the *attB (dif)* integration site [12]. This process facilitates the horizontal gene transfer of toxin *ctxAB*. The ability of phages to cause this type of new toxicity in bacteria is a clear cause for concern and places doubt over their safety profile as a therapeutic agent.

It has also been reported that phages themselves may succumb to resistance mechanisms. The ability of bacteria to become resistant against phages has been recognised for a long time. A 2007 study shed light on one of the mechanisms that allow resistance to occur. This report involved the bacterium *Streptococcus thermophilus* and focused on clustered regularly interspaced short palindromic repeats (CRISPR)-mediated resistance to phage infection. The authors showed that there was a correlation between the difference of spacers in phage-resistant and phage-sensitive mutants of *S. thermophilus*. The addition of new CRISPR spacer units was associated with the development of resistance against phage. It was deduced that these spacers were actually derived from the incoming phage DNA, and that if the spacers were a perfect match to the phage DNA, the bacteria were resistant to that phage. Any discrepancy in the sequence meant the bacteria remained sensitive to phage infection. The study also showed that the removal of CRISPR spacers could reverse the resistance phenotype [13].

Research Success Stories.

Endolysin therapy usually works best with gram-positive bacteria, due to their natural structure of exposed peptidoglycan cell walls. Applying endolysins exogenously, lysis occurs without the need for holins or other partner enzymes. This section will discuss just a few of the many gram-positive bacteria that have become targets for endolysin therapy.

Gram-Positive Targets.

Staphylococcus aureus is a gram-positive bacterium that has been the focus of several endolysin studies in recent years. Multi-drug resistant *S. aureus* (MRSA) is a superbug that can cause infection in a variety of sites on the human body and is often acquired in a hospital setting. These nosocomial infections are difficult to treat and have often been acquired by an individual who is already ill. As treatment of staphylococcal infections with antibiotics is often unsuccessful, endolysin therapy shows promise to tackle this bacterium.

The food industry is another area that has an issue with *Staphylococcus* contamination, particularly in terms of biofilm formation. A 2019 study looked at using

endolysins as an alternative to eliminate *S. aureus* biofilms that form on food contact surfaces. The phage-derived recombinant lysin “LysCSA13”, produced by bacteriophage CSA13 was cloned into an *E. coli* expression system. The resulting protein was used against staphylococcal strains on a number of common surfaces, including polystyrene, glass and stainless steel. LysCSA13 demonstrated strong lytic activity and reduced the biofilm mass by 80-90% across all conditions. Treatment resulted in up to a 3-log unit reduction compared to the untreated control [14]. An added advantage to using endolysins as a biocontrol agent in the food industry and in food production environments is their specificity to bacterial cells. The use of such antimicrobials would have no effect on their human host and little to no effect on non-target bacteria.

Efforts have also been turned to targeting spore-forming gram-positive bacteria such as *Bacillus cereus* and *Clostridium* with phage-derived lysins. *B. cereus* and *Bacillus anthracis* are pathogenic bacilli that can cause serious harm to their hosts in the form of food poisoning and anthrax, respectively. Bacteriophages were historically used as an indicative test for anthrax poisoning due to their high specificity to the bacterium. The *Clostridium* genus includes a number of significant human pathogens including the bacteria responsible for botulism and tetanus, along with the hospital-acquired gut infection due to *Clostridioides difficile* [15].

A 2012 study used an endolysin designated LysB4 that was isolated from the *B. cereus* phage B4. This lysin was characterised as an endopeptidase with a two-domain structure, cleaving the peptidoglycan L-Ala and D-Glu bonds. This endolysin had strong lytic activity against *B. cereus* over a wide range of environmental conditions. It was important to address each of these parameters due to the notorious ability of spore-formers to survive in environmental extremes. Strong lytic activity was observed at pH 8.0-10.0, and the lysin remained active at pH 2.0-10.5 over 30 minutes incubation. A range of temperatures was also examined from 37 ^{o}C to 75 ^{o}C, and the optimal temperature was 50 ^{o}C. However, the protein was inactivated at 55 ^{o}C and above. NaCl concentrations were also tested, with the activity of LysB4 reducing as NaCl concentrations increased, resulting in an approximate 60% decrease of activity at 200 mM NaCl. The performance of LysB4 in each of these tests suggests it as a promising candidate for the biocontrol of spore-forming bacteria [16].

An endolysin has been cloned from the *Clostridium* bacteriophage CD27. This endolysin, CD27L, was initially tested against *C. difficile* cells and showed promising lytic activity. It was subsequently tested against a total of 30 *C. difficile* strains and killed all strains examined. A significant finding of this study was that even though the endolysin is active against *C. difficile*, it showed little to no activity against *Clostridium*-like species that are commensal to the human gut microbiome. This observation means that phage-derived therapy for *C. difficile* infections could be used without risking collateral damage to the surrounding community. This is a property that cannot be matched by the antibiotics that have typically been used to treat these kinds of infections [17]. However, several challenges such as delivery to the target site remain an issue for endolysin therapy, especially in such a complex environment as the human gut.

Other studies have also focused on the potential of endolysins to aid or replace antibiotics in the food production chain. *Listeria monocytogenes* is a problematic pathogen in the production of many foodstuffs such as processed meats and dairy products, particularly soft cheese. PlyP100 is a recombinant endolysin that has been engineered to cope with the environmental conditions in which *Listeria* is often found. These include extremes of heat (pasteurization and heat-treatment steps), high NaCl concentrations and low pH. In addition, it was shown that PlyP100 could continue to act on *L*. *monocytogenes* for up to four weeks in refrigerated storage [18]. This antimicrobial protein shows considerable promise as a specialized, cost-effective, and highly efficient method of biocontrol in terms of food processing, shelf life and storage capacity.

Gram-Negative Targets.

Gram-negative endolysin therapy is not straightforward. Due to the presence of a protective outer membrane, phage lysins do not have a naturally exposed target. The outer membrane is composed of phospholipids and lipopolysaccharides. The lipopolysaccharide molecules are held together by the phosphate bonds present between the acidic sugars. Researchers have been creative in finding a way to get past this hurdle and allow the endolysins to freely access the peptidoglycan cell wall. Pre-treatments have been utilised to overcome the cell membrane barrier. One such method is using EDTA to permeabilise the gram-negative cell membrane. EDTA is a

chelating agent that removes outer membrane-stabilizing divalent cations from their binding site, disrupting the membrane and leaving it vulnerable to lysin attack [19]. Permeabilizing the outer membrane may also be carried out mechanically. High hydrostatic pressure (HPP) is an alternative method of food preservation that can inactivate some microorganisms and enzymes. HPP has been used to make gram-negative bacteria sensitive to molecules such as bacteriocins and anti-microbial peptides by optimising parameters such as time and pressure [20].

An alternative approach to targeting gram-negative bacteria is the fusion of endolysins to components that allow them to gain access to the peptidoglycan layer. One such example is the fusion of the *E. coli* phage lysin Lysep3 with the *Bacillus amyloliquefaciens* endolysin binding domain D8. Lysep3 alone does not possess the catalytic ability to breach the outer membrane of *E. coli*, however when engineered to fuse with the D8 cell wall binding domain it can lyse the bacteria from the outside. The D8 domain has the ability to disrupt the integrity of the outer membrane of the bacteria, thus allowing the enzyme portion to reach the target peptidoglycan membrane [21].

Pseudomonas aeruginosa is a bacterium that is known for its resistance to antibiotics, making it a clear candidate for endolysin therapy. It is often a hospital-acquired infection that can be responsible for pneumonia, urinary tract infections and bacteraemia. In 2017, a phage-derived endolysin LysPA26 was isolated with lytic activity against a multidrug resistant strain of *P. aeruginosa* without the use of an outer membrane permeabilizer. The study published impressive results showing the lysozyme decreased bacterial numbers by up to four log units in 30 minutes. It was also shown to be pH and heat stable. Furthermore, the endolysin had a broad-range of activity, also having an effect against other gram-negative bacteria including *Klebsiella pneumonia*, *Acinetobacter baumannii* and *E. coli*, all of which are known to be multi-drug resistant [22]. LysPA26 is no doubt a strong candidate for the future of endolysin therapy. Being able to selectively target gram-negative bacteria in such an effective manner without the use of antibiotics, as is shown in this study, is indeed possible and could be a first-choice therapy in the future.

Targeting Biofilms.

Bacterial biofilms are notoriously difficult to eradicate. Biofilms form when bacteria attach to a surface and grow as a connected community of cells. This layered formation means that only a certain number of bacteria on the outer layer of the biofilm are vulnerable to antimicrobial agents while the bacteria within are protected and can continue to proliferate. This often establishes a chronic infection resulting in long-term antibiotic use. Due to the nature of this treatment, it is common for biofilm-forming species of bacteria to become resistant to the antibiotics used. Endolysins have recently shown promise as a remedy to these recalcitrant infections.

When *Staphylococcus* strains cause infection in the form of a biofilm, they are often difficult to treat because many of these strains are already multi-drug resistant. In 2018, endolysin LysGH15 was shown to have antimicrobial effects against a number of these strains. In this study *S. aureus*, *S. epidermidis*, *S. haemolyticus* and *S. hominis* were tested both *in vitro* and in a mouse model. In the *in vitro* studies, all *Staphylococcus* strains showed sensitivity to LysGH15 with up to a four-log reduction in bacterial counts in as little as 30 minutes at a concentration of 20 µg/ml. In biofilm formation studies, a dose of 50 µg/ml proved to be effective. Established biofilms of various *Staphylococcus* species were also investigated. In this instance, biofilms were allowed to form for up to 72 hours and were treated with 100 and 150 µg/ml of LysGH15. At the concentration of 150 µg/ml, levels of some established biofilms were significantly reduced by the lysin. Other strains showed a lesser, but still significant, reduction of 51-81% of the biofilm *versus* PBS controls.

In vivo studies were performed to investigate the potential of LysGH15 to protect mice from *S. epidermidis* infections. Mice were first injected with a lethal dose of *S. epidermidis*. One hour post-infection LysGH15 was administered at either 5, 10 or 20 µg/ml. All mice treated with the PBS control died within 48 hours, but the mice given the endolysin made a significant recovery with a decrease in bacteraemia. LysGH15 was also tested for its potential to induce resistance in the bacteria strains used during this study. This was carried out on both vancomycin resistant and vancomycin sensitive strains of *S. epidermidis*. It was reported that no spontaneous resistant mutants were recovered and cells collected from each generation of sub-cultured cells retained their sensitivity to the lysin [23]. This study highlighted many

favourable traits for endolysins against biofilms, including the ability to penetrate the core of the biofilm formation and a lack of resistance emerging against the lysin.

Synergy.

Although there are many cases of individual endolysins showing success as antimicrobials, there have also been instances where more than one lysin have been used together, or a lysin was combined with another class of antimicrobial, to obtain a synergistic effect against infection.

In 2010, a study was published that showed the synergy between a chimeric lysin and a traditional antibiotic against methicillin resistant *S. aureus*. The chimeric lysin ClyS was formed through the fusion of the catalytic domain of phage lysin phiNM3 with the cell wall binding domain of another *S. aureus* endolysin. The resulting lysin was shown to act against methicillin resistant, vancomycin intermediate resistant and sensitive strains of *S. aureus* in a mouse model. However, the study went on to investigate the combination of this engineered phage-derived lysin and oxacillin. *In vivo* experiments showed that in an MRSA septicaemia model, mice injected with the control had a 13% survival rate. In individual treatment groups, the survival rate was between 30-35%. However, when lysin ClyS was combined with oxacillin the survival rate was raised significantly to 80-82% [24].

Cases like these show that, in some scenarios, antibiotics do not need to be replaced entirely but can be paired with molecules like lysin to make them more effective at lower doses. This could give traditional antibiotics a second lease of life.

Safety.

Although a lot of research is being done on the efficacy of endolysins against bacterial targets, not a lot of literature has been published that focuses on their safety profile. However, a 2018 study looked at the safety of two endolysins that target *Streptococcus pneumoniae* [25]. Endolysins PaI and CpI-1 were tested for their toxicity and preclinical safety. Immune monitoring was carried out on human macrophages and pharyngeal cells. Once endolysin was added to the cell lines, changes in gene

expression were monitored over the following six hours. This study also covered *in vivo* toxicity by monitoring IgG levels in mice exposed to lysin over 30 days.

The study found that there were no notable changes in inflammatory gene expression, cellular and non-cellular immune responses or microbiome changes in the mammalian cells or in systemic function. In the mouse studies, IgG levels did rise, however this effect was expected due to the proteinaceous nature of lysins. Despite this response, the immune cells did not have a negative effect on the catalytic ability of the endolysins. Overall, it was shown that endolysins PaI and CpI-1 do not present any significant levels of toxicity. Studies like these are important to create a body of literature around the safety of endolysins and their suitability for clinical use.

Limitations of Endolysin Therapy.

Although there are many positives to the use of phage-derived lysins in a clinical setting, their widespread use is not without challenges.

Drug Delivery Methods.

Although there are a handful of endolysin products on the market, there is a common theme in their method of delivery in that the vast majority are applied topically. While ideal in cases such as skin infections, topical application of lysins is not suitable for many parts of the body, for instance in treating gut infections.

The most common route for gut targeted therapies, and certainly the most popular amongst patients, is the oral delivery route. The majority of drugs on the market are either in tablet or liquid form and are usually taken orally. However, this is not without difficulty. Given that drugs in this form must transit the stomach and digestive system, they are immediately vulnerable to an arsenal of enzymes, different pH levels and mechanical digestion. These can all affect the integrity of the drug molecule and the bioavailability of the drug systemically.

As endolysins are proteinaceous, they can quickly be affected by these processes, rendering them useless. Stomach acid can disrupt the structure of certain endolysins. Proteases such as trypsin, chymotrypsin, pepsin and peptidase act by degrading proteins. Many endolysins possess cleavage sites that can be targeted by

these enzymes [26]. In some instances, the oral route of delivery is not always suitable, for example targeted delivery to the lung. In recent years a number of viable lysins have been isolated to combat lung pathogens but getting them to the site of infection effectively is not a straightforward task. Similar to the human gut, the lungs also have a number of immune cells and enzymes that would quickly target proteins and peptide drugs. Lysins would likely be degraded by proteases or alveolar macrophages [27]. In order to avoid these natural defence systems, endolysins would have to be further engineered.

Not A "One Size Fits All" Concept.

Although a benefit of endolysin and phage therapy is their specificity, there may be a practical issue that is overlooked in the scramble to find an alternative to antibiotics. Of course, this targeted approach is of vital importance in a world where antibiotic-resistance is common, especially when scientists were not ready for the emergence of resistance to the popular broad-spectrum category of antibiotics.

In an ideal world endolysin/phage-related therapy in general would be sufficiently broad spectrum that a patient in a doctor's office could describe their symptoms and a certain phage or a standard cocktail would be prescribed. In specialised settings lysins have shown positive results on specific infections, many of them serious cases. However, the majority of antibiotics are prescribed post-operatively in hospitals or are used by members of the public who have simple illnesses such as tonsilitis, pulmonary and otic infections, to name a few. Is it practical to have such a targeted and personalised approach to these common illnesses? How can we bridge the gap between the use of a precise phage cocktail given to a patient with an already identified pathogen, and the use of antibiotics in an individual with an as-yet unidentified infection? Personalised medicine is constantly evolving and hopefully in years to come will be the gold standard of practice. However, although it has a bright future, it is still in its infancy, and it is unknown when it will become common practice.

Regulatory Body Approval.

In Europe, the European Medical Agency (EMA) is the regulatory authority responsible for deciding the fate of new drugs and therapeutics. This is made up of 50 regulatory authorities across 31 countries. This committee oversees many aspects of a new drug

such as its manufacture and production, marketing, clinical trials and safety monitoring [28]. Endolysins are most commonly produced through recombinant DNA technology. This method of expressing the protein of interest in a host vector, such as *E. coli*, has a specific set of requirements if they are to meet EMA guidelines. Many endolysins showing promise are genetically engineered to have a specific action, and so face more hurdles on the road to production and use in clinical settings. The requirements are listed under 2001/83/EC and EC regulation 726/2004.

Currently there is only one endolysin-based treatment that has made it through EU restrictions for human use and has been accepted as a medical product. Lysin SA.100, otherwise known as Gladskin, is produced by Micreos. Given the amount of research and successful isolation of functional endolysins with viable medical applications, this success rate is extremely low. However, there may be more hope for veterinary-based endolysin products as the restrictions outlined for animal use allow for endolysins to treat certain infections [29].

Looking to the Future – Next Generation Lysins

In recent years a new generation of phage-derived lysins have been proposed. These new and improved lysins may allow us to overcome existing hurdles and make phage lysin therapy more accessible. Modifications of endolysins allow researchers to create an “ideal” lysin that maintains high activity while possessing other desirable traits depending on the circumstances of its use. An ideal lysin might have properties such as thermostability, high solubility to make it bioavailable, may be highly targeted or broad-spectrum depending on the target infection, may be protease resistant if being used in the gut and should be economical to produce. Engineering and specifically designing endolysins gives countless opportunities for a popular, commercially used product to be produced.

Virion Associated Lysins.

Virion-associated peptidoglycan hydrolases (VAPGH) or virion-associated lysins (VAL) are another class of phage-encoded proteins that have antimicrobial properties. These are tail-associated lytic enzymes associated with the structure of the phage tail and are usually termed exolysins. These proteins are active at the initial stages of

infection during adsorption, allowing the phage tail to penetrate the bacterial cell wall. They are found to act specifically on the cell wall of viable and dividing bacteria with a glycosidase or endopeptidase mode of action. These proteins have also been shown to be promising as they have a dual catalytic domain which reduces the risk of the emergence of resistance.

In recent years, VALs have been engineered to target certain bacteria by fusing them into a chimeric lysin. A 2015 study used a chimeric lysin EC300 to target *E. faecalis* during its growth phase [30]. EC300 is a hybrid of an endopeptidase domain of a VAL and a cell wall binding domain of an amidase endolysin. The study showed the engineered chimeric lysin to have a clear advantage over each of its parts alone, and even over commercially available antibiotics. An earlier report also illustrated the therapeutic potential of phage-encoded chimeric proteins [31]. This study involved the joining of *S. aureus* endolysin LysH5 with two fusion proteins between lysostaphin and a virion-associated lysin HydH5. When this hybrid protein was tested against *S. aureus* over time, no resistance to the antimicrobial was detected even after ten cycles of bacterial exposure to the phage lysins. The proven efficacy, together with low resistance development, highlights the potential for virion-associated lysins as an alternative therapy to bacterial infection.

Chimeric Lysins.

Chimeric lysins (chimeolysins) are an engineered form of natural phage endolysins often formed by the shuffling of the catalytic N-terminal domains and cell wall binding domains. This rearrangement of the lysin can give rise to enhanced activity, can make the lysin more soluble or can confer a wider host range. As previously mentioned, artificial lysins have been successful in making to commercial use with products like Gladskin receiving large investments and breaking into the US market. The development of chimeolysins gives a new lease of life to endolysins, allowing them to leave the lab and translate to clinical applications.

A recent study looked at the ability of a phage-derived endolysin HY-133 to tackle *S. aureus* biofilm formation in vascular graft infections [32]. These infections result from biofilm colonisation of prosthetics following surgery and are reported to contain bacteria with 1000-fold greater resistance to antibiotics. The endolysin was tested alongside existing antibiotics daptomycin and rifampin. HY-133 is a

recombinant chimeric lysin formed from the N-terminal domain of the endolysin derived from phage K and the cell wall binding domain from lysostaphin. The study concluded that daptomycin has the strongest effect on the biofilm, but HY-133 has a moderate effect on the graft surface, especially after an incubation time of 18 hrs. Rifampicin was not effective as an antimicrobial for this biofilm. As the use of antibiotics for treatment of bacterial infections could pose a risk of resistance, it could be said that the endolysin would be the most favourable anti-microbial agent in this setting.

ClyF is an example of a chimeric lysin that has been engineered specifically to target MRSA. ClyF is derived from the natural endolysin Pc but was modified by the formation of a hydrophobic cleft in its N-terminal and an immunoglobulin-like structure in the cell wall binding domain. Natural *S. aureus* phage lysins have been reported to display issues with solubility and decreased activity that would greatly hinder antimicrobial action, particularly against biofilms. This lysin was chosen from a screening process and underwent several assessments to test its suitability to act against multidrug resistant *S. aureus* in both sessile and planktonic states. ClyF showed an increase in thermostability and a greater range of tolerance to pH. ClyF was used against bacteraemia and wound infection models in mice and showed good efficacy against planktonic and biofilm MRSA [33].

Another research group has proposed a lysin modification system to increase their efficacy. It has been shown that the addition of a cysteine to the end of the C-terminal of antimicrobial peptides such as lysins can increase their antimicrobial activity. The modification template used for this study was called CTC modification strategy (cysteine to its C-terminus). It was reported that the addition of this additional cysteine can improve the efficacy of lysins against both gram-positive and gram-negative pathogens and may be used regardless of the original lysin sequence. Although the exact mechanism of this approach has yet to be defined and results can vary, the increase in antimicrobial action has been shown to be at least two-fold [34].

Acinetobacter baumannii is another pathogen that has been targeted by a new generation of engineered lytic enzymes derived from phages. A 2016 study saw the modification of *A. baumannii* lysin PlyF307 to create the optimised P307SQ-8C [35]. A sequence of amino acids present in PlyF307 was identified as the full cell wall binding domain of the lysin. This is P307, consisting of eight amino acids that

constitutes the full native C-terminal domain for this endolysin. These eight amino acids were fused to a further eight (SQSREQC) to form the final molecule P307SQ-8C. This second sequence was chosen as screening steps from a previous study predicted that the lysin would have a lower level of activity if this sequence was excluded. The resulting endolysin was tested *in vivo* against skin infections by *A. baumannii*. P307 alone had a significant effect against the bacterial infection but it was demonstrated that the modification of the C-terminal domain did in fact increase lytic activity log-fold. It was reported that the engineered lysin P307SQ-8C was 10-15 times more active than its predecessor.

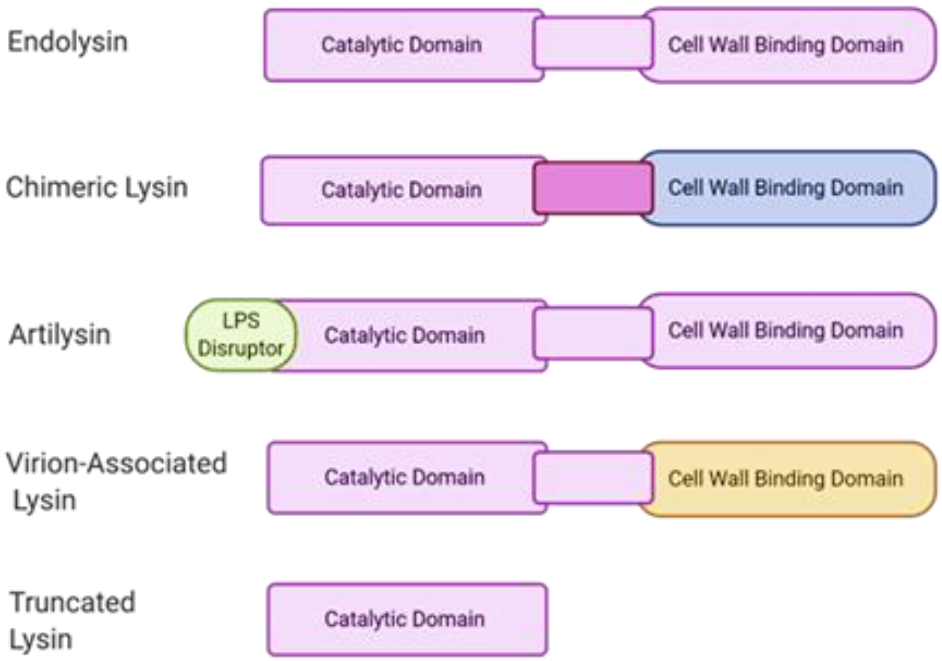


Figure 2. Next Generation Lysins. This illustration shows the basic difference in structure between the following; natural endolysins, the engineered chimeric lysin which involves the linking of the catalytic domain of one lysin to a suitable cell wall binding domain, the Artilysin which involves the fusion of an LPS disrupting molecule to a lysin, VALs which often have dual catalytic active domains and can also be fused to the binding domains of other endolysins, and truncated lysins which have the cell wall binding domain removed to increase activity.

Overcoming Barriers in Drug Delivery Methods.

As previously mentioned, a disadvantage to the use of endolysin therapy is the issue of drug delivery. Endolysins are readily degraded by the body's natural defence mechanisms, and this could lead to poor bioavailability and irreversible damage to the integrity of the protein structure. As we look to the future of endolysin therapy and next generation lysins, it is important to consider ways to overcome this hurdle to make phage-derived proteins a more favourable choice for therapeutics.

Nanoparticles.

Nanotechnology is a relatively new field of science that aims to connect the biological and psychical realms through nanostructures. Nanoparticles are usually small sized nanospheres that are constructed at the molecular level. They are designed to be able to navigate the human body with ease and therefore are excellent candidates as peptide carriers. These spheres can be used to encapsulate a protein-based drug that would otherwise be degraded in its environment. Encapsulation of lysins could be a way forward to allow the endolysin to reach its desired target without being destroyed. It can also allow for a specific dose to be administered at the site of infection. This is essentially a targeted mode of transport for an encapsulated drug.

An example of nanotechnology being adapted to endolysin therapy is the encapsulation of the endolysin LysRODI in pH-sensitive liposomes. This 2020 study highlighted the potential for a lysin to reach its target bacteria with a timed release based on environmental cues [36]. LysRODI is a lysin specific to *S. aureus*. The lysin was encapsulated into liposomes produced by Nanovex Biotechnologies that were designed to be sensitive to pH 5.5 and below. A timed release at this pH could work for human environments such as the skin or in certain food products. A turbidity reduction assay was carried out to confirm that the lysin retained its activity after the encapsulation process. The liposomes containing LysRODI were then tested against biofilms of *S. aureus* strains. The reduction of bacterial counts were not as high as the free form of the lysin, but were still significant against the biofilm formation of all tested *S. aureus* strains.

A 2017 study also encapsulated an endolysin but the release of lysin was thermally triggered. This study targeted skin infections caused by *S. aureus.* The truncated endolysin $CHAP_k$ and bacteriocin lysostaphin were encapsulated in Poly(*N*-

isopropylacrylamide) nanoparticles which were designed to release at a temperature of 37 ºC. This release temperature was chosen as during infection the temperature of the human skin raises from 32 ºC to 37 ºC, and thus the lysin and bacteriocin would only be released in the infected region. Lysostaphin and the endolysin have a synergistic effect against *S. aureus* infection which was not affected by the encapsulation process. The results of this study showed that at 32 ºC *S. aureus* treated with the nanoparticles continued to grow at normal rates while at 37 ºC there was a significant reduction in CFU/ml [37].

Lysocins.

"Lysocins" are a new concept that involves the use of a bacteriocin to deliver a phage-derived lysin across the outer membrane of gram-negative bacteria. This approach has been used to target the ESKAPE pathogen *Pseudomonas aeruginosa*. It has been previously mentioned that endolysins targeting gram-negative species are harder to deploy due to the nature of the bacterial cell. In the parent phage this barrier is overcome by the production of lysins inside the cell and the lysins act to lyse the cell from within. However, lysins applied exogenously are usually unable to reach the inner membrane of gram-negative cells. In 2019 a study used "S-type pyocins" that are naturally produced by gram-negative bacteria such as *Pseudomonas aeruginosa.* These S-type pyocins are bacteriocins encoded by bacteria that favour intraspecies competition in their environments, and so act against other gram-negative species. In this study the domain that allows pyocin S2 to cross the gram-negative outer membrane was utilised. This transporter domain was fused to the *P. aeruginosa* targeting lysin GN4. The resulting "lysocin" was called PyS2-GN4. The study showed PyS2-GN4 to have a narrow spectrum of activity against *P. aeruginosa* with the FpvAI receptor [38].

Current Commercial & Medical Applications of Endolysins.

Although endolysin therapy is a relatively new concept, and the majority of discoveries have not gone beyond the research lab, some companies in both Europe and America have made significant strides in getting their product to the market. Even in its infancy, endolysin therapy has already begun to contribute to medicine and has some promising commercial products coming to the fore. In this review we have discussed

some early endolysin findings, the roadblocks that can be encountered along the path to commercialisation, and what we can expect to see in the future. This section will focus on current applications of endolysins.

Clinical Trials.

At the Rockefeller University, a biotech company called ContraFect has acquired the rights to nine phage-derived lysins. This company's focus is the "molecular treatment of infectious disease", with a strong focus on endolysin therapy. The lysins at this institute specifically target members of the ESKAPE Pathogens including *S. aureus, Streptococcus pneumoniae, Enterococcus faecalis* and *B. anthracis*.

Exebacase (lysin CF-301) is one of ContraFects recombinant lysins that has been designed to target a range of *Streptococcus* and *Staphylococcus* species with an aim to treat infective endocarditis in humans. These species are notorious for forming biofilms on cardiac surfaces and can be difficult to target with traditional antibiotics. This particular lysin was the first of its kind to enter human clinical trials in the U.S. The lysin showed success in Phase II clinical trials, where the recovery rate of MRSA-induced infective endocarditis increased by 42.8% when the lysin was administered alongside the standard antibiotic treatment for this condition [39]. This lysin is expected to enter Phase III human clinical trials in the U.S. ContraFect has reported that upon meeting with the U.S. Food and Drug Administration (FDA) and the end of Phase II it was indicated that the success of Exebacase in Phase III clinical trials could support the Biologics License Application for the commercialisation of this endolysin product. The success of this lysin in clinical trials is so far showing great promise and could pave the way for the approval of endolysin therapy in a clinical and medical setting.

SAL200 is another phage-derived protein that has showed success in clinical trials. A report published in 2017 summarised results of phase I clinical trials in the U.S. In this trial it was reported that SAL200, also called N-Rephasin® SAL200, is suitable to be a drug that is administered intravenously. This endolysin therapeutic is a recombinant form of the lysin SAL-1 targeting antibiotic resistant *Staphylococcus* infections. The 2017 study was focused primarily on the pharmacokinetics of the lysin in the human body, and tolerance for the lysin when administered intravenously. In total, 34 healthy male volunteers completed the study in full. Different doses of SAL200

were administered to different volunteers in a single dose and their condition was monitored. No serious side effects were reported throughout the study, indicating a good safety profile. The endolysin drug also showed good pharmacokinetic characteristics [40]. This was the first study of its kind in the administration of an endolysin-based drug, and the findings indicate that lysin therapy could be a safe and viable therapeutic option, although the authors state further investigations should be carried out.

Commercial Products.

Artilysin® (artificial lysin) is an antimicrobial platform created by biotechnology company Lysando and AiCuris (a spin-off company of Bayer founded in 2006). This platform encompasses the use of muralytic endolysins applied exogenously which can pass through the outer membrane of Gram-negative pathogens. A lipopolysaccharide disruption protein is fused to the catalytic domain or cell wall binding domain of the endolysin without altering its secondary and tertiary structure. This allows the Artilysin to locally penetrate the LPS outer membrane by interfering with its ionic and hydrophobic forces leading to instability of the membrane. The endolysin segment of the moiety may now pass through the outer membrane to target the peptidoglycan layer [41].

Art-175 is an Artilysin developed for the treatment of multi-drug resistant *Pseudomonas aeruginosa*. Art-175 consists of a sheep-derived peptide (SMAP-29) fused to the endolysin KZ144. Due to the success seen with this modified endolysin *in vitro*, the antimicrobial was considered for the treatment of otitis in dog models. Otitis externa results in chronic inflammation of the canine ear canal, often caused by *P. aeruginosa*. Two case studies on two dogs were carried out to test the ability of Artilysins to tackle usually difficult to treat infections. Antibiotics were initially used to treat both dogs, but after 3-6 weeks no signs of improvement were seen. Upon the application of Art-175 in the first subject, the infection and inflammation of the ear was completely healed after six days, with no relapse being reported. These results were mirrored in the second dog who showed recovery from infection eight days after the initial treatment [20].

Micreos is a company whose mission is to provide an alternative to antibiotic treatment. One of their current products on the market is Gladskin, introduced in 2013,

which is a targeted endolysin therapy for *S. aureus* infections that are known to aggravate skin conditions such as eczema and acne. Gladskin is a topical cream containing the endolysin SA.100. A study published in Case Reports in Dermatology reported on the ability of Staphefeckt SA.100 to tackle chronic *S. aureus* related dermatoses [42]. This study followed three individuals, two control groups and one patient receiving the Gladskin treatment. All three subjects had been treated with standard antibiotics and steroids, but the infections persisted with consistent recolonization of the area and resistance to treatment. Treatment with SA.100 resulted in the complete absence of bacterial infection within days, although stopping treatment still resulted in the patient returning to the clinic with recurring symptoms. However, the study concluded that continued use of SA. 100 was capable of continuously suppressing the infection without the occurrence of resistance to the endolysin. It was also found that SA. 100 had a very targeted mechanism of action and did not harm the commensal skin bacteria, unlike the traditional antibiotic solution [42]. Although SA.100 is expressed in recombinant *E. coli*, there is little chance of there being a safety issue during production. *E. coli* is a well-known Gram-negative pathogen, but the strains used for recombinant expression of proteins are usually harmless and are chosen based on their high expression rate, fast growth rate and well-characterised genome [43].

Micreos also have another potential endolysin product in the pipeline in recent years. XZ.700 is a recombinant chimeric lysin that is targeted to drug-resistant species of *Staphylococcus aureus*. The lysin XZ.700 began Phase I/IIa clinical trials in 2020 for the treatment of atopic dermatitis. XZ.700 targets the bacterium that is often responsible for flare ups of this skin condition. This clinical trial will assess the safety of the product on a cohort of 48 patients and is expected to publish results by the end of 2021 [44].

In October of 2020, L'Oréal announced that they will be joining with Micreos to expand their expertise in the skin microbiome and biotechnology field. Under this agreement Micreos will give L'Oréal access to its endolysin research in order to treat skin conditions. This is a significant step forward for endolysin therapy.

Gladskin by Micreos, Exebacase, and the Artilysin platform are certainly success stories when it comes to the commercialisation of endolysin therapy. These

companies have succeeded in creating effective lysins with viable applications and more importantly are continuing their research, advancing clinical trials and seeking future investments. It is worth noting that all of the companies mentioned have other lysins currently under development.

Conclusions.

Endolysin research has certainly grown in interest, particularly over the past ten years. What started out as a decline in phage therapy due to the rise of antibiotics has been revived and is now a growing area of research that has great potential. From a clinical perspective, phage-derived proteins could become a viable treatment option for a number of pathogens, including multidrug resistant bacteria. Not only are lysins proving to be effective antimicrobials alone, but they are also showing synergistic abilities that could give existing antibiotics a new lease of life. Despite the research in favour of endolysins as a therapeutic, they still face challenges to become a commonplace antimicrobial. Initial safety studies to date have indicated that endolysins are suitable for use in medicine but this bank of knowledge is limited, and further research must be carried to ensure phage-derived lysins can be trusted in a clinical setting. Expansion of this research will allow regulatory bodies to make more informed decisions on the commercialisation of endolysin-based products. Nonetheless, the desirable catalytic activity, potential for engineering and targeted mode of action give endolysins a strong advantage over conventional antimicrobials.

Supplementary Materials

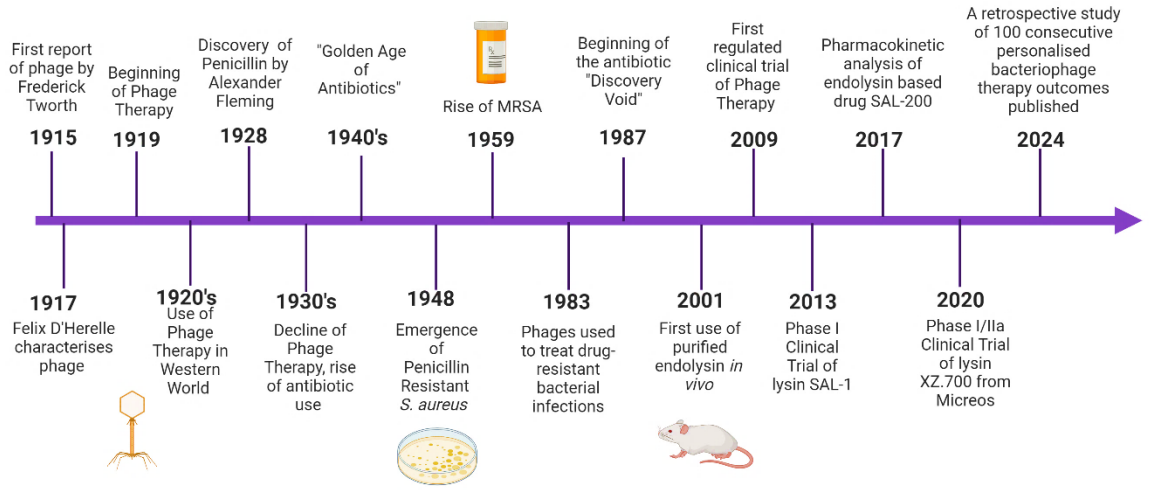


Figure S1. Updated timeline of phage and endolysin therapy.

References

[1] “Antibiotic resistance.” [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance. [Accessed: 26-Feb-2021].

[2] A. L. Clarke, S. De Soir, and J. D. Jones, “antibiotics The Safety and Efficacy of Phage Therapy for Bone and Joint Infections: A Systematic Review,” 2020.

[3] H. H. Kashani, M. Schmelcher, H. Sabzalipoor, E. S. Hosseini, and R. Moniri, “Recombinant Endolysins as Potential Therapeutics against Antibiotic-Resistant Staphylococcus aureus: Current Status of Research and Novel Delivery Strategies,” 2017.

[4] N. Principi, E. Silvestri, and S. Esposito, “Advantages and Limitations of Bacteriophages for the Treatment of Bacterial Infections,” *Front. Pharmacol.*, vol. 10, no. MAY, p. 513, May 2019.

[5] G. P. C Salmond and P. C. Fineran, “A century of the phage: past, present and future,” 2015.

[6] J. Ajuebor, O. McAuliffe, J. O’Mahony, R. P. Ross, C. Hill, and A. Coffey, “Bacteriophage endolysins and their applications,” *Sci. Prog.*, vol. 99, no. 2, pp. 183–199, Jun. 2016.

[7] T. Dörr, P. J. Moynihan, and C. Mayer, “Editorial: Bacterial Cell Wall Structure and Dynamics,” *Front. Microbiol.*, vol. 10, p. 2051, Sep. 2019.

[8] M. Fenton, P. Ross, O. Mcauliffe, J. O’Mahony, and A. Coffey, “Recombinant bacteriophage lysins as antibacterials,” *Bioengineered Bugs*, vol. 1, no. 1. Taylor & Francis, pp. 9–16, 2010.

[9] S. A. Ragland and A. K. Criss, “From bacterial killing to immune modulation: Recent insights into the functions of lysozyme,” *PLoS Pathogens*, vol. 13, no. 9. Public Library of Science, p. e1006512, 01-Sep-2017.

[10] P. Hyman, “Phages for phage therapy: Isolation, characterization, and host range breadth,” *Pharmaceuticals*, vol. 12, no. 1. MDPI AG, 01-Mar-2019.

[11] M. K. Waldor and J. J. Mekalanos, “Lysogenic conversion by a filamentous phage encoding cholera toxin,” *Science (80-.).*, vol. 272, no. 5270, pp. 1910–1913, Jun. 1996.

[12] F. Fan and B. Kan, “REVIEW Survival and proliferation of the lysogenic bacteriophage CTXΦ in Vibrio cholerae,” *Virol. Sin.*, vol. 30, no. 1, pp. 19–25, 2015.

[13] R. Barrangou *et al.*, “CRISPR provides acquired resistance against viruses in prokaryotes,” *Science (80-.).*, vol. 315, no. 5819, pp. 1709–1712, Mar. 2007.

[14] Y. Cha, B. Son, and S. Ryu, “Effective removal of staphylococcal biofilms on various food contact surfaces by Staphylococcus aureus phage endolysin LysCSA13,” 2019. *Food Microbiology*, 84 p. 103245. doi:10.1016/j.fm.2019.103245.

[15] A. Nakonieczna, C. J. Cooper, and R. Gryko, “Bacteriophages and bacteriophage-derived endolysins as potential therapeutics to combat Gram-positive spore forming bacteria,” *J. Appl. Microbiol.*, vol. 119, no. 3, pp. 620–631, Sep. 2015.

[16] B. Son, J. Yun, J. A. Lim, H. Shin, S. Heu, and S. Ryu, “Characterization of LysB4, an endolysin from the Bacillus cereus-infecting bacteriophage B4,” *BMC Microbiol.*, vol. 12, no. 1, p. 33, Mar. 2012.

[17] M. J. Mayer, A. Narbad, and M. J. Gasson, “Molecular characterization of a Clostridium difficile bacteriophage and its cloned biologically active endolysin,” *J. Bacteriol.*, vol. 190, no. 20, pp. 6734–6740, Oct. 2008.

[18] M. L. Van Tassell, L. A. Ibarra-Sánchez, G. P. Hoepker, and M. J. Miller, “Hot topic: Antilisterial activity by endolysin PlyP100 in fresh cheese,” *J. Dairy Sci.*, vol. 100, no. 4, pp. 2482–2487, Apr. 2017.

[19] M. Schmelcher, D. M. Donovan, and M. J. Loessner, “Bacteriophage endolysins as novel antimicrobials,” *Future Microbiology*, vol. 7, no. 10. pp. 1147–1171, Oct-2012.

[20] Y. Briers and R. Lavigne, “Breaking barriers: Expansion of the use of endolysins as novel antibacterials against Gram-negative bacteria,” *Future Microbiology*, vol. 10, no. 3. Future Medicine Ltd., pp. 377–390, 01-Mar-2015.

[21] S. Wang *et al.*, “The antibacterial activity of E. coli bacteriophage lysin lysep3 is enhanced by fusing the Bacillus amyloliquefaciens bacteriophage endolysin binding

domain D8 to the C-terminal region," *J. Microbiol.*, vol. 55, no. 5, pp. 403–408, May 2017.

[22] M. Guo *et al.*, "A Novel Antimicrobial Endolysin, LysPA26, against Pseudomonas aeruginosa," *Front. Microbiol.*, vol. 8, no. FEB, p. 293, Feb. 2017.

[23] Y. Zhang *et al.*, "Antibacterial Effects of Phage Lysin LysGH15 on Planktonic Cells and Biofilms of Diverse Staphylococci," Appl Environ Microbiol. 2018 Jul 17;84(15):e00886-18. doi: 10.1128/AEM.00886-18. PMID: 29776929; PMCID: PMC6052257.

[24] A. Daniel, C. Euler, M. Collin, P. Chahales, K. J. Gorelick, and V. A. Fischetti, "Synergism between a novel chimeric lysin and oxacillin protects against infection by methicillin-resistant Staphylococcus aureus," *Antimicrob. Agents Chemother.*, vol. 54, no. 4, pp. 1603–1612, Apr. 2010.

[25] M. Harhala *et al.*, "Safety Studies of Pneumococcal Endolysins Cpl-1 and Pal," *Viruses*, vol. 10, no. 11, p. 638, Nov. 2018.

[26] B. J. Bruno, G. D. Miller, and C. S. Lim, "Basics and recent advances in peptide and protein drug delivery," *Therapeutic Delivery*, vol. 4, no. 11. NIH Public Access, pp. 1443–1467, Nov-2013.

[27] R. U. Agu, M. I. Ugwoke, M. Armand, R. Kinget, and N. Verbeke, "The lung as a route for systemic delivery of therapeutic proteins and peptides," *Respiratory Research*, vol. 2, no. 4. BioMed Central, pp. 198–209, 12-Apr-2001.

[28] E. Medicines Agency, "The European regulatory system for medicines."

[29] D. Gutiérrez, L. Fernández, A. Rodríguez, and P. García, "Are phage lytic proteins the secret weapon to kill staphylococcus aureus?," *mBio*, vol. 9, no. 1. American Society for Microbiology, 01-Jan-2018.

[30] D. Proença, C. Leandro, M. Garcia, M. Pimentel, and C. São-José, "EC300: a phage-based, bacteriolysin-like protein with enhanced antibacterial activity against Enterococcus faecalis," *Appl. Microbiol. Biotechnol.*, vol. 99, no. 12, pp. 5137–5149, Jun. 2015.

[31] L. Rodríguez-Rubio, B. Martínez, A. Rodríguez, D. M. Donovan, F. Götz, and P. García, "The Phage Lytic Proteins from the Staphylococcus aureus Bacteriophage

vB_SauS-phiIPLA88 Display Multiple Active Catalytic Domains and Do Not Trigger Staphylococcal Resistance," *PLoS One*, vol. 8, no. 5, p. e64671, May 2013.

[32] E. A. Idelevich *et al.*, "Comparative in vitro activity of bacteriophage endolysin HY-133 against Staphylococcus aureus attached to vascular graft surface," *Med. Microbiol. Immunol.*, vol. 209, pp. 51–57, 2020.

[33] H. Yang, H. Zhang, J. Wang, J. Yu, and H. Wei, "A novel chimeric lysin with robust antibacterial activity against planktonic and biofilm methicillin-resistant Staphylococcus aureus OPEN," 2016.

[34] H.-L. Chen, P.-Y. Su, S.-C. Kuo, T.-L. Y. Lauderdale, and C. Shih, "Adding a C-terminal Cysteine (CTC) Can Enhance the Bactericidal Activity of Three Different Antimicrobial Peptides," *Front. Microbiol.*, vol. 9, no. JUN, p. 1440, Jun. 2018.

[35] M. Thandar, R. Lood, B. Y. Winer, D. R. Deutsch, C. W. Euler, and V. A. Fischetti, "Novel Engineered Peptides of a Phage Lysin as Effective Antimicrobials against Multidrug-Resistant Acinetobacter baumannii," 2016.

[36] S. Portilla, L. Fernández, D. Gutiérrez, A. Rodríguez, and P. García, "Encapsulation of the antistaphylococcal endolysin lysrodi in ph-sensitive liposomes," *Antibiotics*, vol. 9, no. 5, May 2020.

[37] H. Hathaway *et al.*, "Thermally triggered release of the bacteriophage endolysin CHAPK and the bacteriocin lysostaphin for the control of methicillin resistant Staphylococcus aureus (MRSA)," *J. Control. Release*, vol. 245, pp. 108–115, Jan. 2017.

[38] R. D. Heselpoth, C. W. Euler, R. Schuch, and V. A. Fischetti, "Lysocins: Bioengineered Antimicrobials That Deliver Lysins across the Outer Membrane of Gram-Negative Bacteria Downloaded from," 2019.

[39] A. Watson, J. T. Oh, K. Sauve, P. A. Bradford, C. Cassino, and R. Schuch, "Antimicrobial activity of exebacase (lysin CF-301) against the most common causes of infective endocarditis," *Antimicrob. Agents Chemother.*, vol. 63, no. 10, Oct. 2019.

[40] S. Y. Jun *et al.*, "Pharmacokinetics and Tolerance of the Phage endolysin-based candidate drug SAL200 after a single intravenous administration among healthy volunteers," *Antimicrob. Agents Chemother.*, vol. 61, no. 6, Jun. 2017.

[41] H. Gerstmans, L. Rodríguez-Rubio, R. Lavigne, and Y. Briers, “From endolysins to Artilysin®s: Novel enzyme-based approaches to kill drug-resistant bacteria,” *Biochem. Soc. Trans.*, vol. 44, no. 1, pp. 123–128, Feb. 2016.

[42] J. E. E. Totté, S. G. M. A. Pasmans, and M. B. Van Doorn, “Case Series Successful Treatment of Chronic Staphylococcus aureus-Related Dermatoses with the Topical Endolysin Staphefekt SA.100: A Report of 3 Cases,” *Case Rep Dermatol*, vol. 9, pp. 19–25, 2017.

[43] I. Permana Maksum, A. Nabiel, S. Ishmayana, and U. M.S. Soedjanaatmadja, “ Safety Aspect of Recombinant Protein Produced by Escherichia coli : Toxin Evaluation with Strain and Genomic Approach ,” in *E. Coli Infections - Importance of Early Diagnosis and Efficient Treatment*, IntechOpen, 2020.

[44] “NTR.” [Online]. Available: https://www.trialregister.nl/trial/8876. [Accessed: 23-Jan-2021].

Chapter 2. Assessing the Therapeutic Potential of an Endolysin Present in Six Novel Phages Targeting *Cutibacterium acnes*.

Ellen Murray, Remi Denise, A.R. Fernandes, Neda Nezam Abadi, Lorraine A. Draper, Paul R. Ross, Colin Hill.

Abstract

Cutibacterium acnes is a gram-positive commensal skin bacterium that has been shown to have pathogenic potential. Certain strains of *C. acnes* have been shown to be associated with the skin condition acne vulgaris. Current treatment options for acne include antibiotic use and harsh topical drugs. These treatment methods often lead to secondary conditions. Depending on the treatment used, these secondary effects can be both physical and psychological, including eczema, skin irritation and depression. Acne is estimated to affect 9.4% of the world's population, and so it is of great importance to find alternative therapeutics for its management. In this study, we isolated six novel phages infecting *C. acnes* (phages 005A1, 005A2, 42-1, 42-2, 42-3, and 42-4). These phages are all of the genus *Pahexavirus* and have high homology to each other but have differing host ranges. The endolysin genes of these phages were identified for biotechnological applications. The six endolysins genes had 100% nucleotide identity and so one candidate was chosen for assessment, LysCa5A2 derived from phage 005A2. However, the expression and purification of LysCa5A2 was compromised by issues surrounding insolubility. Here, we describe the expression of LysCa5A2 in inclusion bodies and the attempt to recover a native protein.

Introduction

Cutibacterium acnes (formerly *Propionibacterium acnes* [1]) is a gram-positive, facultative anaerobe and a commensal of human skin. It is an opportunistic pathogen commonly known for its association with the skin condition acne vulgaris [2]. *C. acnes* has also been linked with implant-related infections, including neurological shunts [3], intra-ocular lenses [4], cardiac devices [5], prosthetic shoulder joints [6] and non-hardware-associated vertebral osteomyelitis [7].

The skin is a diverse and rich source of microbes. These microbes not only exist on the surface of the skin but also populate the deeper layers of the epidermis, such as the dermis, follicles, and sebaceous glands [8]. Acne vulgaris is a chronic inflammatory condition in the skin's pilosebaceous unit. This is a region of the dermis that comprises a hair follicle, hair shaft and its associated sebaceous gland [3]. Acne lesions form when keratinocytes lining the hair follicle fall away and cause a clog in the pore, creating a microcomedone. An increase in sebum production can facilitate the colonisation of these lesions with *C. acnes*. As the bacteria proliferate, inflammatory mediators are produced, which drive inflammation of the lesion [9]. This condition is estimated to affect approximately 9.4% of the world’s population, making it the eighth most prevalent disease worldwide [10]. From the ages of 15 – 17, nearly all individuals will experience acne to some degree, but up to 20% of these cases are moderate to severe [11]. This is a multifaceted chronic skin condition. However, increased sebum production, microbial composition, abnormal cell maturation, and migration are the most common influencing factors. Individuals with acne vulgaris do not have higher levels of *C. acnes* than the average person. Instead, the strain of *C. acnes* colonising the skin causes acne. Some strains are commensal and contribute to skin health, whereas others carry genetic features that have pathogenic potential [12]. Currently, there are eight phylotypes of *C. acnes*. IA1 and IA2 phylotypes are predominantly found in acne lesions. Type IB is associated with healthy skin [13]. Type IA1 *C. acnes* is more virulent and has been shown to result in a higher production of β-defensin 2 from cultured keratinocytes and higher lipase activity levels than other phylotypes associated with healthy skin [14].

Current treatments for acne vulgaris are harsh and often come with undesirable side effects. *C. acnes* infections are regularly treated with topical and oral broad-spectrum antibiotics. Such treatments can disrupt the skin microbiome, altering the composition of the skin's commensal bacteria. Using antibiotics for these infections comes with the risk of the emergence of antibiotic-resistant strains of *C. acnes*. Resistance to several antibiotics has already been reported, including erythromycin, clindamycin, and doxycycline [15]. Other drugs used to manage the condition include the direct application of benzoyl peroxide, salicylic acid, azelaic acid, and topical retinoids. These drugs are often effective in reducing the appearance of acne but have a range of adverse effects associated with their use. Concentrations of 2.5% to 10.0% benzoyl peroxide can be prescribed. However, higher concentrations can lead to redness and burning and can even bleach clothing, bedding and hair that comes in contact with the treated skin [16]. Topical retinoids are a popular treatment option. An example is tretinoin, which increases the turnover of follicular epithelial cells and speeds up the shedding of corneocytes (cells on the outermost layer of the epidermis). It is reported that follicles become less anaerobic due to this process, causing a reduction in the proliferation and function of *C. acnes* [17].

Combining the treatment options listed above is also an alternative for patients dealing with the condition. However, the ability of *C. acnes* to form biofilms can enable the bacteria to resist even these combined antimicrobial agents. [18]. Long-term acne treatments using low doses of antimicrobials and antibiotics have contributed to increased resistance among *C. acnes* strains. This is illustrated by a significant increase in resistant strains, from 34.5% in 1991 to 55.5% in 2000 [19].

Treatment compliance for acne vulgaris is poor due to discontinuation of treatment when early improvement is seen, a perception of worsening acne and the onset of side effects [20]. Many patients experience the onset of eczema and other secondary skin conditions when using topical treatments. While high concentrations of retinoids may effectively eliminate bacterial infections, they can also compromise the skin barrier, leading to flaking, discomfort, and pain for the patient. Oral medications like Accutane, commonly prescribed for acne management, have been linked to adverse effects on mental health [21]. Despite the skincare industry's significant financial power, current treatment options have many unintended consequences, both physical and mental. A new generation of therapy is clearly

needed to revolutionise traditional treatment approaches and enhance patient outcomes.

The use of bacteriophages (phages) and their endolysins (lysins) has gained much traction in recent years as an alternative to antibiotic treatment regimes. Phages and their lysins allow us to specifically target individual bacteria and edit microbiome composition. Phage therapy has clear therapeutic potential, but biological, production and regulatory challenges limit its application in mainstream medicine [22]. Endolysins could provide a solution to overcome these barriers. Endolysins (lysins) are phage-derived lytic proteins classed as peptidoglycan hydrolases. Depending on their mechanism of action, they target specific bonds within the peptidoglycan layer of the bacterial cell wall. In nature, they come into play at the end of the phage replication cycle. Endolysins are paired with another protein called holin. Holin is a small protein that accumulates in the bacterial cell during the phage infection. When they reach a specific level, the holin creates pores in the cell membrane, allowing endolysins to target the peptidoglycan. This degradation of the peptidoglycan disrupts the bacterial membrane potential, resulting in lysis [23]. For biotechnological applications, endolysins can be applied exogenously (particularly in the case of gram-positive targets), and this lytic activity can be replicated from outside the cell.

To date, three endolysins from *C. acnes* phages have been characterised in the literature. Endolysin CAP 10-3 is derived from phage CAP 10-3. This endolysin was shown to target *C. acnes* KCTC 3314 in a dose-dependent manner [24]. *C. acnes* endolysin *Pa*Ami1 is derived from *Cutibacterium* phage PAC1 [25]. *Pa*Ami1 was shown to have lytic activity against *C. acnes* peptidoglycan. Two variants of *Pa*Ami1 were also generated, which involved the fusion of the N-terminus to two antimicrobial peptides. These variants had improved lytic activity against the target strain and also displayed activity against *Enterococcus faecalis* and *Enterococcus faecium*. LysP11 is derived from *Cutibacterium* phage P1.1. LysP11 is especially interesting as the protein is produced in the plant *Nicotiana bethamiana* using an *Agrobacterium*-mediated transient expression strategy. LysP11 did not have lytic activity against *C. acnes* but was active against *Erysipelothrix rhusiopathiae*. While not associated with acne *E. rhusiopathiae* also causes inflammation of lesions in humans. *E. rhusiopathiae* is a pathogenic bacterium that typically causes infections due to a wound. This infection becomes a localised cutaneous disease characterised by inflamed skin

lesions. All currently characterised lysins have an N-acetylmuramoyl-L-alanine amidase mode of action.

In this study, we present the isolation of six novel phages infecting *C. acnes* (005A1, 005A2, 42-1, 42-2, 42-3, and 42-4). These phages are taxonomically placed in the genus *Pahexavirus*. The newly isolated phages had a close homology at the nucleotide and amino acid levels with a genome size of 29 – 30 kb. They presented differing host ranges against *C. acnes* strains. The endolysin gene in each phage was identified. These six endolysins were found to have 100% similarity on an amino-acid level. The endolysin from phage 005A2, LysCa5A2, was assessed for its potential as a therapeutic for acne vulgaris. LysCa5A2 was cloned and recombinantly expressed in *E. coli*. Downstream analysis revealed that the production of LysCa5A2 was produced as an insoluble protein. No lytic activity was observed under its current production conditions.

Results

Host Range.

Six phages infecting *C. acnes* were isolated from a screen of samples ranging from skin swabs, farm samples and sewage samples. *C. acnes* DSM 1897 was the strain used as the host bacterium. In a host range analysis, the ability of the phages to infect other *C. acnes* strains was assessed (**Table 1**). There are type strains procured from DSMZ, Germany, representing type I and type II *C. acnes* strains –namely DSMZ 1897, 16379, 30738, 30753 and 30919. All isolated phages were able to infect *C. acnes* DSM 1897 and *C. acnes* DSM 30738. All phages, except for phage 42-3, were shown to infect *C. acnes* DSM 16379. Only phages 005A1 and 42-4 could infect *C. acnes* DMS 30919.

Phage Visualisation and Morphology.

TEM morphological analysis was carried out on phage 005A1 (**Figure 1(A)**) and phage 42-1 (**Figure 1(B)**). Phage 005A1 had a head-to-tail length of 199.2 nm. The non-contractile tail was 138 nm in length and 9.2 nm in diameter. The head of phage 005A1 was icosahedral in shape and had a diameter of 48.6 nm. Phage 42-1 had a head-to-tail length of 213 nm. The tail measured 149 nm in length and 9.9 nm in diameter. The diameter of the phage head measured 54 nm, and it was also icosahedral in shape. The tails of both phages appear to be flexible. These are typical features of the *Siphoviridae* family.

C. acnes phages 005A1, 005A2, 42-1, 42-2, 42-3, and 42-4 gave rise to a mixed plaque morphology. The plaques were irregularly shaped and generally of two sizes – a larger plaque with a halo and a smaller plaque with a halo. To ensure that the morphology presented by these phages was not due to a contaminating phage, they were separated and propagated based on their size. This was carried out by a “phage T-streak” method in which a phage lysate was T-streaked on solid agar and overlaid using the double-layer method. From this, plaques were isolated based on their size and the technique was repeated. Plaque assays were performed using individual small or large plaque lysates. Multiple rounds of this separation and propagation showed that the mixed morphology arose from a pure phage lysate (**Figure 1 (C)**). The images

shown in **Figure 1 (C)** were prepared with phage 005A2 against *C. acnes* DSM 1897 and were photographed after two days of growth under anaerobic conditions.

Genome Analysis.

Circular genomes were assembled for each of the newly isolated phages. During the assembly process, a number of troubleshooting steps were employed. It was evident that contaminating reads were present in the sequencing data. To minimise errors, the reads were decontaminated from DNA of the following sources: human, phage phiX, *E. coli*, host *C. acnes*, and *C. acnes* plasmids. Following these steps, six circular genomes were created and annotated for further analysis. It is possible that some errors are still present in these resulting genomes, and to confirm the following findings, the DNA extractions and sequencing of these phages should be repeated.

The genomes of *C. acnes* phages isolated in this study ranged from 29 – 30 kbp in size. Genomes were annotated using Pharokka, phold, and phynteny to increase annotation levels. These annotation tools are based on different databases. The use of all three provided a wide range of annotations based on protein family databases (Pharokka), structural homology (phold) and synteny amongst known viruses (phynteny) [26] [27] [28].

VIRIDIC was used to calculate the intergenomic similarities between the isolated phages (**Figure 2**) [29]. The output of this tool is based on the percentage identity between two genomes determined by BLASTN. VIRIDIC revealed that the *C. acnes* phages had high intergenomic similarity on a nucleotide level, ranging from 90% to 100%. Phages 42-1, 42-2, and 42-3 had 100% genetic similarity, suggesting that they are the same phage. This conclusion is not supported by their host ranges, as they have a differing infection profile. It is possible that as phages 42-1 and 42-2 have the same host range, they are the same phage. However, 42-3 has a different host range. The similarity seen in VIRIDIC analysis could be due to errors in sequencing and assembly. Phage 005A2 was the most dissimilar to the other phages, with the lowest homology to phage 42-4.

The major functional proteins of this genome could be divided into four groups based on different functions: DNA, RNA and nucleotide metabolism, head and packaging, lysis, tail morphogenesis and transcriptional regulation. Sequencing reads were assembled into circular genomes. All genomes have a GC content of 54% -

54.2%. Genome maps for phages 005A2 and 42-4 were created with Proksee [30]. These two phages were chosen as representatives as they were the most dissimilar to each other (90%). The annotation was inferred from phold analysis (**Figure 3 (A)** and **(B)**). Phage 005A2 is 27,743 bp in size, and 42-4 is 27,644 bp. The right-hand side of both genomes (regions 7 – 13 kbp) encoded for genes with functions associated with phage structure. The left-hand side of the genomes (region 14 – 23 kbp) is dominated by genes associated with DNA, RNA, and metabolism. Both phages encoded an endolysin and a holin gene at ~ 13 kbp.

Whole genome comparisons were performed with the *C. acnes* phage genomes obtained in this study against members of the different genus clusters within the candidate *Pahexavirus* family identified by the taxmyPHAGE tool [31]. This tool was developed to assign taxonomy to genus and species-level phages based on the current classified International Committee on Taxonomy of Viruses (ICTV) genomes. From here, the tool indicates if your genome represents a new species. TaxmyPHAGE analysis on the six *C. acnes* phages revealed they all belonged to the class *Caudoviricetes*, genus *Pahexavirus*, and represented a new species of this genus. TaxmyPHAGE grouped the phage with 57 other phages that would comprise this new species. The genus *Pahexavirus* was established by the ICTV in 2016. The placement of these *C. acnes* phages in the genus *Pahexavirus* is consistent with other studies that genomically characterised *C. acnes* phages [32].

A phylogenetic tree was prepared to assess the phylogenetic relationships between our six *C. acnes* phages and the other 57 from the database. This was created using the FoldTree snakemake workflow. FoldTree allows for the visualisation of trees from protein structures [33]. The annotations for this tree were generated using Colabfold [34]. Colabfold provides complex protein folding based on the AlphaFold model. Once the protein structures were determined, they were compared against each other using DIAMOND (double index alignment of next-generation sequencing data), an alignment tool [35]. The analysis was performed against 57 genomes from TaxmyPHAGE. These genomes were normalised, so each genome in the figure starts with the large terminase (TerL). The resulting tree (**Figure 4**) shows genomes with Colabfold annotations highlighted in the legend and percentage identity between genomes. The overall genome organisation and genetic synteny were found to be largely conserved across all the clusters except for three genomes at the bottom of

the tree (MH479925, MH479913 and MK967380). These three genomes are considerably larger (~ 60 – 70 kbp) and show a low percentage identity to each other and the other genomes of this species. Regarding the genomes isolated during this study, phages 005A2, 42-1, 42-2, 42-3, and 42-4 are grouped, but 005A2 is more similar to other species members. The majority of our phage genomes show a high percentage identity to their relatives, with some variations at the beginning and ends of the genomes. This observation may be due to issues in the assembly of the phage genomes. The genomes of *C. acnes* phages are highly conserved. The endolysins of *C. acnes* phage genomes are of particular interest to this study, and it can be seen from the phylogenetic analysis that these genes (highlighted in red) are highly conserved at a structural level (>80% percentage identity).

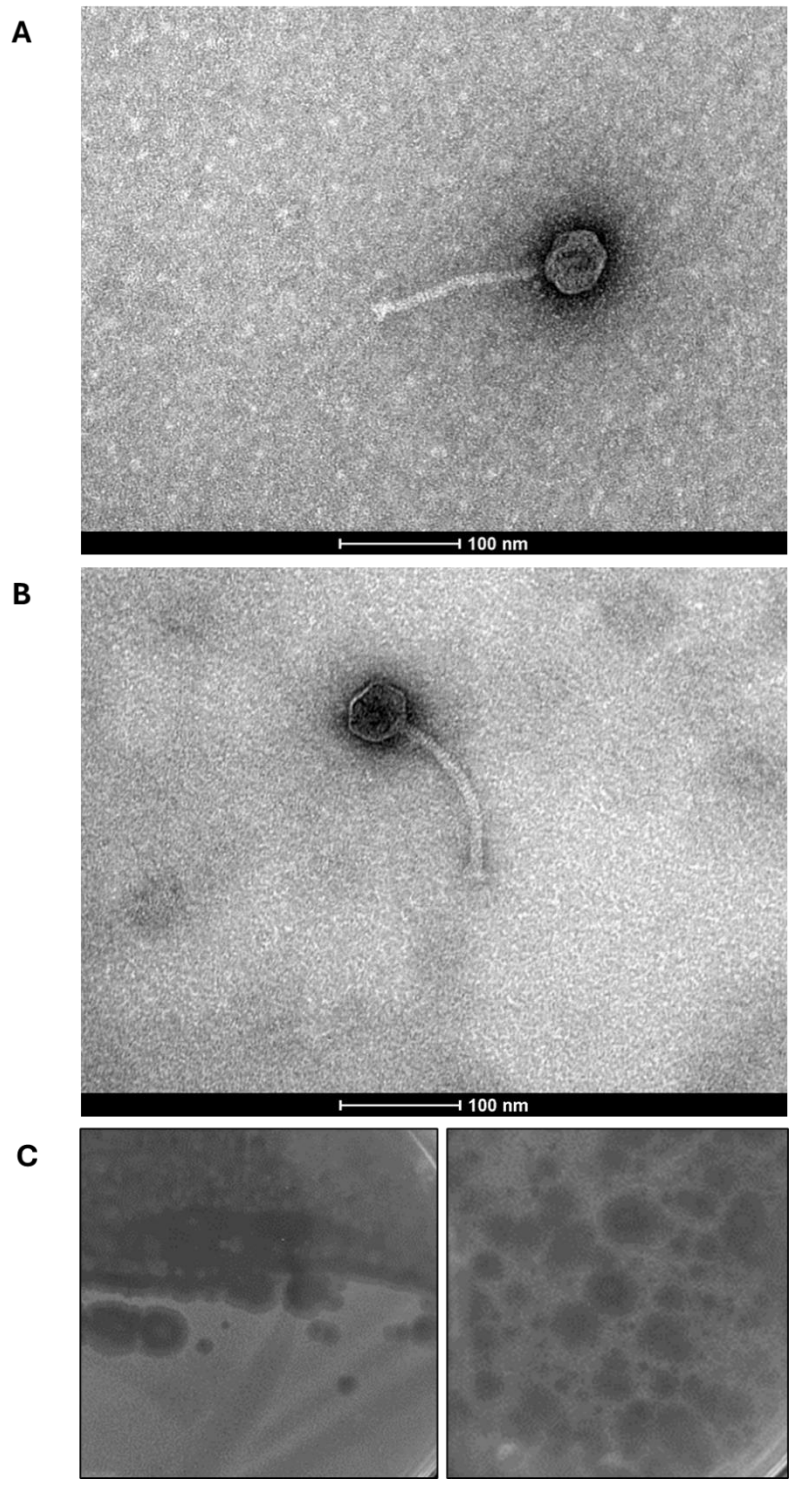


Figure 1. Visual and morphological analysis of *C. acnes* phages and plaques. A) Transmission Electron Micrograph (TEM) of Phage 005A1, scale=100 nm. **B)** TEM of phage 42-1, scale=100 nm. **C)** Phage morphology assessed by the phage T-streak method and traditional plaque assay, respectively. These techniques were carried out with phage 005A2 applied to a lawn of *C. acnes* DSM 1897.

Table 1. Host range of *C. acnes* phages. Phage infection against *C. acnes* phages.

Species	Strain	φ 005A1	φ 005A2	φ 42-1	φ 42-2	φ 42-3	φ 42-4
C. acnes	1897	+	+	+	+	+	+
	16379	+	+	+	+	-	+
	30919	+	-	-	-	-	+
	30738	+	+	+	+	+	+

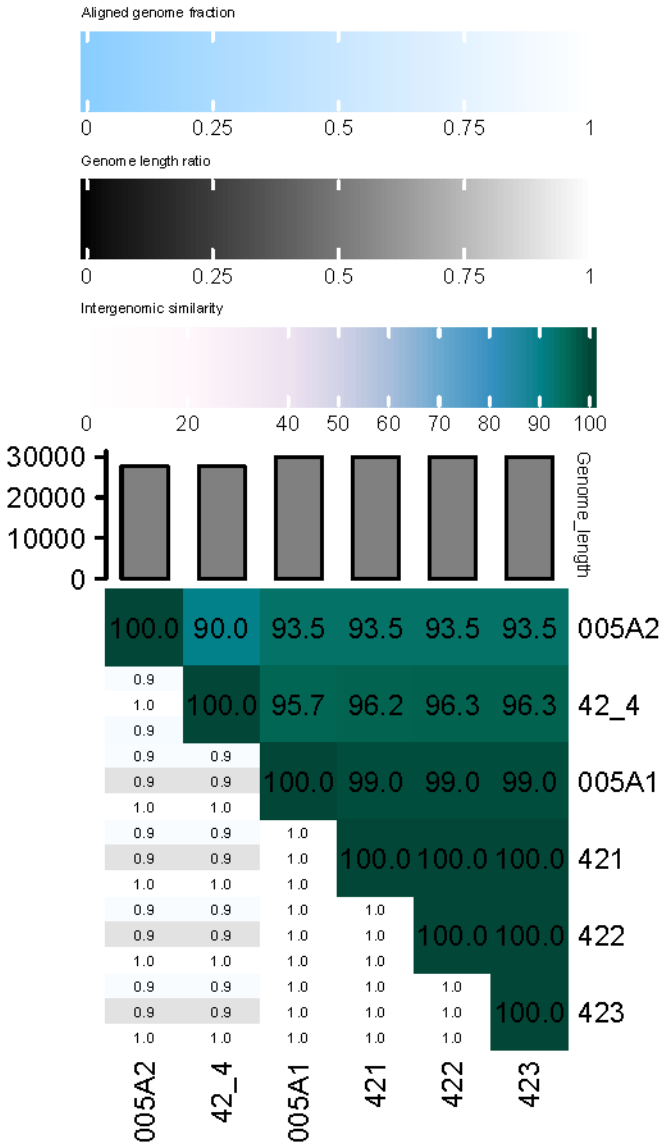


Figure 2. VIRIDIC analysis of *C. acnes* phage genomes. VIRIDIC was used to calculate the intergenomic similarities between *C. acnes* phages 005A1, 005A2, 42-1, 42-2, 42-3, and 42-4. VIRIDIC analysis was carried out on a nucleotide level.

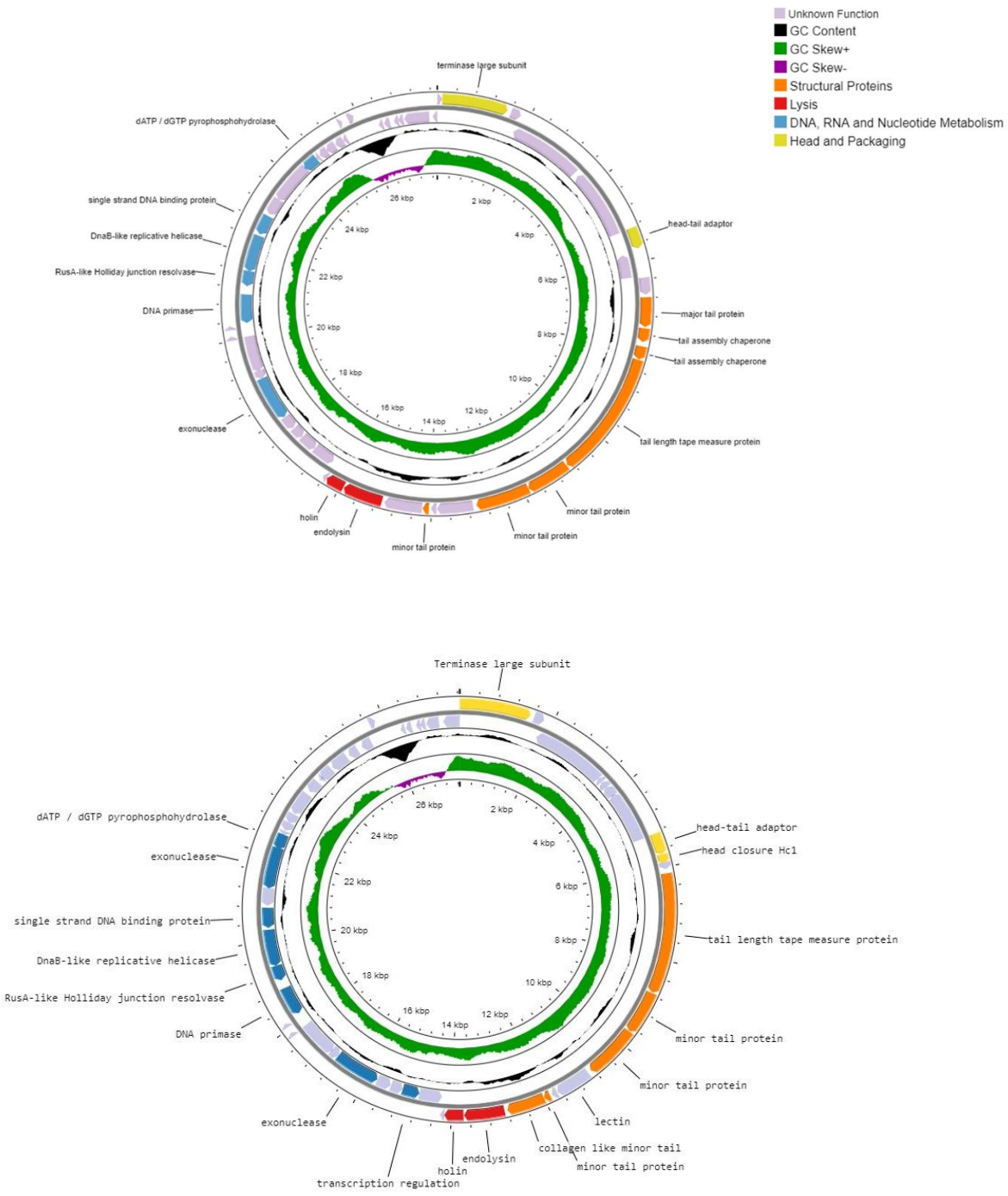


Figure 3. Genome Map of *C. acnes* phages A) 005A2 and B) 42-4. The innermost ring represents the GC skew (green for the positive strand and purple for the negative strand), while the middle ring (black) displays the GC content. The outermost circle illustrates the coding genes (CDS) with phold's predictions of functions as labels. The colouration of the CDS corresponds to their general functions, as indicated in the legend. Genes whose function could not be determined remain unlabelled.

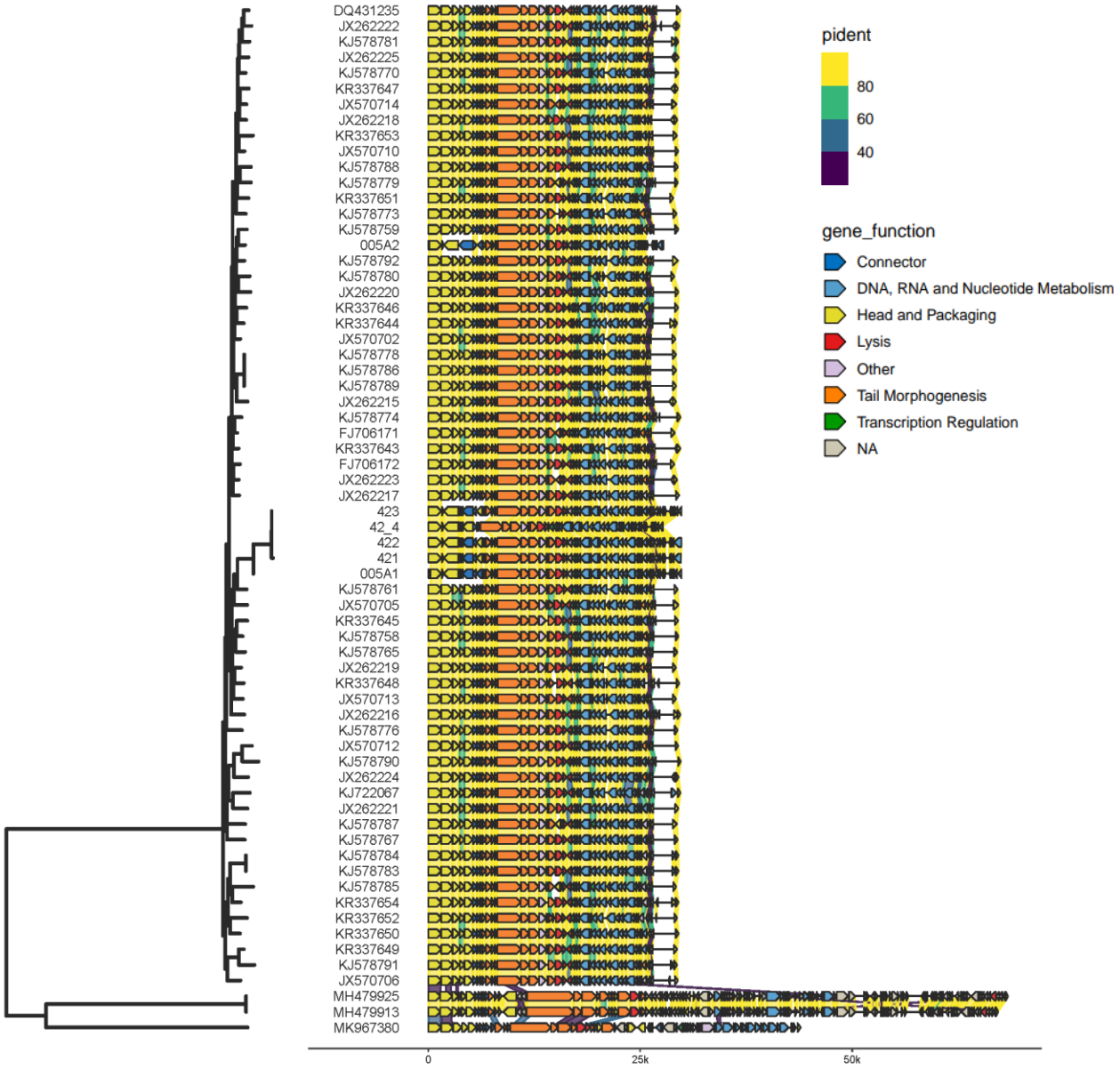


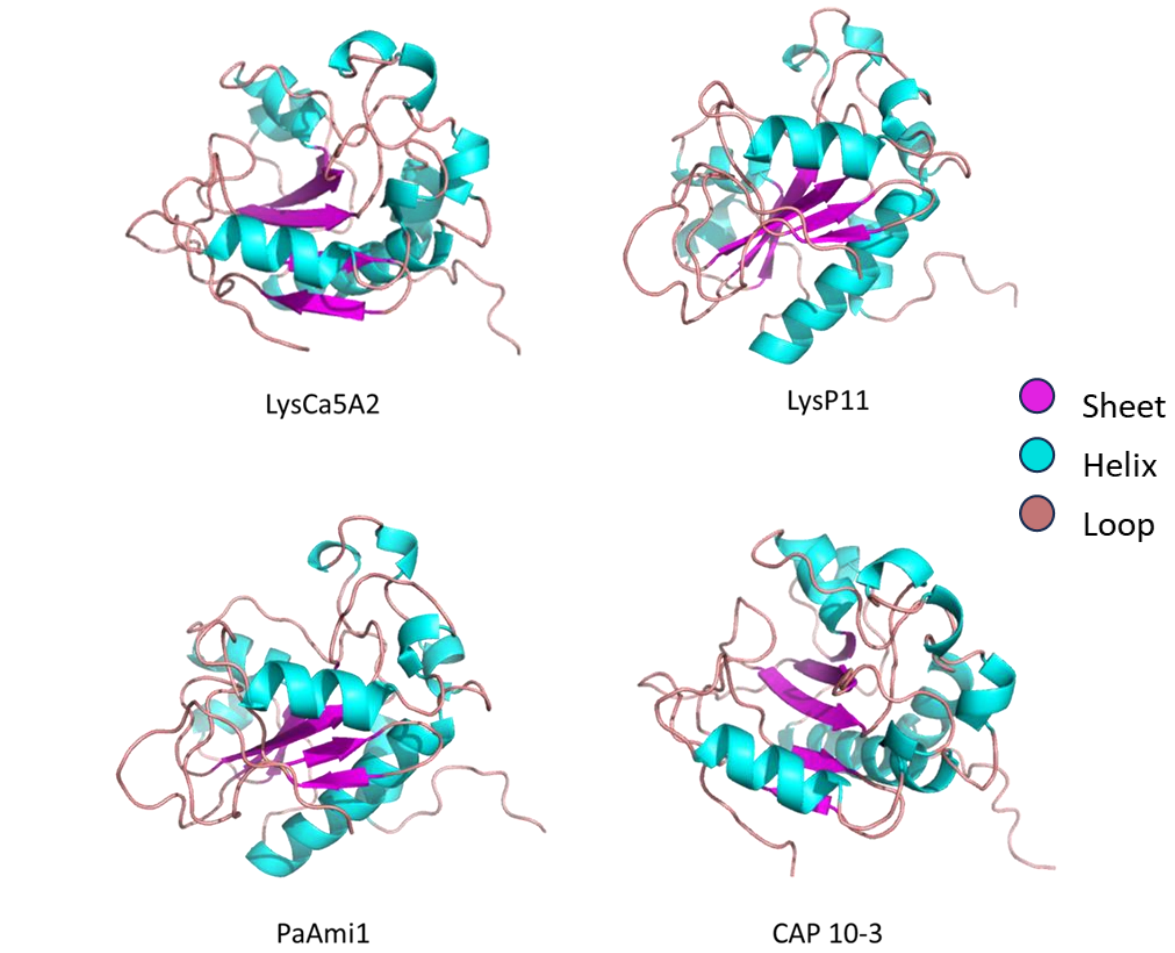
Figure 4. Whole genome comparisons of *C. acnes* phages isolated in this study against members of the different genus clusters within the candidate *Pahexavirus* family identified by TaxMyPhage. The figure shows the percentage of identity of different proteins compared to the closest phylogenetic phage genome. The comparison was performed using DIAMOND. The colour between two genes represents the percentage of identity between both at the protein level. The colour of the gene is based on the annotation given by Colabfold. The figure was created using the R packages gggenomes and ggtree.

Endolysins of C. acnes Phages.

The endolysins of the six newly isolated phages were identified with BLASTp. Using Clustal Omega, the six endolysins were compared on an amino acid level, and the percentage identity revealed they have 100% identity. Endolysin from phage 005A2 was investigated for biotechnological applications. This endolysin is 284 amino acids in size. Bioinformatic tools were used to assign functions to the domains of LysCa5A2. InterproScan analysis predicted a two-domain structure [36]. The N-terminal domain was predicted to be the catalytic domain that was assigned an N-acteylmuramoyl-L-alanine amidase mode of action. The catalytic domain was annotated to span amino acids 1 – 180. InterproScan analysis did not report any hits from the database for the C-terminal domain of LysCa5A2. HHpred analysis mirrored these results, with the top hit (99.63% probability) being N-acteylmuramoyl-L-alanine amidase. Region 247 – 284 of the endolysin (the end of the C-terminal domain) had homology to a peptidoglycan recognition protein (E-value 0.00012) of Drosophila melanogaster. To further assign functions to the domains, Foldseek was employed [37]. This server assesses protein homology based on protein structure. The top hit for this search also annotated the N-terminal domain to have an amidase mode of action but also reported that it has structural homology to the autolysin of *Staphylococcus aureus*. As for the C-terminal domain, the top homology was to an uncharacterised protein. Lower-confidence hits included “receptor protein serine/threonine kinase” and a range of other proteins of unknown function. Based on this analysis, we conclude that the N-terminal of LyCa5A2 is a catalytic domain possessing an amidase mode of action. We propose that the C-terminal of LysCa5A2 is likely to be a cell wall binding domain.

Comparative analysis was carried out on the *C. acnes* endolysin from LysCa5A2 and the other characterised *C. acnes* endolysins from the literature – *Pa*Ami1, CAP10-3 and LysP11. Structural analysis was performed on each endolysin using the AlphaFold3 server [38]. The active domains of the endolysins were visualised using PyMOL (**Figure 5 (A)**). The active domains are shown alone, as AlphaFold failed to assign structures of high confidence to the C-terminal domains of the *C. acnes* endolysins. The N-terminal domain of LysCa5A2 displays two anti-parallel β-sheets with a single β-sheet between. These sheets are surrounded by five α-helices. The structure of *C. acnes* endolysins appears to be well conserved, as the other *three C. acnes* lysins displayed the same structure. To assess the amino acid

composition of the four endolysins, multiple sequence alignments and percentage identity matrices were carried out using Clustal Omega [39]. The multiple sequence alignments were visualised with SeaView4 (**Figure 5 (B)**). The percentage identity matrix was visualised with Heatmap2 (**supplementary figure S1**). The multiple sequence alignments show that the composition of the endolysins is quite conserved, with two regions showing distinct deviations (position 1 and position 193). At position 1, *Pa*Ami1 has a methionine start codon. The other lysins start with an alternative start codon, valine. At position 193, LysCa5A2 and LysP11 lack a serine. The percentage identity matrix shows that the endolysins are all >94% similar on an amino acid level. LysCa5A2 has the highest homology to LysP11 (94.72%) and the lowest homology to lysin CAP 10-3 (93.66%).




```
                1
WIW76974.1      MVRFIPAAHH SAGSNSPVNR VVIHATCPDV GFPSASRKGR AVSTANYFAS PSSGGSAHYV  CDIGETVQC LSESTIGWHA
LysCa2          -VRYIPAAHH SAGSNSPVNR VVIHATCPDV GFPSASRKGR AVSTANYFAS PSAGGSAHYV  CDVGETVQC LSESTIGWHA
YP_006906607.1  -MRFIPAAHH SAGSNSPVNR VVIHATCPDV GFPSASRKGR AVSTANYFAS PSSGGSAHYV  CDISETVQC LSESTIGWHA
YP_009214894.1  -MRFIPAAHH SAGSNSPVNR VVIHATCPDV GFPSASRKGR AVSTANYFAS PSSGGSAHYV  CDVGETVQC LSEGTIGWHA

               81
WIW76974.1      PPNPHSLGIE ICADGGSHAS FRVPGHAYTR EQWLSPQVWP A VEKAAVLC RQLCDKHGVP KRKLSVADLK AGKRGVCGHV
LysCa2          PPNPHSLGIE ICADGGSHAS FRVPGHAYTR EQWLDPRVWP A VERAAVLC RRLCDKYNVP KRKLSVADLK AGRRGICGHV
YP_006906607.1  PPNPHSLGIE ICADGGSHAS FRVPGHAYTR EQWLDPRVWP A VEKAAILC RRLCDKYNVP KRKLSAADLK AGRRGVCGHV
YP_009214894.1  PPNPHSLGIE ICADGGSHAS FRVPGHAYTR EQWLDPRVWP A VEKAAILC RRLCDKYNVP KRKLSAADLK AGRRGVCGHV

              161
WIW76974.1      DVTDAWHQSD HDDPGPWFPW DK FMAVVCG GSSDSGELTV ADVKALHDQI KQLSAQLTGS VNKLHHDVGV VQVQNGDLGK
LysCa2          DVTDAWHQSD HDDPGPWFPW DK FMAVVCG GSGS-EELTV ADVKALHDQI KQLSAQLTGS VNRLHHDVGV VQVQNGDLGK
YP_006906607.1  DVTDAWHQSD HDDPGPWFPW DR FMAVVNG GSGS-EELTV ADVKALHDQI KQLSAQLTGS VNKLHHDVGV VQVQNGDLGK
YP_009214894.1  DVTDAWHQSD HDDPGPWFPW DK FMAVVCG GSGESGELTV ADVKALHDQI KQLSAQLSGS VNKLHHDVGV VQVQNGDLGK

              241
WIW76974.1      RVD ALSWVK NPVTGKLWRT KDALWSVWYY VLECRSRIDR LESTVNGLKK
LysCa2          RVD ALSWVK NPVTGKLWRT KDALWSVWYY VLECRSRIDR LESAVNDVKK
YP_006906607.1  RVD ALSWVK NPVTGKLWRT KDALWSVWYY VLECRSRIDR LESTVNGLKK
YP_009214894.1  RVE ALSWVK NPVTGKLWRT KDALWSIWYY VLECRSRIDR LESAVNGLKK
```

Figure 5. Comparative Analysis of *C. acnes* Endolysins. A) Structural analysis of *C. acnes* endolysin active domains, secondary structure highlighted. **B)** Multiple sequence alignment of *C. acnes* endolysins. LysP11: YP_006906607.1, CAP 10-3: WIW76974.1, *Pa*Ami1: YP_009214894.1.

Cloning and Recombinant Expression of LysCa5A2.

Amplification of the endolysin gene resulted in PCR products of 877 bp. Amplicons were cloned into the pET28b(+) expression vector and were subsequently introduced into *E. coli* BL21(DE3) for recombinant expression. The construct was designed with an N-terminal 6xHis-tag for protein purification. The predicted size of the protein and 6xHis-tags is ~ 33 kDa. The isoelectric point is calculated to be 9.620. The solubility of LysCa5A2 was assessed with an in-silico tool ProteinSol [40]. This tool compares the query protein's solubility score to the database's average solubility (0.45). LysCa5A2 was found to be less soluble than the average protein, scoring 0.35 (**Supplementary Figure S4**).

E. coli BL21(DE3) cells containing the vector were induced with Isopropyl ß-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.1 mM. To assess overexpression, crude *E. coli* was harvested before and after induction. Samples were also taken of the soluble and insoluble fractions after the cells were lysed for purification. These samples were assessed by SDS-PAGE (**Figure 6 (B)**). In the SDS-PAGE analysis, the expected band of overexpression (~ 33 kDa) was visible (LysCa5A2 with 6xHis-tag) in lane 2 (crude *E. coli* post-induction), lane 3 (soluble fraction) and lane 4 (insoluble fraction). It was observed that the vast majority of LysCa5A2 was present in the insoluble fraction of cell lysate. This suggests that during expression, the protein may have formed inclusion bodies. In an effort to increase yield from the soluble fraction, a 3-litre protein expression was established. The soluble fraction of this preparation was harvested and purified. Purification of the soluble fraction was carried out by affinity chromatography through the ÄKTA start protein purification system (Cytiva) with the Cytiva Talon Crude purification columns. The purity of the yield was assessed by SDS-PAGE (4-12% gradient gel). The resulting chromatogram and SDS-PAGE from the ÄKTA start Unicorn software can be seen in **Supplementary Figure S2**. While a peak was observed in the chromatogram, SDS-PAGE analysis indicated that the His-tagged protein was not recovered during purification. The fractions of the purification were concentrated using size-exclusion spin columns (10K MWCO), and the yield was assessed for activity by spot assay against *C. acnes* DSM 1897. No lysis was observed.

Challenges of insoluble proteins and assessment of lytic activity.

Due to the unsuccessful attempt to purify bioactive protein from the soluble fraction of cell lysate, efforts were made to recover active protein from the insoluble fraction. The insoluble pelleted was resuspended in binding buffer adjusted to a pH of 6 (50 mM sodium phosphate, 150 mM NaCl). This suspension was assessed for activity by traditional spot assay, but no lysis was observed.

Lytic activity was further assessed using zymogram analysis. This is a technique used to detect enzyme activity in a polyacrylamide gel. In a zymogram, proteins are run under denaturing conditions (SDS) against autoclaved cells (in this case, *C. acnes* DSM 1897). After electrophoresis, the zymogram gel is treated to remove SDS and allow the proteins to renature and regain their activity. The SDS-PAGE (**Figure 7 (A))** and zymogram **(Figure 7 (B)**) were run with lanes containing a ladder, crude *E. coli* before induction with IPTG, after induction with IPTG, the soluble cell lysate and the insoluble cell lysate, respectively. The zymogram gel was refolded in 25 mM Tris-HCl pH 7.5, with the addition of 0.3% Triton-X 100. The zymogram was refolded at 37 ^{o}C for 72 hours. The gel was subsequently stained to assess for any zones of lysis. No lytic activity was observed.

We assumed that LysCa5A2 is inactive when it is in inclusion bodies. N-lauroylsarcosine (Sarkosyl) (Thermo Fischer) was used as a solubilising agent to try to recover the protein from inclusion bodies. A solubilising buffer was prepared containing 50 mM sodium phosphate, pH 6, and 1% w/v sarkosyl. An insoluble pellet from the purification was harvested by centrifugation, washed, and resuspended in the solubilisation buffer. The protein was left to in this buffer, shaking, for 2 hours. The suspension was centrifuged, and the supernatant was harvested and assessed by SDS-PAGE (**Figure 6 (C)**). Here, the protein was observed to be present in the soluble fraction (**Figure 6 (C)**, lane 2), indicating that the solubilisation buffer was effective. As sarkosyl is a detergent that can disrupt bacteria cell membranes, it needed to be removed from the protein suspension before lytic activity could be assessed. Pierce protein concentrators (Thermo Fischer) with a 10K molecular weight cut-off (MWCO) were used to perform a buffer exchange. The suspension was centrifuged in 15-minute intervals until the sarkosyl-containing buffer was removed. The suspension was exchanged into 50 mM sodium phosphate, 150 mM NaCl and 20% glycerol. However,

on the final wash step, when the sarkosyl was completely removed, the protein precipitated out of the solution. To assess the activity of the remaining solution, spot assays and turbidity reduction assays (**supplementary figure S5**) were performed with *C. acnes* DSM 1897, but no lysis was observed.

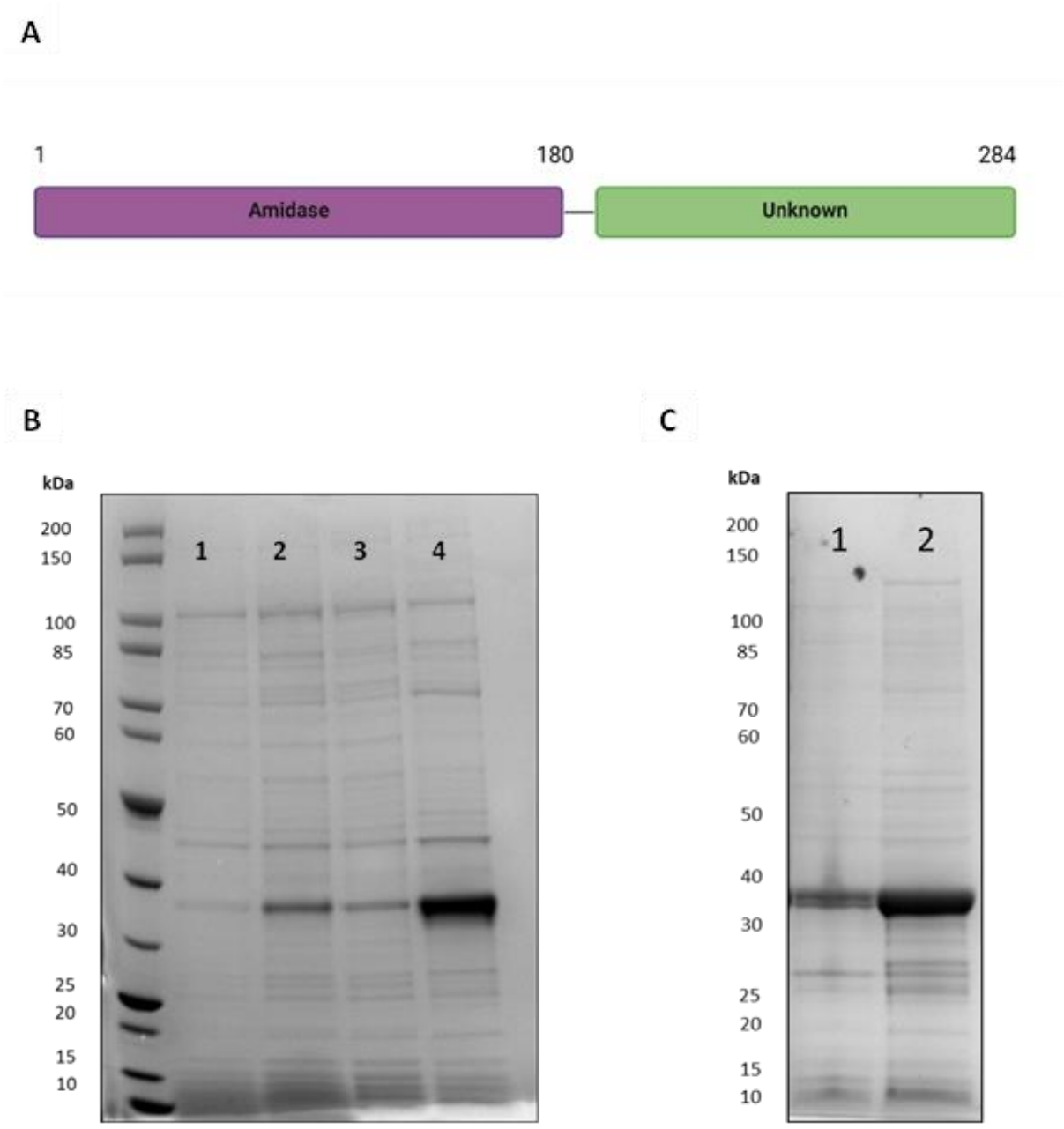
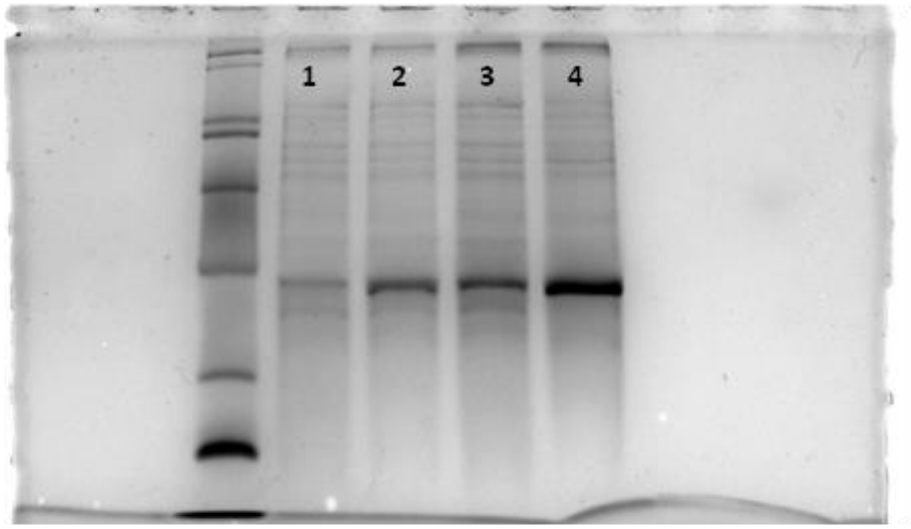


Figure 6. Cloning and Recombinant Expression of LysCa5A2. A) Domains of LysCA5A2. N-terminal catalytic domain (1-180) and C-terminal putative binding domain (180-284). **B)** Protein expression assessed by 4-12% gradient SDS-PAGE. Lane 1: crude *E. coli* before IPTG induction. Lane 2: crude *E. coli* post-IPTG induction. Lane 3: soluble fraction of the cell lysate. Lane 4: insoluble fraction of the cell lysate**.** **C)** Solubilisation of inclusion bodies, lane I: insoluble fraction, lane 2: soluble fraction.

A



B

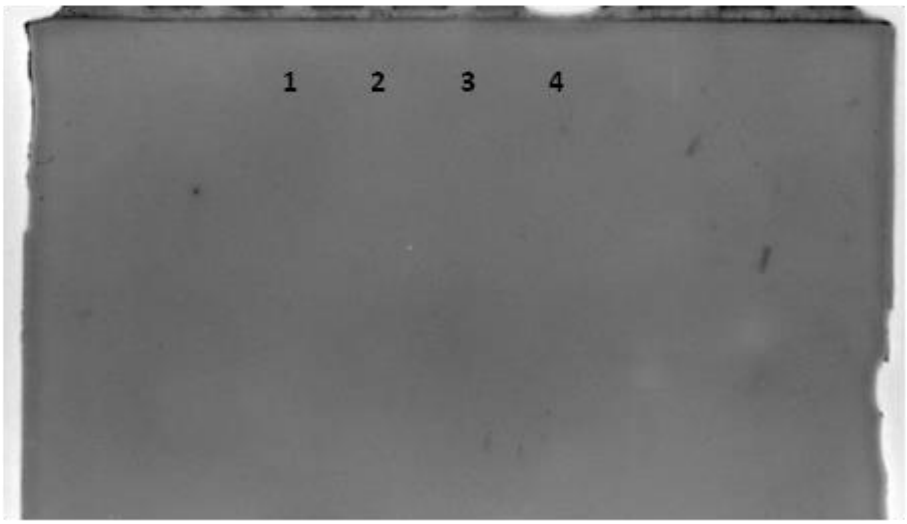


Figure 7. Zymogram analysis of LysCa5A2 against *C. acnes* DMS 1897. A) SDS-PAGE with a ladder, Lane 1: crude *E. coli* before IPTG induction. Lane 2: crude *E. coli* post-IPTG induction. Lane 3: soluble fraction of the cell lysate. Lane 4: insoluble fraction of the cell lysate. **B)** zymogram analysis with samples loaded in the same pattern.

Discussion

C. acnes is a significant contributor to the skin condition acne vulgaris. This bacterium has strong links to the inflammation associated with acne. Current treatment plans are lacking and are associated with a plethora of side effects [20]. These off-target effects, which can manifest as skin irritation and secondary conditions, not only pose physical discomfort but also have a profound impact on the mental well-being of patients. This underscores the critical need for alternative treatments to address this widespread condition.

In this study, we present the isolation of six novel phages infecting *C. acnes* (005A1, 005A2, 42-1, 42-2, 42-3, and 42-4). These phages are taxonomically placed in the genus *Pahexavirus*. The newly isolated phages had a close homology at the nucleotide level with a genome size of 29 – 30 kbp. The endolysin gene in each phage was identified and investigated for biotechnological applications. These endolysins were found to have 100% similarity on an amino-acid level. The endolysin from phage 005A2, LysCa5A2, was assessed for its potential as a therapeutic for acne vulgaris by targeting *C. acnes* DMS 1897. LysCa5A2 was cloned and recombinantly expressed in *E. coli*. Downstream analysis revealed that LysCa5A2 was produced as an insoluble protein, likely in inclusion bodies. Inclusion bodies are aggregates of protein with non-native conformations [41]. Despite a number of assays being performed, no lytic activity against *C. acnes* DMS 1897 was observed.

In 2012, a study was published that assessed the genetic diversity of *C. acnes* phages [42]. Eleven phages were isolated, and their genomes were sequenced. This study reported that even though the phages had a wide geographical spread, their genetic diversity was low. The nucleotide identity of the phages was > 85%. They were reported to have significant similarity in GC content, gene content and genome length. A pangenome and comparative analysis of *C. acnes* strains was published in 2013 [43]. This study reported that while the pangenome is open, relatively small and expanding slowly. The analysis was carried out on 82 *C. acnes* strains, including subtypes IA, IB, II, and III. The conserved regions shared by all the sequenced strains accounted for 88% of the average genome. The lack of diversity in strains of *C. acnes* may suggest why the phages of *Pahexavirus* also have low genetic diversity. These previous findings were consistent with the phylogenetic analysis of our study. Of the

57 strains assessed, the majority had > 80% nucleotide identity, and the genes were largely conserved. The GC content of the newly isolated phage genomes was also consistent with other characterised *C. acnes* phages. For example, *C. acnes* phage CAP 10-3 has a genome with 53.86% GC content and 29 kbp in length [32].

Applications of *C. acnes* phages include topical applications of *C. acnes* phage to acne-like lesions and synergy with bacteriocins. In a 2023 study, topical application of eight phages was used in an in vivo mouse model [44]. The phages were applied to acne-like lesions in a gel matrix. It was reported that the phage treatment could effectively reduce both the bacterial load and the inflammation in induced lesions. Synergy is an effective technique that allows for two antimicrobial agents to have a greater effect when used together. Combined use of phage PAP 1-1 and Nisin was shown to have significant antimicrobial action against *C. acnes* KCTC 3314 [45]. *C. acnes* phages isolated during this study could yet prove to be therapeutic candidates. Future perspectives of this work should focus on the lytic activity of the phages, the potential for synergy and applications in a gel formulation. The development of a phage suspension in a vehicle for topical application, such as a cream of hydrogel, would be beneficial to skin experiencing acne. These lesions can often be irritated and have a damaged surrounding skin barrier. A cream suspension can provide hydration, is easy to apply, and is minimally irritating to the skin [46].

This study focused on the biotechnological potential of the newly isolated phages. As the six endolysin genes had 100% nucleotide similarity, the endolysin from phage 005A2 was chosen as a representative candidate. This endolysin, LysCa5A2, was cloned, expressed, and tested for lytic activity against *C. acnes*. Due to issues with the expression of a soluble protein, LysCa5A2 has not displayed any lytic activity to date. We conclude that the LysCa5A2 is being improperly folded during protein production. This misfolding likely affects its active site, rendering it enzymatically inactive. To assess how this lysin fits amongst currently described lysins in the literature, an *in silico* comparative analysis was performed. Three endolysins (*Pa*Ami1, CAP 10-3 and LysP11) have been shown to be enzymatically active. All four endolysins that were compared have the same predicted mechanisms of action, are >94% similar on a nucleotide level, and have similar protein structures. Much like the parental phage, the endolysins are highly conserved.

In the literature, some challenges of insolubility during protein production were reported. *Pa*Ami1 was introduced into a selection of nine different *E. coli* expression strains. *E. coli* BL21 SHuffle® T7 Express cells were selected as the final expression cells as they are suitable for cysteine-rich proteins. Much like LysCa5A2, *Pa*Ami1 was expressed as an insoluble protein in inclusion bodies. In this case, the denatured protein was solubilised and recovered using dialysis [25]. The endolysin of phage CAP 10-3 was cloned into a pET-15b vector and expressed in *E. coli* BL21(DE3). The authors of this study did not report any issues with solubility and recorded lytic activity from the supernatant of cell lysate [32]. LysP11 did not report any activity against *C. acnes* strains but did have activity against *E. rhusiopathiae*. To functionally characterise the C-terminal domain of LysP11, this study involved the fusion of the putative CBD of LysP11 to an enhanced green fluorescent protein (eGFP). Binding analysis with the fluorescent protein revealed that it bound with high efficiency to *E. rhusiopathiae* cells.

Insoluble proteins are a common bottleneck in recombinant protein production. *E. coli* is a popular organism that is regularly used for the expression of high yields due to its fast growth rate and genetic accessibility [47]. While these are favourable traits, the speed at which protein is produced in *E. coli* can result in improper folding under certain conditions [48]. A simple fix to address this problem can focus on slowing down the growth rate of *E. coli* by lowering the vessel temperature, adding a reduced amount of IPTG, and reducing the expression time. In this study, we adopted these methods. The expression was carried out at a lower temperature (18 ^{o}C), IPTG was added at a 10-fold lower concentration than the other endolysins described in this thesis (0.1 mM), and the cells were harvested after 14 hours of incubation instead of 24 hours. Unfortunately, these measures did not seem to have a significant effect on the solubility of LysCa5A2. We suggest that going forward, two strategies should be utilised. One approach is to clone LysCa5A2 into a more suitable vector and transform it into a specialised expression strain. As mentioned in the *Pa*Ami study, a range of vectors and cells were assessed until an optimal system was designed. A range of fusion tags are readily available in pET vectors that can be used for solubility improvement. These include thioredoxin (Trx-tag) and glutathione-S-transferase (GST) [49] [50]. The second approach would be to use the existing expression system and recover an active protein from inclusion bodies. While the solubilisation of

LysCa5A2 in this study showed promise, the methods need to be optimised. The use of a one-step size-exclusion method to remove the denaturant likely led to protein aggregates and a low yield. A stepwise dialysis involving the gradual removal of the denaturant could improve the recovery of native protein [41].

During the comparative analysis, multiple sequence alignments showed that LysCa5A2 and LysP11 had the highest homology. Both lysins had an additional amino acid, serine, at position 193. It is possible that because of this similarity, LysCa5A2 may also present an interesting spectrum of action. To assess whether LysCa5A2 does not have activity against *C. acnes* and can instead act upon another candidate strain, an expanded spectrum of activity analysis should be carried out. Another method to identify a host strain could be to adopt the use of a green fluorescent fusion protein. Creating a GFP fusion C-terminal of LysCa5A2 could allow for the screening of candidate strains by fluorescent microscopy.

This study has highlighted the hurdles that can arise during *in silico* analysis of sequencing data, and the issues that can occur during protein production. While decontaminating steps in assembly pipelines can be an effective troubleshooting step, DNA extractions should be performed with aseptic techniques in mind. As an expression system, *E. coli* has limitations. While well-characterised vectors like the pET28b(+) are often a first choice for protein expression, a range of vectors should be tested for the specific requirements of each target protein.

Materials and Methods

Bacterial Strains and Culturing Conditions.

C. acnes strains were obtained from DSMZ, Germany, and were routinely grown in tryptone soy broth (TSB) (Merck, Ireland), 1.5% TSB agar and 0.4% TSB soft agar for overlay assays, incubated at 37°C anaerobically. These strains included *Cutibacterium acnes* 1897, *Cutibacterium acnes* 16379, Cutibacterium acne 30738, *Cutibacterium acnes* 30753, and *Cutibacterium acnes* 30919. Anaerobic conditions were maintained using an anaerobic jar supplemented with an AnaeroPack (Thermo Scientific).

Sampling, Phage Screening and Purification.

Skin swab samples were used for initial screening. Samples from 65 subjects were collected, 19 samples from a farm, and 30 sewage samples were collected. Skin swab samples were collected from the faces, armpits, and feet of healthy male and female subjects (protocol-APC071). The swabs were suspended in 3ml of SM buffer and set aside for 30 minutes.

Screening for potential phages against *C. acnes* was performed by three-step enrichment of sample – 1 mL in an exponential phase, 9 mL culture. The contents were centrifuged (Thermo Scientific) at 5000 x g for 15 mins, and the supernatant was passed through a 0.45 µm filter (Millipore). Plaque and spot assays were performed using the filtrate to examine the presence of phage. Harvesting of the phages was performed using the supernatant in a three-step agar overlay method. Isolated plaques were selected, re-suspended in 300 µL of SM buffer and centrifuged at 13,000 x g for 5 mins.

The harvested phages were propagated by transferring 200 µL of the supernatant into an actively growing *C. acnes* culture and concentrated by re-suspending 400 mL log phase culture in 10 mL broth and adding 5 mL of phage lysate. The titre of the concentrated phages was calculated, and the samples were stored at 4°C for further analysis.

Transmission Electron Microscopy.

C. acnes phages were purified for TEM imaging. The phage suspensions were applied to a step gradient of 5 M and 3 M caesium chloride (CsCl) solutions, followed by centrifugation using an F65L-6×13.5 rotor (Thermo Scientific) at 105,000 g for 2.5 hours at 4°C. The visible band containing viral particles was collected and subjected to CsCl clean-up steps. 5 µL aliquots of the concentrated viral fraction were applied to Formvar/Carbon 200 Mesh, Cu grids (Electron Microscopy Sciences, PA, USA) with subsequent removal of the excess sample by blotting. Grids were then negatively contrasted with 0.5% (w/v) uranyl acetate and examined at UCD Conway Imaging Core Facility (University College Dublin, Dublin, Ireland) by Tecnai G2 12 BioTWIN transmission electron microscope.

DNA Extraction and Sequencing of C. acnes Phages.

The viral DNA extraction protocol previously described by Shkoporov et al was used with slight modifications [51]. To 10 mL of the phage lysate, 5M NaCl and PEG-8000 were added to give a final concentration of 0.5M and 10% w/w. The samples were centrifuged at 5000 x g for 20 minutes. The supernatant was discarded, and the pellets were re-suspended in 400 µL of SM buffer. An equal volume of chloroform was added, and the sample was centrifuged at 2500 x g for 5 mins. The aqueous layer was transferred into a clean Eppendorf tube and treated with 8 U TURBO DNase (Ambion/ThermoFisher Scientific) and 20 U of RNase I (ThermoFisher Scientific) for 1 hour at 37°C. The enzymes were then inactivated at 70 °C for 10 mins. Proteinase-K (20µL) and 0.5% SDS treatment was performed at 56 °C for 20 mins. The viral particles were lysed using a 100 µL Phage Lysis Buffer (4.5 M guanidinium isothiocyanate, 44 mM sodium citrate pH 7.0, 0.88% sarkosyl and 0.72% 2-mercaptoethanol) at 65°C for 20 minutes. The lysates were extracted using a two-step phenol/chloroform/isoamyl alcohol 25:24:1 (Fisher Scientific) wash step. The final DNA contents were passed through a DNeasy Blood and Tissue Kit (Qiagen) to obtain purified viral DNA products.

The viral DNA was quantified using a Qubit dsDNA HS Assay kit (ThermoFisher Scientific).1 µL of the DNA was subjected to multiple displacement amplification (MDA) using Illustra GenomiPhi V2 kit (GE Healthcare) according to the manufacturer's instructions. The samples were normalised to 0.2 ng/µL for tagmentation using the

Nextera XT protocol (Illumina). All samples were run on an Agilent Technologies 2100 Bioanalyzer using the high-sensitivity DNA chip. The prepared libraries were then sequenced on an Illumina HiSeq 4000 platform with 2 x 150 nt paired-end chemistry at GATC Biotech AG, Germany.

Genome Analysis.

The quality of the raw paired-end reads was analysed using FastQC (v0.11.5). Trimming and filtering of the reads were performed with fastp (to filter based on minimum length requirement, the option --length_required 60 was used. To enable automatic detection adaptor sequences for PE data, the option --detect_adapter_for_pe was used) [52]. Human, mouse, and contaminating reads were removed from the whole metagenome with kraken2 [53] (the option --unclassified-out was used to extract unmapped reads). This was followed up with a step to remove PhiX reads with bbmap (option k-32 hdist=1 was used to remove all reads that have a 31-mer match to PhiX, allowing one mismatch). Subsequent steps were taken to decontaminate reads from *E. coli, C. acnes* bacterial genome, and *C. acnes* plasmids. The trimmed, filtered, and decontaminated reads were then assembled into contigs using SPAdes (v1.13.1) (run in –meta option, and -k 21,33,55 option to specify k-mer sizes used for the assembly process. These settings were utilised to improve the quality and accuracy of the assembly) [54]. All *C. acnes* phage sequences were annotated with Pharokka v1.5.0 [26] and further annotated based on protein structures using phold [27] and phage synteny using phynteny, all with default parameters [28].

A range of *in silico* tools assessed the genome analysis of the *C. acnes* phages isolated in this study. TaxmyPHAGE [31] was used to assign taxonomy to *C. acnes* phages. This was run with default settings. Colabfold v 1.5.2 was used to predict the protein structure of the large terminase (TerL) [34]. To create a phylogenetic tree, FoldTree v1 [55] was employed. This was created with no outgroup but was rooted in the FoldTree algorithm. First, Pharokka v 1.7.0 [26] was used to translate proteins and reorder them to start with the TerL (option –dnaapler to reorder the genome to start with TerL). To create the comparison for the figure, blastn [56] was used to search nucleotide all vs all, and DIAMOND v 0.9.25 sequence aligner was used with BLASTp for proteins all vs all (option –outfmt 6 –k0 to report all alignments for the query

sequence) [35]. The resulting figure was created using the R packages gggenomes v 1.0.0 and ggtree v 3.19.

VIRIDIC v1.1 was used to calculate virus intergenomic similarities of the *C. acnes* phage genomes against each other, using default settings [29]. A map of the individual phage genomes was prepared using Proksee, using the annotations as prepared in the previous steps [30].

Morphological Analysis of the Phages.

Phage plaque assays and spot assays were carried out using the double-layer plate method as previously described [57]. This method was also used to assess the host range of the *C. acnes* phages. Any phage lysate harvested from this method was filtered through a 0.45 µM filter to remove residual bacterial debris and stored at 4 °C.

Separation of plaques to purify the phages based on plaque morphology was carried out using a protocol adapted from Methods in Molecular Biology [57]. In this "phage T-streak" method, the phage lysate was streaked across a solid agar plate using a standard loop. This streak was allowed to dry before being overlaid with semi-solid (0.4%) agar containing a 2% inoculum of *C. acnes* DSM 1897. Once the lawn had grown, single plaques were harvested and resuspended in SM buffer. This technique was repeated until a pure lysate was obtained.

In silico Characterisation of Endolysins Genes.

Structural analysis of endolysin genes was carried out using the AlphaFold3 server [38]. The accession numbers of other characterised *C. acnes* endolysins used for comparative analysis are as follows – LysP11: YP_006906607.1, CAP 10-3: WIW76974.1, *Pa*Ami1: YP_009214894.1. Homology searches were performed for LysCa5A2 in HHpred [58] and InterPro Scan [36]. Homology searches based on protein structure were carried out using Foldseek [37]. Multiple sequence alignments and percentage identity matrices were generated by Clustal Omega [39]. The alignment was visualised using the SeaView4 software [59]. The percentage identity matrix was visualised using Heatmap2 on the Galaxy Europe Server.

Cloning.

Putative *C. acnes* endolysin genes in the phage genomes were identified using BLASTp from NCBI [59]. The endolysin gene from phage 005A2, LysCa5A2, was

amplified from a pure phage lysate. Amplification was carried out with KOD Hot Start DNA Polymerase (Merck). The primers used for amplification of LysCa5A2 are as follows with restriction sites underlined: forward primer 5'-TCGAATTCGatggtgagatacattccagcggc–3', reverse primer 5'–ATAAAGCTTatcacttttttcacatcgttgac – 3' . These primers introduced an alternative start codon of a methionine. The PCR products were cleaned and concentrated using Zymo Research Clean and Concentrate DNA Kit (Zymo Research). The PCR product and the plasmid pET28b(+) were digested using restriction enzymes Fast Digest EcoRI and HindIII (Thermo Scientific) and ligated with T4 ligase (Rapid Ligation Kit, Thermo Scientific) to create an N-terminally 6xHis-tagged construct of pET28b(+) with insert LysCa5A2. The construct was introduced into *E. coli* "One Shot" Top10 chemically competent cells (Invitrogen) by heat-shock transformation. Transformants were screened for positive clones, which were propagated for plasmid extractions. Extracted plasmids positive for the correct vector size were introduced into One Shot BL21(DE3) chemically competent cells (Invitrogen) for protein expression. Sanger sequencing was utilised to confirm the integrity of the construct pET28b(+) LysCa5A2.

Recombinant Protein Expression and Purification.

Protein expression was carried out by the following method. A large preparation of *E. coli* BL21(DE3) cells containing pET28b(+)-LysCa5A2 (three litres) was grown in LB broth supplemented with 50 µg/mL Kanamycin and incubated at 37 ^{o}C at 150 rpm shaking. Cells were grown to an OD of 0.6, and expression was induced by the addition of 0.1 mM final concentration IPTG (Sigma Aldrich). Cultures were incubated, shaking, overnight in temperature-controlled room that was set to 18 ^{o}C. Cells were harvested the following day by centrifugation at 4,500 rpm for 30 minutes. Cell pellets were lysed using sonication on ice, pelleted again by centrifugation and the supernatant was extracted for protein purification. The insoluble cell pellet was stored for further analysis.

The Akta Start Purification System (Cytiva) facilitated the purification of proteins in the soluble fraction using the Talon Crude Purification 5 mL columns (Cytiva). Purification fractions were assessed for quality and purity by SDS-PAGE (4-12% gradient gel) (Invitrogen) and imaged with the iBright500 imaging system.

Protein Solubilisation

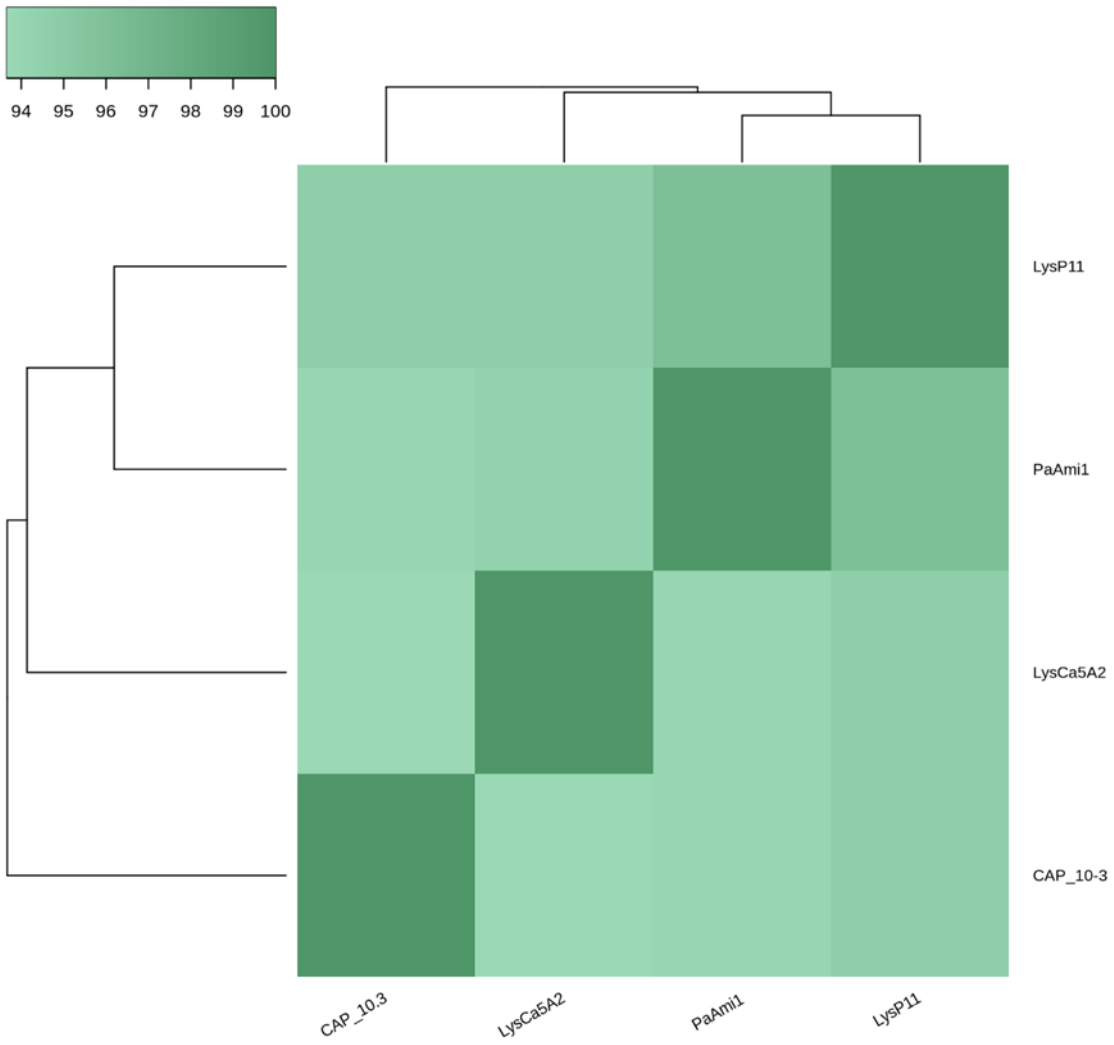
To recover protein from the insoluble fraction of the cell lysate, a solubilisation step was performed. 1mL of insoluble protein suspension was harvested by centrifugation at 5,000 x g for 10 minutes. The insoluble pellet was washed three times with PBS. The pellet was then resuspended in 3 mL 50 mM sodium phosphate buffer with 1% w/v N-lauroylsarcosine (sarkosyl) (Sigma Aldrich). This suspension was left to shake at room temperature for two hours. The suspension was centrifuged at 10,000 xg for 5 minutes. The supernatant was harvested and assessed by SDS-PAGE.

To perform a buffer exchange to remove sarkosyl from the buffer, pierce protein concentration columns were used (Thermo fisher). These columns had a 10k molecular weight cut-off (MWCO). The soluble supernatant was washed through the spin columns and the waste buffer was discarded. The replacement buffer was washed through three times before the protein was resuspended.

Lytic Activity of Endolysin LysCa5A2.

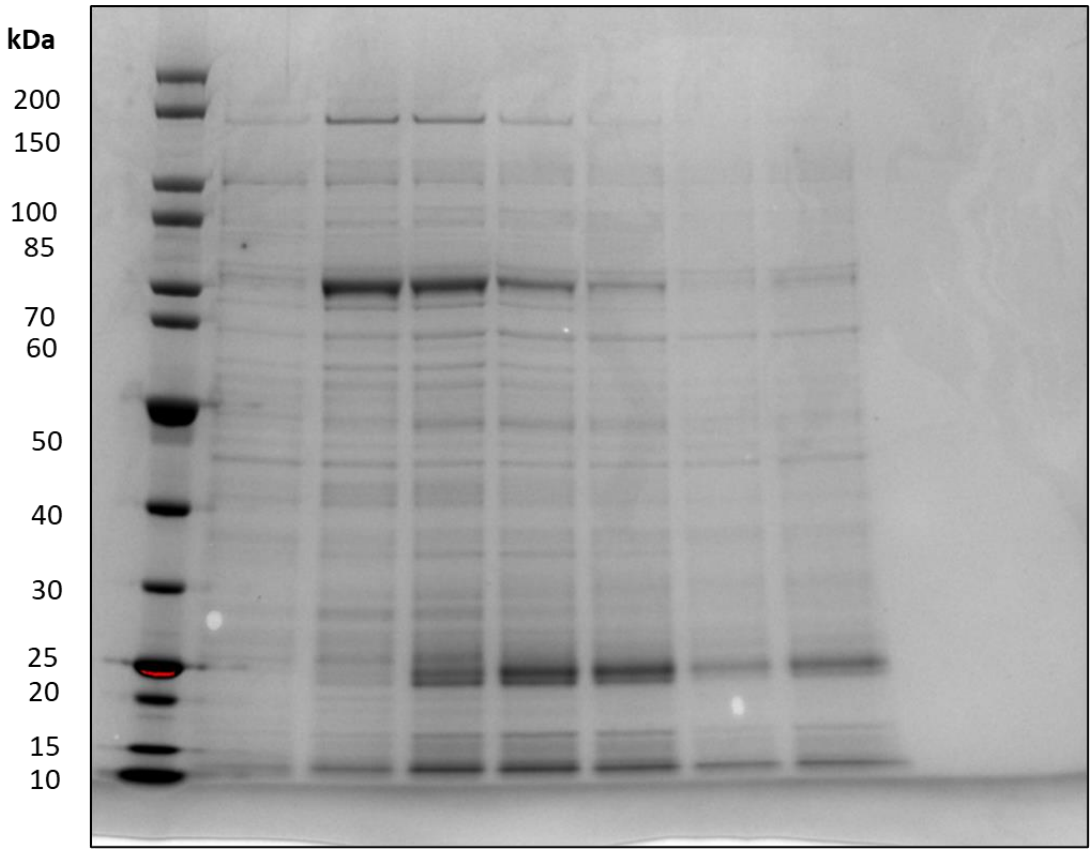
Zymogram analysis was used to assess activity in the soluble and insoluble fractions of LysCa5A2. This method was carried out as previously described, with modifications [60]. *C. acnes* DSM 1897 was prepared as a 250 mL culture. The cell preparation was autoclaved. Autoclaved cells and cell debris were harvested by centrifugation at 3,000 x g for 20 minutes. The cell pellet was washed three times with PBS before being resuspended in a final volume of 600 µL sodium phosphate buffer.

Turbidity reduction assays were used to assess the lytic activity of soluble protein. An overnight culture of the *C. acnes* DMS1897 was grown in liquid media to an OD600 of ~ 1.0. The cells were inactivated and then pelleted by centrifugation at 3000 x g for 5 minutes. The cell pellet was washed three times in PBS before being resuspended to an OD600 of 0.8. A 96-well plate was prepared to 200 µL final volume of cell suspension with the addition of purified protein or a negative control of buffer. Changes in optical density were recorded every 5 minutes over the course of 2 hours at 37 ^{o}C using a microplate reader (Thermo Fischer).

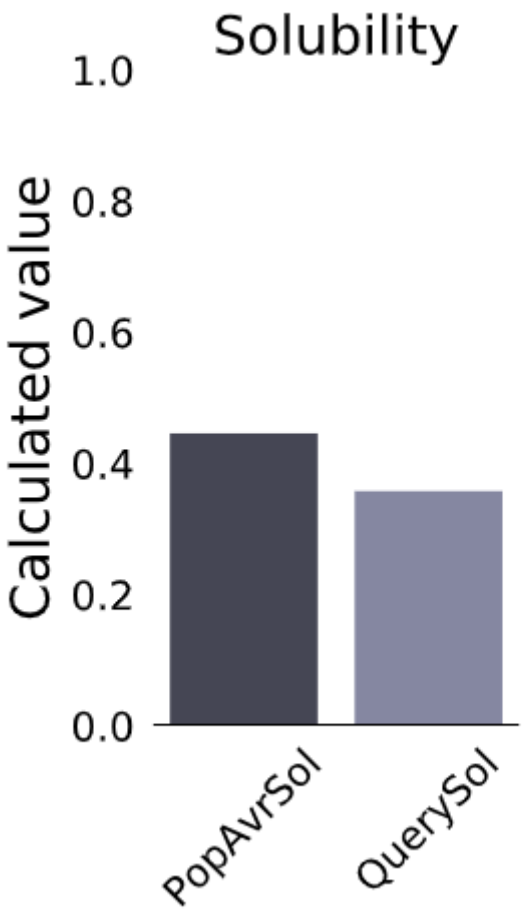


Supplementary Figure S1. Percentage Identity Matrix of *C. acnes* phages. Generated with Clustal Omega and Visualised with Heatmap2.

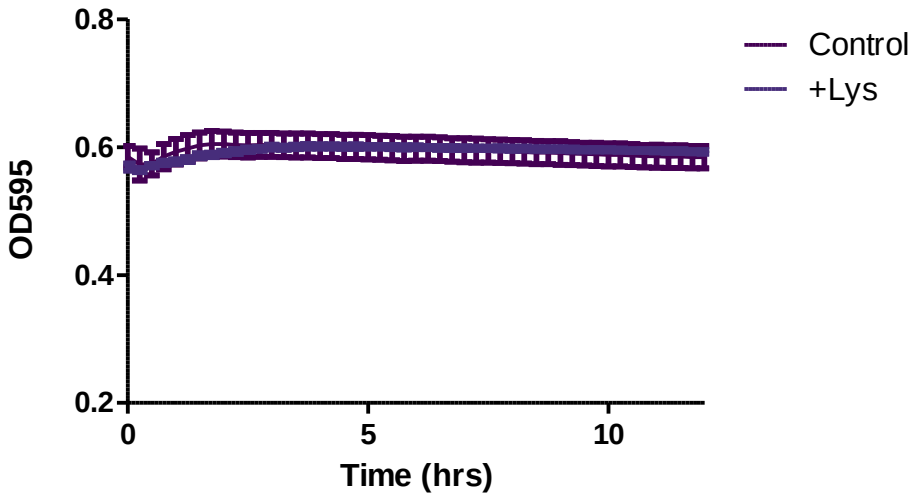
Supplementary Figure S2. Chromatogram of LysCa5A2 purification.



Supplementary Figure S3. **SDS-PAGE analysis of Akta purification fractions.** Each lane contains a sample taken from the purification process. Fractions should consist of a pure protein band at an expected size – in the case of LysCa5A2 at ~ 33 kDa.



Supplementary Figure S4. Protein solubility prediction for LysCa5A2.



Supplementary Figure S5. Turbidity reduction assay against *C. acnes* DSM 1897.

Author Contributions

EM and CH conceptualised the study. ARF isolated the phages and performed TEM analysis, DNA extractions and wet lab sequencing preparations. RD and EM assembled the phage genomes. RD and EM carried out bioinformatic analysis. EM performed endolysin *in silico* analysis and wet lab work. EM, CH, and PR wrote the manuscript.

References

[1] C. F. P. Scholz and M. Kilian, “The natural history of cutaneous propionibacteria, and reclassification of selected species within the genus Propionibacterium to the proposed novel genera Acidipropionibacterium gen. nov., Cutibacterium gen. nov. and Pseudopropionibacterium gen. nov,” Int J Syst Evol Microbiol, vol. 66, no. 11, pp. 4422–4432, Nov. 2016, doi: 10.1099/IJSEM.0.001367.

[2] B. Dréno, S. Pécastaings, S. Corvec, S. Veraldi, A. Khammari, and C. Roques, “Cutibacterium acnes (Propionibacterium acnes) and acne vulgaris: a brief look at the latest updates,” Journal of the European Academy of Dermatology and Venereology, vol. 32, pp. 5–14, Jun. 2018, doi: 10.1111/JDV.15043.

[3] R. Agarwal, O. P. Katare, and S. P. Vyas, “The pilosebaceous unit: a pivotal route for topical drug delivery,” Methods Find Exp Clin Pharmacol, vol. 22, no. 2, pp. 129–133, 2000, doi: 10.1358/MF.2000.22.2.796082.

[4] B. J. Fowler, D. Miller, X. Yan, N. A. Yannuzzi, and H. W. Flynn, “Postoperative Endophthalmitis Caused by Cutibacterium (Formerly Propionibacterium) Acnes: Case Series and Review,” Case Rep Ophthalmol, vol. 12, no. 1, p. 1, Jan. 2021, doi: 10.1159/000510208.

[5] A. El Rafei et al., “Cardiovascular Implantable Electronic Device Infections due to Propionibacterium Species,” Pacing Clin Electrophysiol, vol. 39, no. 6, p. 522, Jun. 2016, doi: 10.1111/PACE.12845.

[6] H. H. Vilchez et al., “Prosthetic Shoulder Joint Infection by Cutibacterium acnes: Does Rifampin Improve Prognosis? A Retrospective, Multicenter, Observational Study,” Antibiotics, vol. 10, no. 5, p. 475, May 2021, doi: 10.3390/ANTIBIOTICS10050475.

[7] M. Passerini et al., “Cutibacterium acnes in spine tissue: characteristics and outcomes of non-hardware-associated vertebral osteomyelitis,” J Bone Jt Infect, vol. 8, no. 2, p. 143, Apr. 2023, doi: 10.5194/JBJI-8-143-2023.

[8] G. Condrò, M. Guerini, M. Castello, and P. Perugini, “Acne Vulgaris, Atopic Dermatitis and Rosacea: The Role of the Skin Microbiota—A Review,” Biomedicines

2022, Vol. 10, Page 2523, vol. 10, no. 10, p. 2523, Oct. 2022, doi: 10.3390/BIOMEDICINES10102523.

[9] G. F. Webster, “The pathophysiology of acne.,” Cutis, vol. 76, no. 2 Suppl, pp. 4–7, Aug. 2005, Accessed: Jun. 14, 2024. [Online]. Available: https://europepmc.org/article/med/16164150

[10] A. H. S. Heng and F. T. Chew, “Systematic review of the epidemiology of acne vulgaris,” Sci Rep, vol. 10, no. 1, Dec. 2020, doi: 10.1038/S41598-020-62715-3.

[11] K. Bhate and H. C. Williams, “Epidemiology of acne vulgaris,” British Journal of Dermatology, vol. 168, no. 3, pp. 474–485, Mar. 2013, doi: 10.1111/BJD.12149.

[12] B. Dréno, S. Pécastaings, S. Corvec, S. Veraldi, A. Khammari, and C. Roques, “Cutibacterium acnes (Propionibacterium acnes) and acne vulgaris: a brief look at the latest updates,” Journal of the European Academy of Dermatology and Venereology, vol. 32, pp. 5–14, Jun. 2018, doi: 10.1111/JDV.15043.

[13] M. P. Cros et al., “New insights into the role of Cutibacterium acnes-derived extracellular vesicles in inflammatory skin disorders,” Scientific Reports |, vol. 13, p. 16058, 123AD, doi: 10.1038/s41598-023-43354-w.

[14] I. Cavallo et al., “Skin dysbiosis and Cutibacterium acnes biofilm in inflammatory acne lesions of adolescents,” Sci Rep, vol. 12, no. 1, p. 21104, Dec. 2022, doi: 10.1038/S41598-022-25436-3.

[15] E. Alkhawaja, S. Hammadi, M. Abdelmalek, N. Mahasneh, B. Alkhawaja, and S. M. Abdelmalek, “Antibiotic resistant Cutibacterium acnes among acne patients in Jordan: a cross sectional study,” BMC Dermatol, vol. 20, no. 1, Dec. 2020, doi: 10.1186/S12895-020-00108-9.

[16] W. I. Worret and J. W. Fluhr, “Acne therapy with topical benzoyl peroxide, antibiotics and azelaic acid,” JDDG: Journal der Deutschen Dermatologischen Gesellschaft, vol. 4, no. 4, pp. 293–300, Apr. 2006, doi: 10.1111/J.1610-0387.2006.05931.X.

[17] D. Rigopoulos, D. Ioannides, D. Kalogeromitros, and A. D. Katsambas, “Comparison of topical retinoids in the treatment of acne,” Clin Dermatol, vol. 22, no. 5, pp. 408–411, Sep. 2004, doi: 10.1016/J.CLINDERMATOL.2004.03.013.

[18] A. V. Gannesen et al., “Composition of the biofilm matrix of cutibacterium acnes acneic strain RT5,” Front Microbiol, vol. 10, no. JUN, p. 456374, Jun. 2019, doi: 10.3389/FMICB.2019.01284/BIBTEX.

[19] J. J. Leyden, M. Wortzman, and E. K. Baldwin, “Antibiotic-Resistant Propionibacterium acnes Suppressed by a Benzoyl Peroxide Cleanser 6%,” vol. 82, p. 417, 2008.

[20] K. A. Habeshian and B. A. Cohen, “Current Issues in the Treatment of Acne Vulgaris,” Pediatrics, vol. 145, no. Supplement_2, pp. S225–S230, May 2020, doi: 10.1542/PEDS.2019-2056L.

[21] M. Ludot, S. Mouchabac, and F. Ferreri, “Inter-relationships between isotretinoin treatment and psychiatric disorders: Depression, bipolar disorder, anxiety, psychosis and suicide risks,” World J Psychiatry, vol. 5, no. 2, p. 222, Jun. 2015, doi: 10.5498/WJP.V5.I2.222.

[22] A. P. Fabijan, J. Iredell, K. Danis-Wlodarczyk, R. Kebriaei, and S. T. Abedon, “Translating phage therapy into the clinic: Recent accomplishments but continuing challenges,” PLoS Biol, vol. 21, no. 5, p. e3002119, May 2023, doi: 10.1371/JOURNAL.PBIO.3002119.

[23] E. Murray, L. A. Draper, R. P. Ross, and C. Hill, “The advantages and challenges of using endolysins in a clinical setting,” Viruses, vol. 13, no. 4, 2021, doi: 10.3390/v13040680.

[24] J. I. Kim, M. A. Hasnain, and G. S. Moon, “Expression of a recombinant endolysin from bacteriophage CAP 10-3 with lytic activity against Cutibacterium acnes,” Scientific Reports 2023 13:1, vol. 13, no. 1, pp. 1–9, Sep. 2023, doi: 10.1038/s41598-023-43559-z.

[25] C. Varotsou, G. E. Premetis, and N. E. Labrou, “Characterization and Engineering Studies of a New Endolysin from the Propionibacterium acnes Bacteriophage PAC1 for the Development of a Broad-Spectrum Artilysin with Altered Specificity,” Int J Mol Sci, vol. 24, no. 10, p. 8523, May 2023, doi: 10.3390/IJMS24108523/S1.

[26] G. Bouras, R. Nepal, G. Houtak, A. J. Psaltis, P. J. Wormald, and S. Vreugde, “Pharokka: a fast scalable bacteriophage annotation tool,” Bioinformatics, vol. 39, no. 1, Jan. 2023, doi: 10.1093/BIOINFORMATICS/BTAC776.

[27] “GitHub - gbouras13/phold: Phage Annotation using Protein Structures.” Accessed: Jun. 11, 2024. [Online]. Available: https://github.com/gbouras13/phold

[28] “GitHub - susiegriggo/Phynteny: Predict the function of phage hypothetical proteins using an LSTM model trained with Phage Synteny.” Accessed: Jun. 11, 2024. [Online]. Available: https://github.com/susiegriggo/Phynteny

[29] C. Moraru, A. Varsani, and A. M. Kropinski, “VIRIDIC-A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses,” Viruses, vol. 12, no. 11, Nov. 2020, doi: 10.3390/V12111268.

[30] J. R. Grant et al., “Proksee: in-depth characterization and visualization of bacterial genomes,” Nucleic Acids Res, vol. 51, no. W1, pp. W484–W492, Jul. 2023, doi: 10.1093/NAR/GKAD326.

[31] “GitHub - amillard/tax_myPHAGE: Workflow to assign Taxonomy to a bacteriophage genome isolate.” Accessed: Jun. 15, 2024. [Online]. Available: https://github.com/amillard/tax_myPHAGE/tree/main

[32] J. I. Kim, M. A. Hasnain, and G. S. Moon, “Expression of a recombinant endolysin from bacteriophage CAP 10-3 with lytic activity against Cutibacterium acnes,” Scientific Reports 2023 13:1, vol. 13, no. 1, pp. 1–9, Sep. 2023, doi: 10.1038/s41598-023-43559-z.

[33] “GitHub - DessimozLab/fold_tree: snakemake pipeline for creating trees from sequence sets.” Accessed: Jun. 15, 2024. [Online]. Available: https://github.com/DessimozLab/fold_tree

[34] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, “ColabFold: making protein folding accessible to all,” Nat Methods, vol. 19, no. 6, p. 679, Jun. 2022, doi: 10.1038/S41592-022-01488-1.

[35] B. Buchfink, K. Reuter, and H. G. Drost, “Sensitive protein alignments at tree-of-life scale using DIAMOND,” Nat Methods, vol. 18, no. 4, pp. 366–368, Apr. 2021, doi: 10.1038/S41592-021-01101-X.

[36] P. Jones et al., “InterProScan 5: genome-scale protein function classification,” vol. 30, no. 9, pp. 1236–1240, 2014, doi: 10.1093/bioinformatics/btu031.

[37] M. van Kempen et al., “Fast and accurate protein structure search with Foldseek,” Nature Biotechnology 2023 42:2, vol. 42, no. 2, pp. 243–246, May 2023, doi: 10.1038/s41587-023-01773-0.

[38] J. Abramson et al., “Accurate structure prediction of biomolecular interactions with AlphaFold 3,” Nature 2024, pp. 1–8, May 2024, doi: 10.1038/s41586-024-07487-w.

[39] F. Sievers et al., “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega,” Mol Syst Biol, vol. 7, p. 539, 2011, doi: 10.1038/MSB.2011.75.

[40] M. Hebditch, M. A. Carballo-Amador, S. Charonis, R. Curtis, and J. Warwicker, “Protein-Sol: A web tool for predicting protein solubility from sequence,” Bioinformatics, vol. 33, no. 19, pp. 3098–3100, Oct. 2017, doi: 10.1093/BIOINFORMATICS/BTX345.

[41] H. Yamaguchi and M. Miyazaki, “Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies,” Biomolecules, vol. 4, no. 1, p. 235, Mar. 2014, doi: 10.3390/BIOM4010235.

[42] L. J. Marinelli et al., “Propionibacterium acnes bacteriophages display limited genetic diversity and broad killing activity against bacterial skin isolates,” mBio, vol. 3, no. 5, 2012, doi: 10.1128/MBIO.00279-12/SUPPL_FILE/MBO005121334SF04.TIF.

[43] S. Tomida et al., “Pan-genome and comparative genome analyses of Propionibacterium acnes reveal its genomic diversity in the healthy and diseased human skin microbiome,” mBio, vol. 4, no. 3, 2013, doi: 10.1128/MBIO.00003-13/SUPPL_FILE/MBO002131499SF05.PDF.

[44] A. Rimon et al., “Topical phage therapy in a mouse model of Cutibacterium acnes-induced acne-like lesions,” Nature Communications 2023 14:1, vol. 14, no. 1, pp. 1–18, Feb. 2023, doi: 10.1038/s41467-023-36694-8.

[45] M. H. Han, S. A. Khan, and G. S. Moon, “Cutibacterium acnes KCTC 3314 Growth Reduction with the Combined Use of Bacteriophage PAP 1-1 and Nisin,”

Antibiotics 2023, Vol. 12, Page 1035, vol. 12, no. 6, p. 1035, Jun. 2023, doi: 10.3390/ANTIBIOTICS12061035.

[46] R. Y. K. Chang, S. Morales, Y. Okamoto, and H. K. Chan, “Topical application of bacteriophages for treatment of wound infections,” Transl Res, vol. 220, p. 153, Jun. 2020, doi: 10.1016/J.TRSL.2020.03.010.

[47] G. L. Rosano and E. A. Ceccarelli, “Recombinant protein expression in Escherichia coli: Advances and challenges,” Front Microbiol, vol. 5, no. APR, p. 79503, Apr. 2014, doi: 10.3389/FMICB.2014.00172/BIBTEX.

[48] M. Gutiérrez-González et al., “Optimization of culture conditions for the expression of three different insoluble proteins in Escherichia coli,” Scientific Reports 2019 9:1, vol. 9, no. 1, pp. 1–11, Nov. 2019, doi: 10.1038/s41598-019-53200-7.

[49] D. Esposito and D. K. Chatterjee, “Enhancement of soluble protein expression through the use of fusion tags,” Curr Opin Biotechnol, vol. 17, no. 4, pp. 353–358, Aug. 2006, doi: 10.1016/J.COPBIO.2006.06.003.

[50] S. Costa, A. Almeida, A. Castro, and L. Domingues, “Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: The novel Fh8 system,” Front Microbiol, vol. 5, no. FEB, p. 77021, Feb. 2014, doi: 10.3389/FMICB.2014.00063/BIBTEX.

[51] A. N. Shkoporov et al., “Reproducible protocols for metagenomic analysis of human faecal phageomes,” Microbiome, vol. 6, no. 1, p. 68, Apr. 2018, doi: 10.1186/S40168-018-0446-Z/FIGURES/5.

[52] S. Chen, Y. Zhou, Y. Chen, and J. Gu, “fastp: an ultra-fast all-in-one FASTQ preprocessor,” Bioinformatics, vol. 34, no. 17, pp. i884–i890, Sep. 2018, doi: 10.1093/BIOINFORMATICS/BTY560.

[53] D. E. Wood, J. Lu, and B. Langmead, “Improved metagenomic analysis with Kraken 2,” Genome Biol, vol. 20, no. 1, pp. 1–13, Nov. 2019, doi: 10.1186/S13059-019-1891-0/FIGURES/2.

[54] A. Bankevich et al., “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing,” Journal of Computational Biology, vol. 19, no. 5, p. 455, May 2012, doi: 10.1089/CMB.2012.0021.

[55] “GitHub - DessimozLab/fold_tree: snakemake pipeline for creating trees from sequence sets.” Accessed: Jun. 11, 2024. [Online]. Available: https://github.com/DessimozLab/fold_tree

[56] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” J Mol Biol, vol. 215, no. 3, pp. 403–410, 1990, doi: 10.1016/S0022-2836(05)80360-2.

[57] S. Sillankorva, “Isolation of Bacteriophages for Clinically Relevant Bacteria,” Methods in Molecular Biology, vol. 1693, pp. 23–30, 2018, doi: 10.1007/978-1-4939-7395-8_3.

[58] J. Söding, A. Biegert, and A. N. Lupas, “The HHpred interactive server for protein homology detection and structure prediction,” Nucleic Acids Res, vol. 33, no. Web Server issue, p. W244, Jul. 2005, doi: 10.1093/NAR/GKI408.

[59] M. Gouy, S. Guindon, and O. Gascuel, “SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building,” Mol Biol Evol, vol. 27, no. 2, pp. 221–224, Feb. 2010, doi: 10.1093/MOLBEV/MSP259.

[60] L. Rodríguez-Rubio, D. M. Donovan, B. Martínez, A. Rodríguez, and P. García, “Peptidoglycan Hydrolytic Activity of Bacteriophage Lytic Proteins in Zymogram Analysis,” Methods in Molecular Biology, vol. 1898, pp. 107–115, 2019, doi: 10.1007/978-1-4939-8940-9_9.

Chapter 3. LysH1 is a Novel *Enterococcus faecium* Phage Endolysin with Activity Against *Ruminococcus gnavus* in a Simplified Human Gut Consortium.

Ellen Murray, Natalia S. Ríos Colombo, David Hourigan, Andrei S. Bolocan, Lorraine A. Draper, Neda Nezam Abadi, Paul R. Ross, Colin Hill.

Abstract

Antibiotic resistance is a significant public health concern. To slow the rate of resistance development, we urgently need to develop alternative antimicrobial strategies. Endolysins are phage-derived peptidoglycan hydrolases that could be promising alternatives or adjuncts to traditional antibiotics. Here, we identified, cloned, and expressed a novel *Enterococcus faecium* bacteriophage (phage) lysin, LysH1. Structural analysis of this endolysin revealed a two-domain structure composed of a catalytic domain and a previously uncharacterised domain. Binding assays with a green fluorescent protein hybrid fusion allowed us to functionally characterise this unknown domain as the cell wall binding domain of LysH1. This is the first instance of experimental characterisation of the cell wall binding domain of an *E. faecium* endolysin. The lytic activity of LysH1 was assessed against a panel of strains and revealed that it has cross-species activity against *Ruminococcus gnavus* (also known as *Mediterraneibacter gnavus*), a bacterium associated with inflammatory bowel disease (IBD). In a simplified human intestinal microbiota, LysH1 can target *R. gnavus* without impacting the other members of the consortium. This suggests that LysH1 could be a therapeutic candidate for the management of *R. gnavus* in patients with IBD.

Introduction

Enterococcus faecium and *Ruminococcus gnavus* are important bacterial species within the human gut microbiome, each playing distinct roles in gut health. *E. faecium* is often noted for its development of multidrug resistance, especially in hospital settings, while *R. gnavus* is a mucin-degrading bacterium associated with inflammatory bowel disease (IBD).

E. faecium is a gram-positive facultative anaerobe and an opportunistic pathogen. It is commonly found in the digestive tracts of humans and animals, in soil and other natural environments. Historically, *E. faecium* was considered a relatively harmless commensal within the human gastrointestinal tract, only sporadically causing disease. However, over the last two decades, it has quickly become one of the leading causes of nosocomial infections worldwide, particularly among immunocompromised patients. Disease states caused by *E. faecium* include endocarditis, bacteraemia, urinary tract infections, and wound infections [1]. This transition from commensal to a pathogen is mainly due to the acquisition of antibiotic resistance genes (ARGs). Vancomycin-resistant enterococci (VRE) are now considered to be serious opportunistic pathogens associated with higher mortality [2]. VRE strains often represent a group of enterococci that have acquired multi-drug resistance.

E. faecalis and *E. faecium* are the most clinically relevant enterococci. *E. faecalis* is more pathogenic than *E. faecium*, but *E. faecium* is more resistant to antibiotics and accounts for the majority of VRE infections [3]. Antibiotic resistance in *Enterococcus* spp. is generally divided into two classes – intrinsic resistance and acquired resistance. Intrinsic resistance occurs when a bacterium is naturally resistant due to elements of its genome, whereas acquired resistance occurs in response to sporadic mutations or the acquisition of external genetic material, such as the uptake of plasmids [4]. A recent review outlined enterococcal genomics in the context of therapeutics [5]. As part of this comprehensive review, the authors discussed the epidemiology of *E. faecium* strains, reporting that *E. faecium* has divided into two distinct clades – one of clinical isolates (Clade A) and the second comprising community-related isolates (Clade B). Clade A subtype 1 *E. faecium* strains have significant genome plasticity and dominate the hospital population of disease-causing

E. faecium. It is likely that new genotypes/phenotypes will continue to emerge, adding to the challenge of controlling these infections [6].

E. faecium is now well documented to display resistance to β-lactam antibiotics, penicillin, carbapenem, ampicillin, and aminoglycosides, to name but a few [7]. *E. faecium* has even been shown to be resistant to desiccation, with more than 30% survival rates [8]. Current treatments for VRE infections include daptomycin and linezolid [9]. Linezolid is a first-line treatment for infections, but resistance to the drug is increasingly being reported [10]. Daptomycin is a last-line antibiotic used to treat multidrug-resistant *E. faecium* infections. This antibiotic has an increased effect against *Enterococcus* spp. compared to previous treatments. However, there are now reports of resistance in *Enterococcus* species to this drug. Although occurrence of this resistance is rare, it is a concerning development [11] [12]. The rise in these infections is also having a significant economic effect, with costs in Europe amounting to up to €13,157 per patient with a VRE infection [13]. This increase in cost is reported to be due to pharmaceuticals, nursing staff, medical products and assistant medical technicians.

IBD is a lifelong inflammatory illness of the gastrointestinal tract. It is an umbrella term that encompasses Crohn's disease and ulcerative colitis (UC) [14]. Crohn's disease can occur anywhere in the gastrointestinal tract, whereas UC is usually limited to the colon and rectum [15]. IBD patients experience a chronic relapse of symptoms that include abdominal pain, diarrhoea, weight loss, fatigue, and psychological stress [16] [17]. Pathogenesis of IBD is multifaceted. It encompasses a number of factors, including environmental factors, gene variants, microbiome composition, barrier function, and immune response dysregulation [15] [18]. While these factors have been widely researched, the exact pathogenesis of IBD remains unclear, and so treatment of IBD is broadly unsatisfactory. Therapeutics generally focus on the management of symptoms to improve quality of life.

In recent years, a number of studies have linked IBD to the gut microbiome. Several murine studies have investigated the effect of mutations in genes linked to IBD. However, in many cases, genetically pre-disposed mice do not contract IBD if they are raised germ-free. The development of IBD is dependent on the presence of microbes. This highlights the important dialogue between the immune system and the

microbiome [19]. One microbe that has been discussed in an IBD context is *R. gnavus*. *R. gnavus* is a gram-positive anaerobic bacterium that is found in the gut of 90% of individuals. In IBD patients, particularly those with Crohn's disease, an increase in the abundance of the bacterium correlates with a flare-up in symptoms [20] [21]. *R. gnavus* is reported to produce an inflammatory polysaccharide – a complex glucorhamnan polysaccharide with a rhamnose backbone and glucose side chains. This polysaccharide induces the secretion of TNF-α, an inflammatory cytokine, by host immune cells [22]. *R. gnavus* is also a mucin-degrading bacterium. The intestinal mucus layer of the gut is important to maintain barrier function, and the degradation of this layer can lead to bacterial invasion which can result in inflammation. As *R. gnavus* utilises mucin as a nutrient source, this is another aspect in which the transient overrepresentation of *R. gnavus* could result in the activation of an immune response [23] [24]. Antibiotics can be used for the management of IBD and, in some cases to alter microbiome composition. However, while antibiotic treatment can ease symptoms of colitis their continued use can lead to a risk of resistance and leave patients vulnerable to *C. difficile* infections [25]. To specifically modulate the microbiome of patients with minimal off-target effects, the development of new antimicrobial strategies is essential.

Bacteriophage (phage) derived enzymes such as lysins could be considered as an alternative or adjunct therapy. Lysins are classed as phage-encoded peptidoglycan hydrolases. Phages require these enzymes during infection. Along with another protein, holin, these lysins build up inside the host bacterium while the phage progeny are being assembled. When holin levels reach a certain threshold, they create pores in the cytoplasmic membrane, allowing the lysins to access the host peptidoglycan. Lysins specifically target components of the peptidoglycan structure (depending on their mechanism of action) and degrade it, leading to the ultimate lysis of the bacteria and the release of phage progeny into the surrounding environment [26]. This lytic event can be harnessed as a therapeutic by applying purified lysin exogenously and allowing it to degrade the peptidoglycan layer, particularly in gram-positive bacteria.

Several lysins targeting *Enterococcus* strains have been reported. However, the majority of these target *E. faecalis* [27] [28]. Lysins that target both strains also exist. One example is the broad-acting phage lysin PlyV12 targeting both *E. faecalis* and *E. faecium*. This lysin also targeted pathogenic group B *Streptococcus* strains,

including *S. pyogenes* [29]. Endolysins from *E. faecium* phages are rare, with only three examples currently described in the literature. One is from a prophage identified in the genome of *E. faecium* 11C6_DIV0699. This lysin, EfAmi1, was shown to have lytic activity of varying degrees against a selection of ESKAPE pathogens. Its highest activity was against *E. faecium*, but it also displayed weaker activity against *S. aureus, Acinetobacter baumannii, E. faecalis, S. pyogenes* and *C. difficile* cells [30]. Endolysin LysEFm5 was isolated from *E. faecium* phage IME-EFm5. This endolysin was shown to have high lytic activity, with zinc ion dependence, against *E. faecium* [31]. Endolysin Ply113 was isolated from an *E. faecium* phage and has displayed rapid lytic activity against *E. faecium*, *E. faecalis* and *S. aureus*. Ply113 was also shown to have anti-biofilm activity, decreasing bacterial loads in a murine peritoneal septicaemia model [32]. In terms of phage activity against *Ruminococcus*, only one case has been described. This study was published in 2023, and it reports six temperate phages that infect *R. gnavus* [33]. Currently, there are no endolysins targeting *R. gnavus* in the published literature, but the endolysins derived from these phages are described in **Chapter 4** of this thesis.

In this study, we describe the isolation of LysH1, a novel endolysin isolated from the genome of *E. faecium* phage H1 that has cross-species activity against *R. gnavus*. LysH1 is a two-domain endolysin. The N-terminal catalytic domain was predicted to have an amidase mode of action. Homology searches for the C-terminal domain indicated that this was a domain of unknown function, but through the creation of a green fluorescent fusion protein (GFP) of the C-terminal domain, we have functionally characterised this to be the cell wall binding domain of LysH1. Due to the strong lytic activity of LysH1 against *R. gnavus*, its potential as a novel therapeutic was assessed in a simplified gut community. LysH1 was shown to target *R. gnavus* JCM6515^{T} in this defined population. The community structure was also maintained in the presence of endolysin. We propose that LysH1 could be a promising candidate for the management of *R. gnavus* in patients with IBD.

Results

In Silico Analysis of LysH1.

Lysin H1 was identified in the genome of enterococcal phage H1 (accession number YP_009788405.1). LysH1 is a protein of 406 amino acids with a two-domain structure. InterPro analysis assigned LysH1 an N-acetylmuramoyl-L-alanine amidase mode of action, specifically in the protein family Amidase type 2, with a Zinc-dependent N-terminal catalytic domain [34]. HHpred analysis indicated that while the N-terminal catalytic domain of the protein showed homology to known enterococcal endolysins (99.86% probability), the C-terminal domain, which is thought to be the cell-wall binding domain, has no significant homology to the available databases [35]. The Foldseek search tool was used to search for homology to LysH1 based on protein structure [36]. The top hits returned from this search were Autolysin LytO from *Staphylococcus aureus,* N-acetylmuramoyl-L-alanine amidase of *E. faecium* and an autolysin of *Streptococcus pneumoniae*.

The predictive tool AlphaFold [37] was used to carry out structural analysis of LysH1 (**Figure 1(B)**). AlphaFold produces a per-residue confidence score for each protein (pLDDT). Much like the results of the previously mentioned tools, AlphaFold revealed a 2-domain structure of high confidence (pLDDT above 70), with a flexible linker region between the two domains. The N-terminal catalytic domain ranges from residues 1-166. Structural predictions show that it comprises two anti-parallel β-sheets flanked by α-helices. In total, there are seven α-helices in this domain, distributed around the β-sheets. The C-terminal domain (residue 175-406) appears to be comprised of two repeating regions that are made up of seven anti-parallel β-sheets. These groups of β-sheets are divided by a small α-helix (three in total). The structural prediction of LysH1, highlighting its secondary structures, can be seen in the supplementary materials (**Figure S1**).

The BLAST local alignment search tool was used to compare LysH1 to known endolysins [38]. The top 30 hits from BLASTp were brought forward to Clustal Omega to create multiple sequence alignments of LysH1 to other proteins of high homology, and this output was visualised with SeaView4 **(Figure S2)** [39] [40]. The proteins with the highest homology to LysH1 were mostly amidase domain endolysins from other *Enterococcus* phages. However, proteins from other species, including

Ruminococcus, Mediterraneibacter, and *Faecalibacterium*, are also present in these hits of the highest similarities. From the multiple sequence alignment, we can see that most similarities lie within the catalytic domains, and the C-terminal domains display less homology.

To assess the similarity of LysH1 to a larger dataset of proteins, a Sequence Similarity Network (SSN) was generated (**Figure 1 (A)**). This network assessed the homology between LysH1 and other members of the InterPro protein family database. LysH1 was blasted versus UniProt with an e-value of 1e-05, with a maximum of 1000 sequences retrieved. Sequences were then clustered using an all-vs-all blast and visualised in Cytoscape with the layout directed with the parameter %ID. In total 215 amidases were returned and used in the network. LysH1 (highlighted in yellow in **Figure 1(A)**) is most closely clustered with other proteins derived from *Enterococcus, Ruminococcus* and *Clostridium*.

Cloning and Expression of Whole Endolysin and its Putative Cell Wall Binding Domain.

Three constructs were created from LysH1. These constructs were made to functionally characterise both the N-terminal and C-terminal domains. For the C-terminal domain, two truncated forms of LysH1 were generated – LysH1 cell wall binding domain (CBD) and LysH1 CBD-GFP (**Figure 2(A)**). Each construct was designed with N-terminal 6xHis-tags for purification with affinity chromatography. The structural prediction by AlphaFold 2 of the LysH1 CBD-GFP hybrid protein can be seen in supplementary **Figure S5**.

Amplification of endolysin encoding genes by PCR resulted in products of 1221 bp for the whole endolysin LysH1 and 747 bp for LysH1-CBD. LysH1 and LysH1-CBD were cloned into the pET-28b(+) vector and were subsequently introduced into *E. coli* BL21(DE3) for recombinant protein expression (**Figure 2 (A)**). LysH1 CBD-GFP was cloned into the pHTP9 expression vector containing the GFP-fusion protein and then introduced into *E. coli* Trans-α for protein expression. All constructs were overexpressed in 3 litre preparations. Purification was carried out by affinity chromatography through the ÄKTA start protein purification system (Cytiva) with the Cytiva Talon Crude purification columns. The purity of the yield was assessed by SDS-PAGE (4-12% gradient gel) (**Figure 2(B)**). The resulting chromatogram from the ÄKTA

start Unicorn software can be seen in **Figure S3**. The size of each expressed protein, including their 6xHis-tags, are as follows: LysH1 – 48.9 kDa, LysH1 CBD – 27.6 kDa, LysH1 CBD-GFP – 54.60 kDa.

Spectrum of Activity of LysH1.

A collection of 65 strains was used to assess the spectrum of activity of LysH1, where all strains were evaluated for lytic activity by traditional spot assays. All strains were tested using this method, with the exception of *Faecalibacterium prausnitzii*, which was assessed in liquid culture due to its limited ability to grow as lawns. **Table 1** records the spectrum of activity across a variety of strains, including a selection of relevant gut members, encompassing both pathogenic and non-pathogenic strains. **Table 2** reports the spectrum of activity across a panel of VRE strains composed of both *E. faecium* and *Enterococcus avium* strains. These strains are clinical isolates from Cork University Hospital (CUH), Cork, Ireland. *E. faecium* and *E. faecalis* strains that were not obtained from CUH were tested for the presence of *vanA* and *vanB* resistance genes by PCR (**Table S1**). LysH1 displayed a narrow spectrum of activity, with lytic activity observed against only five *Enterococcus* strains at varying levels of efficiency – *E. faecium* ATCC 700221, *E. faecium* 374iF1, *E. faecium* CUH-VRE 14, *E. faecium* CUH-VRE 6, and *Enterococcus durans*. The most potent activity against any enterococcal strain was observed against *E. faecium* ATCC 700221. Strong lytic activity was also observed in spot assay against a panel of *R. gnavus* strains, including JCM56515^{T}, CCUG 57137, CCUG 54531, and PS/160. Weak activity (a hazy or weak zone) was observed against *R. gnavus* CCUG 49994 and *Ruminococcus torques* ATCC 27756.

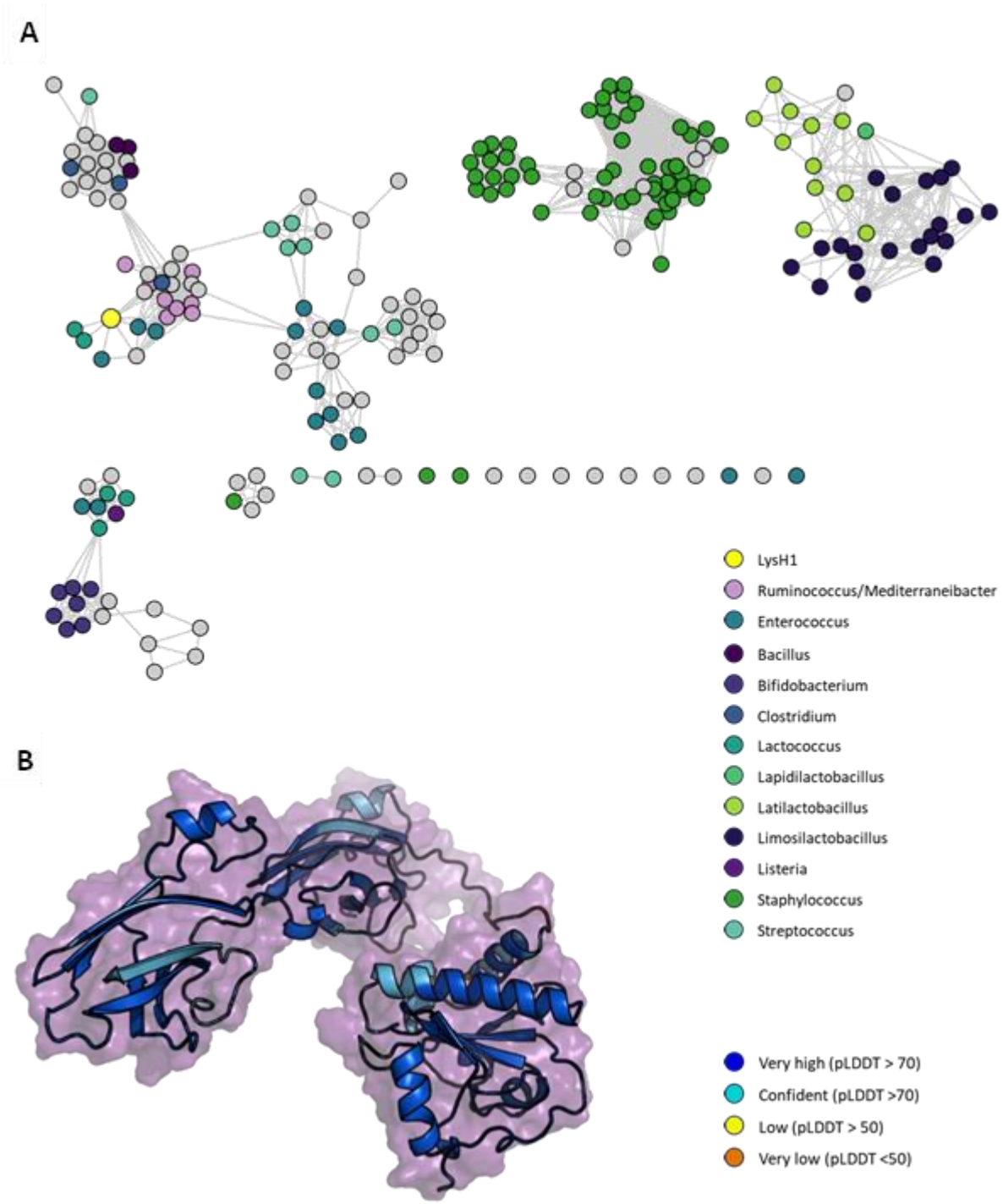


Figure 1. A) Sequence similarity network for LysH1 (yellow). **B)** AlphaFold structure prediction for LysH1. Confidence scores (pLDDT) are indicated in the legend.

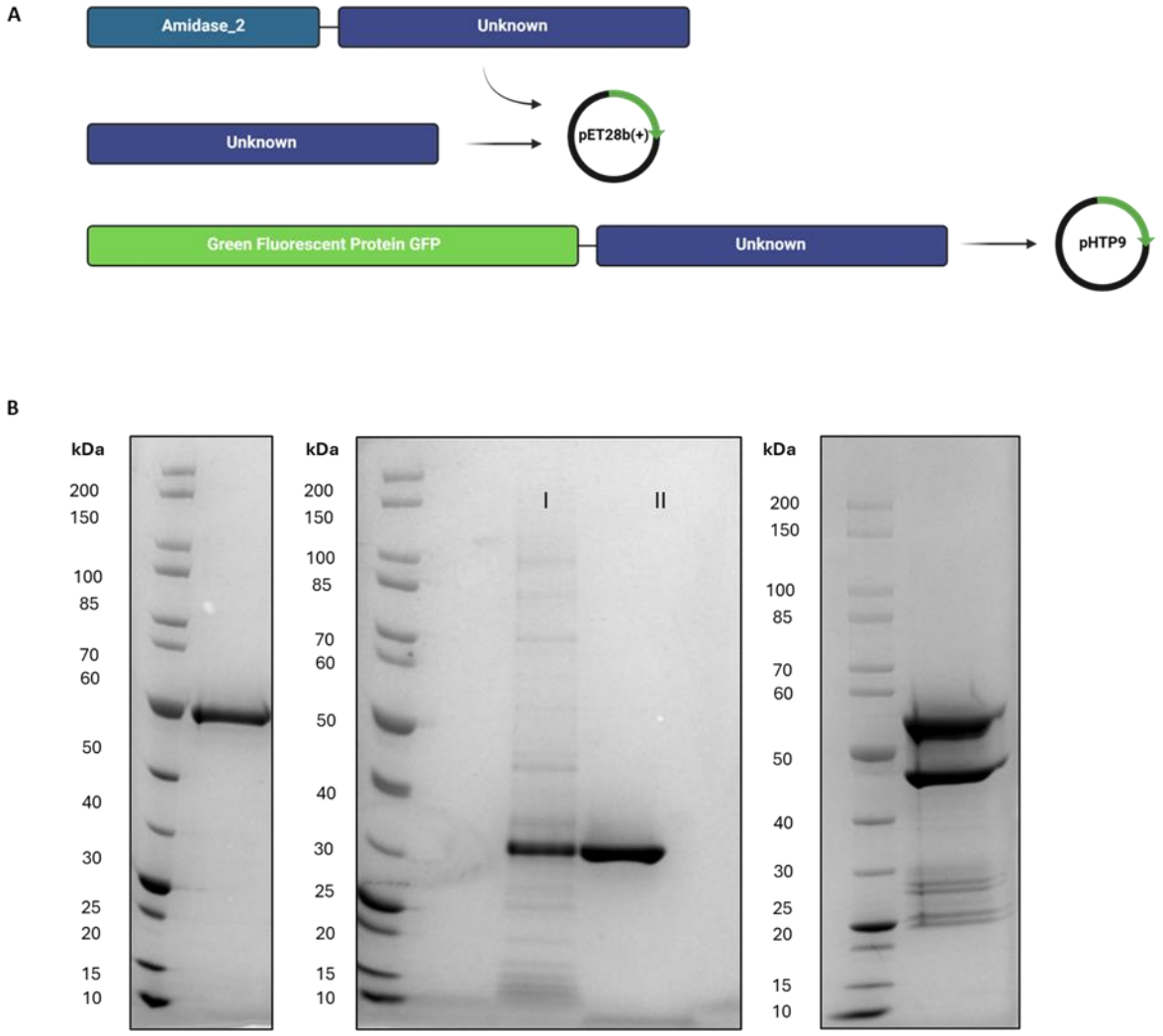


Figure 2. Constructs of LysH1 as cloned into expression vectors. A) Domains of endolysins LysH1 as assigned by *in silico* analysis and the vectors they were cloned into for recombinant expression. The putative cell wall binding domain of LysH1 is denoted by the “unknown” region. **B)** SDS-PAGE analysis of protein production. Proteins were run on a 4-12% gradient gel. Here we show purified LysH1 (left), the truncated CBD before (I) and after (II) purification (middle), and the GFP hybrid protein of CBD (right).

Table 1. Spectrum of activity of LysH1 against a number of relevant strains. - indicates no activity observed, + indicates a clear zone of lysis, and +/- indicates a weak or hazy zone of lysis.

Species	Strain	Lytic Activity
Enterococcus faecium	APC 1025	-
	APC 1031	-
	APC 1032	-
	APC 1035	-
	APC 1037	-
	APC 1038	-
	APC 1042	-
	ATCC 700221	+
	918H1	-
	374iF1	+/-
	D0 TX16	-
Enterococcus faecalis	OG1RF	-
	ATCC 29212	-
	V583	-
	APC 1309	-
Enterococcus durans	5133	+/-
Enterococcus casseliflavus	5053	-
Escherichia coli	LF82	-
Ruminococcus gnavus	JCM6515^{T}	+
	CCUG 57137	+
	CCUG 54531	+
	CCUG 49994	+/-
	CCUG 54531	-
	CCUG 57208	-
	PS/160	+
Ruminococcus torques	ATCC 27756	+/-
Blautia producta	DSM 2950	-
Staphylococcus aureus	DSM 28766	-
Streptococcus agalactiae	COH31rs	-
Listeria monocytogenes	EGDe	-
Lactiplantibacillus plantarum	WCFS1	-
Bifidobacterium longum	ATCC 15707	-
Phocaeicola vulgatus	DSM 1447	-
Faecalibacterium prausnitzii	A2-165	-
Clostridium difficile	APC 43	-

Table 2. Spectrum of activity of LysH1 against a number of VRE strains isolated from patients at Cork University Hospital (CUH). "-" indicates no activity observed, "+" indicates a clear zone of lysis, and "+/-" indicates a weak or hazy zone of lysis.

Species	Strain Number	Lytic Activity
Enterococcus faecium	CUH-VRE 1	-
	CUH-VRE 2	-
	CUH-VRE 3	-
	CUH-VRE 4	-
	CUH-VRE 5	-
	CUH-VRE 6	+/-
	CUH-VRE 7	-
	CUH-VRE 8	-
	CUH-VRE 9	-
	CUH-VRE 10	-
	CUH-VRE 11	-
	CUH-VRE 12	-
	CUH-VRE 13	-
	CUH-VRE 14	+/-
	CUH-VRE 15	-
	CUH-VRE 16	-
	CUH-VRE 17	-
	CUH-VRE 18	-
	CUH-VRE 21	-
	CUH-VRE 22	-
	CUH-VRE 23	-
	CUH-VRE 24	-
	CUH-VRE 25	-
	CUH-VRE 26	-
	CUH-VRE 27	-
	CUH-VRE 28	-
Enterococcus avium	CUH-VRE 19	-
	CUH-VRE 20	-
	CUH-VRE 29	-
	CUH-VRE 30	-

Lytic Activity of LysH1.

To further assess the therapeutic potential of LysH1, turbidity reduction tests and kill curves were performed against two candidate strains, *E. faecium* ATCC 700221 and *R. gnavus* JCM6515^T. The ability of LysH1 to lyse inactivated stationary phase cultures can be seen in **Figure 3 (A)** and **(B)**. Inactivated cells treated with LysH1 at varying final concentrations were incubated for 120 minutes at 37 ^oC, with an optical density reading taken every 15 minutes. In the case of both candidates, LysH1 was capable of halving the optical density within two hours of its addition. LysH1 activity was more efficient against *R. gnavus* JCM6515^T, reaching this level of reduction in under 60 minutes.

LysH1 is also capable of acting against actively growing cells **(Figure 3 (C-F))**. The strains were incubated for 24 hours at 37 ^oC, and an optical density reading was taken every 30 minutes. A 1% inoculum of each strain was added to media containing either a control buffer or lysH1 at a range of concentrations (**Figure 3 (C, D)**). *E. faecium* ATCC 700221 appears to follow a trend in response to treatment with LysH1. The bacteria proliferate for up to ~ 6 hours in the presence of endolysin, following which the optical density decreases. It is interesting that this lytic response does not seem to occur until the cells have reached a certain growth stage, regardless of the concentration of endolysin used. Once this density is reached, lysis occurs in a dose-dependent manner. The opposite observation occurs in *R. gnavus* JCM6515^T, where we see that the growth of this bacterium is completely inhibited in the presence of endolysin.

Next, log phase cultures of each strain were added to media containing either a control buffer or lysH1 at a range of concentrations (**Figure 3 (E, F)**). Endolysin treatment was applied to *E. faecium* ATCC 700221 cells (**Figure 3 (E)**). These cells were left to grow to an OD_{600} of ~0.6 before being treated with LysH1 at a range of concentrations. The lysis of cells occurs in a dose-dependent manner, with the greatest lytic activity recorded at 100 µg/mL. When the 24-hour incubation was complete, cells were taken from the wells and cultured on solid agar. The following day, growth was observed on these plates, indicating that the endolysin had not completely killed the culture. To ensure these cells were not LysH1-resistant mutants, spot assays were repeated using these colonies as an inoculum, but lysis was still

observed. Lytic activity of LysH1 against log phase *R gnavus* JCM6515^T was also assessed (**Figure 3 (F)**). Cells were grown to an OD_{600} of ~0.4 and were then treated with LysH1 at a range of concentrations. At every concentration tested, there was a clear reduction in optical density in as little as 1 hour, regardless of the dose applied. In a further assay, the minimum inhibitory concentration (MIC) of LysH1 against *R. gnavus* JCM6515^T was determined to be 1.5 µg/mL.

The lytic activity of LysH1 was examined against a second strain of *E. faecium*, 374iF1. This strain was isolated from a healthy human faecal sample in our laboratory. In spot assays, LysH1 only produced a weak and hazy zone against *E. faecium* 374iF1. Turbidity reduction assays against inactivated cells showed LysH1 having strong lytic activity, with a reduction in optical density within 1 hour (**Figure S4 (A)**). Lytic activity was also tested against actively growing cells, with endolysin-containing medium inoculated with 1% *E. faecium* 374iF1. Cells were incubated for 24 hours at 37 ^oC. *E. faecium* 374iF1 mirrored the trend of strain ATCC 700221, where growth occurred to a specific optical density (in this case OD_{600} ~0.4) before lytic activity was observed (**Figure S4 (B)**).

Biochemical Characterisation of LysH1.

LysH1 was subjected to a range of conditions to assess its stability at different temperatures and in different buffer compositions. These assays were carried out with *E. faecium* ATCC 700221 as the target bacterium. In **Figure 4 (A)** and **(B)**, optimum pH was determined. The Expasy “compute pI” tool [41] was used to obtain the theoretical isoelectric point (pI) for LysH1, which was 9.42. First, cells of *E. faecium* were exposed to the buffers at a varying range of pH (2-11) as a control assay to assess whether pH alone could have a lytic effect against *E. faecium* **Figure 4 (A)**. From this, we can see a slight lytic effect at pH levels 9 and 10. Then, cells of *E. faecium* were washed and resuspended in endolysin-containing buffer at a range of pH and at a lysin concentration of 100 µg/mL (**Figure S4 (B)**). From this, we can determine that LysH1 is more effective in alkaline buffers, with the best lytic activity recorded at pH 8 - 11.

The remaining tests focused on temperature stability and NaCl concentrations. LysH1 was exposed to a range of temperatures (4 ^oC to 90 ^oC) for 1 hour before application to *E. faecium* cells at a concentration of 100 µg/mL. Lytic activity was

retained from 4 to 60 ^{o}C, with loss of function observed from 75 ^{o}C and above (**Figure 4 (C)**). To assess the optimum salt concentration for LysH1, cells of *E. faecium* were resuspended in endolysin-containing buffer at varying concentrations of NaCl (0 mM to 200 mM). The lytic activity of LysH1 was equal across all concentrations **(Figure 4 (D)**). From these assays, we determined that LysH1 should be stored in a buffer of pH 8 and 50 mM NaCl for optimum activity.

Fluorescent Binding Assays of LysH1 CBD-GFP.

To assign a function to the putative cell wall binding domain at the C-terminal of LysH1 (LysH1 CBD), a hybrid GFP C-terminal domain was created (LysH1 CBD-GFP). Binding assays with LysH1 CBD-GFP were performed on a panel of strains (**Figure 5 (A)**). The GFP-fusion protein was incubated with the test bacteria at a final concentration of 50 µg/mL for 45 minutes, shaking. The cells were harvested and washed three times to remove any unbound GFP. Binding was assessed by fluorescent microscopy. This assay confirmed that the C-terminal domain of LysH1 is involved in binding. Strains that had previously shown to be strongly acted upon by LysH1, *E. faecium* ATCC 700221 and *R. gnavus* JCM6515^{T} are seen to be brightly decorated with GFP. LysH1 has only weak activity against *E. faecium* 374iF1, and this is reflected in a weaker binding assay. Cells of this strain show a low efficiency of binding, which indicates why activity is not as potent against this target. Strains that had shown no lysis in response to LysH1 treatment, *E. faecalis* OG1RF and S*. aureus* DSM 28766, did not display any fluorescence.

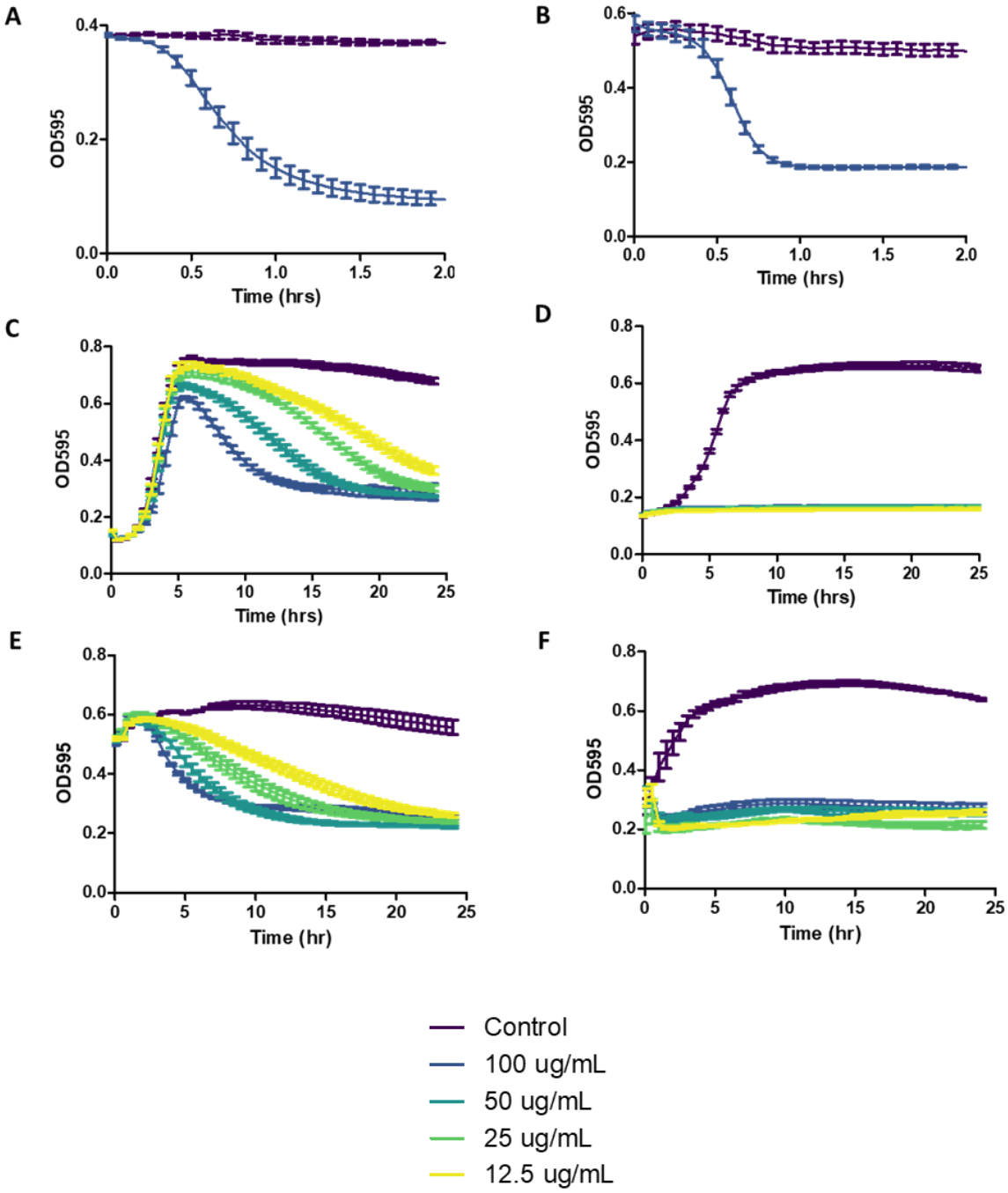


Figure 3. Lytic activity of LysH1 against *E. faecium* and *R. gnavus*. A) Turbidity reduction assay of LysH1 versus *E. faecium* ATCC 700221. **B)** Turbidity reduction assay of LysH1 versus *R. gnavus* JCM6515^{T}. **C)** *E. faecium* ATCC 700221 growth curve in LysH1-containing media at varying concentrations. **D**) *R. gnavus* JCM6515^{T} growth curve in LysH1-containing media at varying concentrations. **E)** Treatment of log phase culture *E. faecium* ATCC 700221 with LysH1. **F)** Treatment of log phase culture *R. gnavus* JCM6515^{T} with LysH1. All assays had a negative control (Control) treatment of buffer alone.

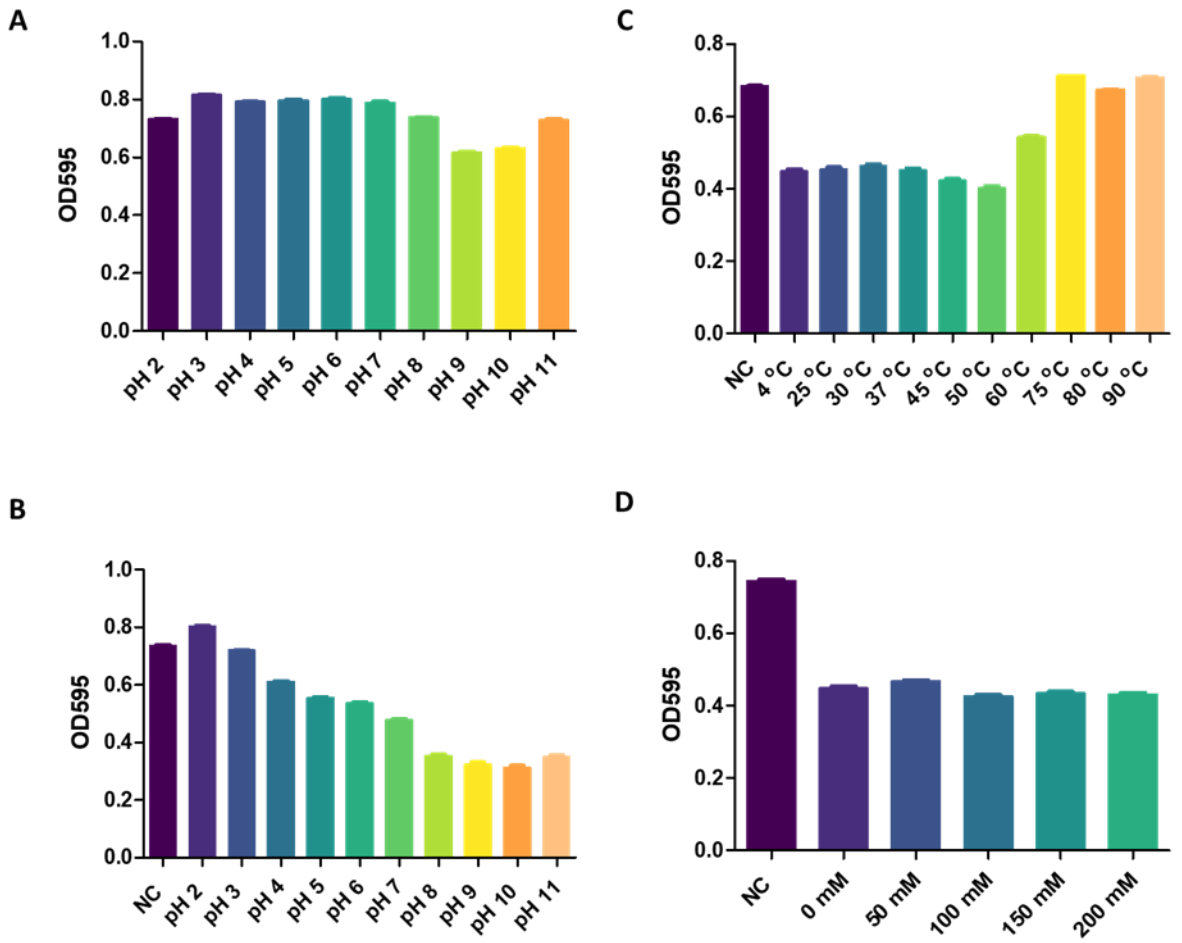


Figure 4. Biochemical characterisation of LysH1. These graphs are based on a turbidity reduction method where *E. faecium* ATCC 700221 was treated with a final concentration of 100 µg/mL LysH1. **A)** Control assay for pH against *E. faecium* ATCC 700221. This was to test the lytic effect of pH against the strain in the absence of endolysin. **B)** The activity of LysH1 at a range of pHs. **C)** The activity of LysH1 after exposure to a range of temperatures. **D)** The activity of LysH1 in a range of salt concentrations.

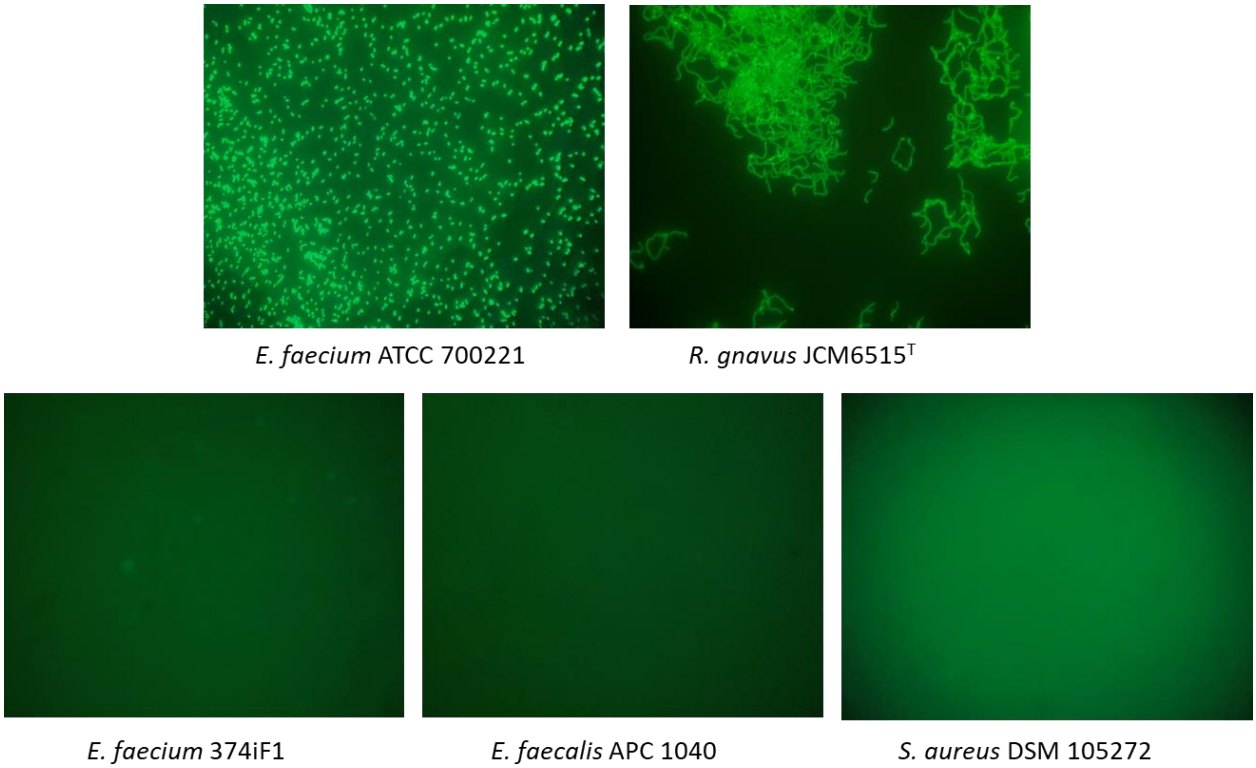


Figure 5. Binding assays of LysH1-CBD-GFP. The GFP-fusion protein of LysH1 CBD facilitated binding assays against a panel of strains. The purified protein was added to strains at a final concentration of 50 µg/mL. Fluorescent microscopy analysis showed that LysH1 binds with high affinity to strains with high lytic activity. Strains with weak lytic activity displayed reduced binding affinity.

The Effect of LysH1 on a Simplified Human Gut Community.

A simplified human intestinal microbiota (SIHUMI) was established to assess the effect of endolysin treatment on the structure of a bacterial community. This community consists of seven strains – *Escherichia coli* LF82, *E. faecalis* OG1RF, *Lactiplantibacillus plantarum* WCFS1, *Phocaeicola vulgatus* DSM 1447*, Bifidobacterium longum* ATCC 15707, *Faecalibacterium prausnitzii* A2-165 and *R. gnavus* JCM6515^{T}. The target strain for the LysH1 treatment is *R. gnavus* JCM6515^{T}. Host range data has previously shown that LysH1 acts specifically against *R. gnavus* in this defined cohort.

These strains were inoculated into the broth and allowed to grow for six hours before intervention with endolysin treatment. The consortium was split into a test group that received LysH1 at a concentration of 50 µg/mL and a control group that received the same volume of buffer. The culture mix was grown for up to 48 hours under strict anaerobic conditions. Samples of 1 mL were collected at set time points (0h, 6h, 24h, 48h) for DNA extractions. Quantification of genome copies per mL of each member in the community was determined by qPCR with strain-specific primers (supplementary **Table S2**).

The impact of LysH1 on the defined bacterial community can be seen in **Figure 6 (A)** and **(B).** The bacteria community structure in the absence of endolysin is shown in **Figure 6 (A)**. Here, we can see that the more dominant members of the SIHUMI, *E. coli*, *E. faecalis* and *L. plantarum*, proliferate relatively quickly. These dominant strains reach ~9 log genome copies per mL. *B. longum, R. gnavus*, *F. prausnitzii* and *P. vulgatus* rise at a slower rate and remain at lower levels (between ~ 4 – 6 log genome copies per mL). The order of the members remained stable for the 48-hour incubation period. The SIHUMI with endolysin treatment is displayed in **Figure 6 (B)**. With treatment of LysH1 at the 6-hour timepoint, the community structure reflects that of the control group and remains stable for the duration of the incubation.

To verify that LysH1 was active against *R. gnavus* JCM515^{T} and that changes could be quantified by qPCR, a monoculture of *R. gnavus* was established in the same way as the SIHUMI. *R gnavus* monoculture was split into a control group and a test group at the 6-hour timepoint. The control was supplemented with buffer, and the test was treated with 50 µg/mL of LysH1 (**Figure 6 (C)**). The genome copy number per mL

of *R. gnavus* in a monoculture setting was higher than that of the levels seen in the SIHUMI, peaking at ~9 log genome copies in the control group. In the presence of LysH1, *R. gnavus* genome copy numbers do not increase – indicating that the growth of the bacterium was inhibited. It is clear that LysH1 effectively targets *R. gnavus*, but we do not observe an obvious change in numbers in the community setting. This may be a limitation of qPCR, as it does not differentiate between living and dead cells. Also, as *R. gnavus* exists as a less dominant member of the defined community, we may not be able to track cell death as the log number of copies does not reach as high a level as can be seen in the monoculture. Instead, the qPCR analysis may be showing us the log copy number that previously existed in the sample before the addition of lysin, and in a setting where the growth of *R. gnavus* is already suppressed, the lytic action of the lysin may not be visible.

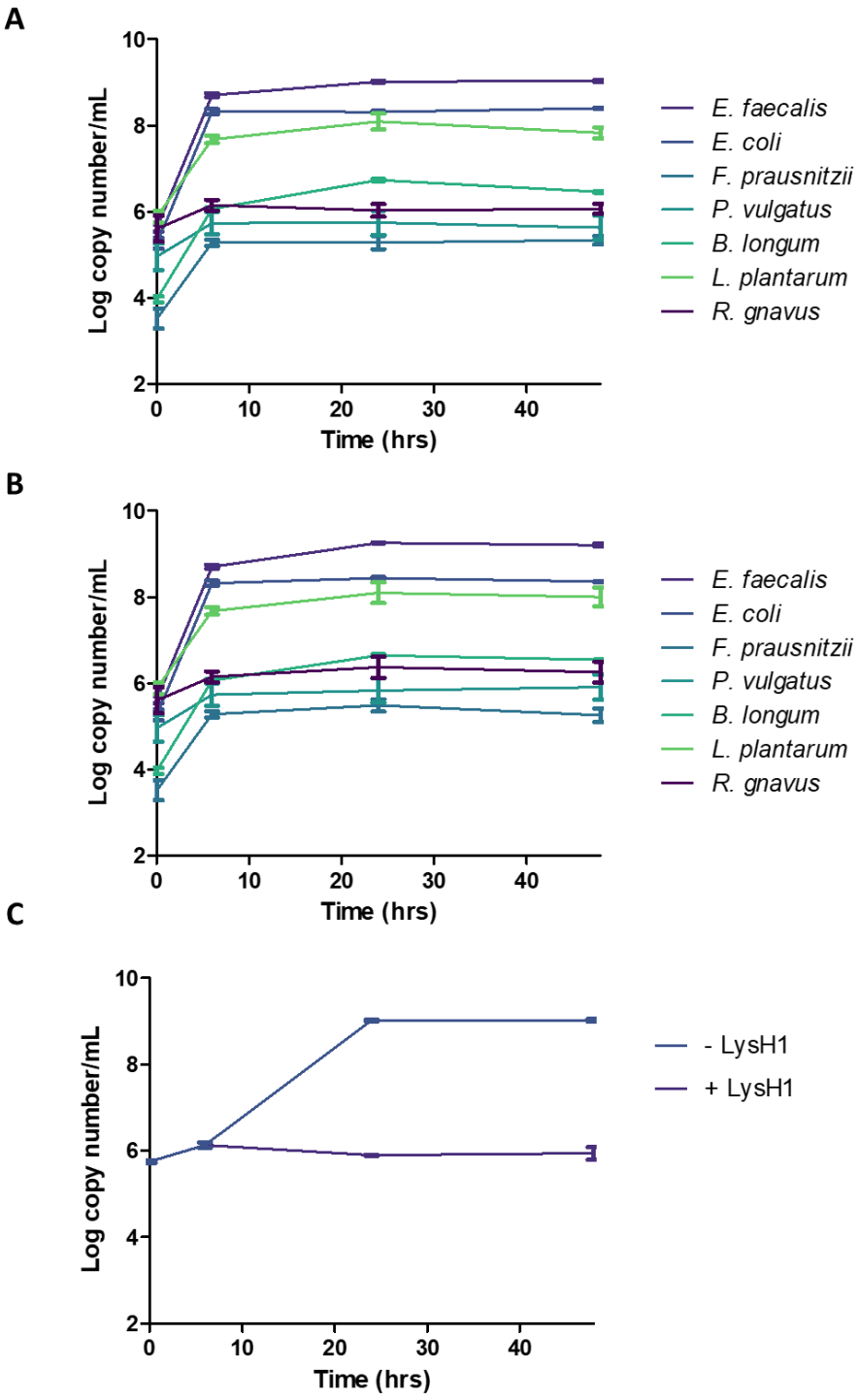


Figure 6. Impact of LysH1 on a simplified gut community. A) SIHUMI consortium in the absence of endolysin. **B)** SIHUMI consortium with LysH1 added at T6, at a final concentration of 50 µg/mL. **C)** *R. gnavus* JCM6515^{T} monoculture in the absence of endolysin, or with LysH1 added at T6, at a final concentration of 50 µg/mL.

Discussion

The growing threat of antibiotic resistance means that the need for alternative therapeutics is of the utmost importance. *E. faecium* is associated with life-threatening infections such as bacteraemia and endocarditis. This microbe has evolved rapidly over the past decades to acquire resistance mechanisms to a range of antibiotics, making it a complicated bacterium to target [5]. *R. gnavus* is a bacterium that, in recent years, has been reported to have correlations with IBD. Blooms of *R. gnavus* in the gut of IBD patients have been linked with a flare-up of symptoms [42]. *R. gnavus* is sometimes targeted by antibiotics to control these flares. As IBD is a lifelong illness, the continued use of antibiotics for the management of symptoms could lead to a rise in resistance levels [25].

In this study, a new endolysin, LysH1, from *E. faecium* phage H1 was identified and cloned into *E. coli* for recombinant expression. Furthermore, the cell wall binding domain of this endolysin was functionally characterised through the use of a GFP-hybrid version of the protein that exhibited strong binding to its target strains. LysH1 was initially produced to target *E. faecium*. However, host range analysis of 65 strains revealed that the endolysin has cross-species activity targeting *R. gnavus*. Interestingly, LysH1 displayed more efficient lytic activity against this species. To assess the potential therapeutic applications of LysH1, we investigated its effect against *R. gnavus* in a simplified human gut consortium. Further analysis should be carried to confirm that LysH1 did not disrupt the community structure, but initial findings indicating that it has therapeutic potential for the management of IBD.

Structural and *in silico* analysis of LysH1 revealed a two-domain structure consisting of an N-terminal catalytic domain with an N-acetylmuramoyl-L-alanine amidase mode of action and an uncharacterised C-terminal domain. To assign function to this previously uncharacterised domain, we created a truncated form of LysH1 comprising the C-terminal domain fused to a green fluorescent protein, LysH1 CBD-GFP. Binding assays assessed by fluorescent microscopy confirmed that this domain was, in fact, the cell wall binding domain of LysH1. It was shown to bind specifically to bacteria that were shown to be targets for lytic activity. The authors of EfAmi1 lysin hypothesised that the C-terminal domain of the endolysin was involved in carbohydrate binding from *in silico* predictions [30]. However, this is the first

experimental evidence to functionally characterise the cell wall binding domain from an *E. faecium* phage endolysin.

Assessing the lytic activity of LysH1 has led to a number of interesting observations. This lysin was initially isolated to target *E. faecium*. Host range analysis showed that activity towards *E. faecium* strains was very narrow – with lytic activity only recorded in four *E. faecium* strains of 39 tested. In these four strains, only one was lysed efficiently, ATCC 700221. *E. faecium* ATCC 700221 is a VRE strain positive for the plasmid-borne *vanA*-type resistance [43]. Binding assays with LysH1 CBD-GFP showed that LysH1 has strong binding to *E. faecium* ATCC 700221 but reduced binding to *E. faecium* 374iF1, which was only weakly lysed. This strongly suggests that the reason for having limited lytic activity of LysH1 against this strain is due to poor binding efficiency. Unlike other characterised *E. faecium* endolysins, LysH1 did not display any lytic activity against a panel of *E. faecalis* strains but had weak activity against *E. durans*.

The lytic activity of LysH1 was further characterised against *E. faecium* ATCC 700221 to assess the potential of LysH1 as a therapeutic for VRE infections. A turbidity reduction assay using inactivated *E. faecium* cells showed promising lytic action. The optical density of *E. faecium* cells was rapidly reduced. However, lytic activity against actively growing cells displayed a different pattern. In both a 1% inoculum of cells and log-phase culture, the density of cells rises to a certain level before lysis occurs. This is an unusual observation, but one such reason for this interesting pattern could be the structural similarities between LysH1 and autolysins. Foldseek analysis, which is based on 3D structure, reported that LysH1 had the greatest structural homology to autolysins of *Staphylococcus* and *Streptococcus* in its N-terminal domain. LytA is a major autolysin (and an amidase) of *Streptococcus pneumoniae*. It has been shown that the cell wall is only susceptible to extracellular LytA when cells reach the stationary phase or when wall synthesis is inhibited [44]. Growth profiles of *S. pneumoniae* in the presence of LytA resemble that of *E. faecium* in the presence of LysH1. Perhaps the lytic activity of LysH1 is similarly dependent on the growth phase, which is why we are seeing this unusual lysis pattern. It is also possible that the activity of LysH1 against *E. faecium* is simply not effective enough to kill the actively dividing bacteria. The antimicrobial action of LysH1 against *E. faecium,* while interesting, is not favourable for a potential therapeutic. When endolysin kill curve assays were completed, the wells

of the 96-well plate were streaked on fresh media to assess the survival of *E. faecium* post-treatment. This resulted in strong re-growth of the bacteria, highlighting that LysH1 (at least alone) may not be suitable for the management of *E. faecium* infections. Future work could be carried out to look for synergy between LysH1 and other antimicrobials to effectively target *E. faecium*. An example of such a strategy is the enterococcal lysin Ply2660 and its synergy with antimicrobial peptide cathelicidin LL-37 [45]. The combination of these antimicrobial agents against *E. faecalis* gave a greater bactericidal effect than either one alone.

Endolysins are generally favoured for their specificity. However, we have demonstrated that LysH1 is an endolysin that has a lytic effect on its natural target *E. faecium*, acting on both antibiotic-sensitive *E. faecium* and vancomycin-resistant strains, but also has the ability to lyse *R. gnavus*. Other reported endolysins that target *E. faecium* have displayed extended host ranges, spanning species-specific to cross-species activity. *E. faecium* endolysin LysEFm5 has lytic activity across a panel of 23 *E. faecium* strains, but no antibacterial activity was detected against other species [31]. On the other hand, Endolysin Ply113, isolated from *E. faecium* phage, displayed varying levels of activity across *E. faecium, E. faecalis, C. difficile* and *S. aureus* [30]. Cross-species activity against *Ruminococcus* spp. has not previously been reported. In order to understand this cross-species activity, bioinformatic analysis was carried out to search for homology between LysH1 and other available database proteins. A sequence similarity network was generated that showed a number of *Ruminococcus* enzymes that had shared nucleotide level similarity to LysH1.

From the panel of *Ruminococcus* strains tested, four strains of *R. gnavus* and one strain of *R. torques* were lysed by LysH1. The highest lytic activity was against *R. gnavus* JCM6515^T. *In vitro* analysis allowed us to investigate this interaction further. The creation of purified LysH1 CBD-GFP facilitated binding assays that showed LysH1 can bind with high efficiency to *R. gnavus* JCM6515^T. Lytic activity tests with this strain were more promising than those of *E. faecium* strains. LysH1 displayed rapid and effective lytic activity against inactivated and log-phase cells. It also was efficient in the inhibition of growth. The MIC of LysH1 against *R. gnavus* JCM6515^T was determined to be 1.5 µg/mL. These traits combined indicated that LysH1 could be a promising candidate for the management of *R. gnavus* JCM6515^T.

Several studies have assessed the impact of endolysin application on microbiome structure [46] [47] [48]. To investigate the effect of LysH1 on a defined community, we established a simplified human intestinal microbiota (SIHUMI). The SIHUMI is a well-defined human microbiome consortium that has been used in previous studies as a simplified proxy for the gut microbiome [49] [50] [51]. This consortium of seven key gut members provided a simplified model for the activity of LysH1. Prior to setting up the community, the strains were tested individually for LysH1 sensitivity. We established that LysH1 only had lytic activity against *R. gnavus* JCM6515^T, but the assessment of activity in the community setting allowed us to investigate the population dynamics of specifically targeting one strain (**Figure 6**). As presented in the results section, LysH1 did not have an effect on population structure. However, we also did not see a change in the levels of *R. gnavus* genome copies in the test group versus the control group. This may be a limitation of qPCR, combined with the population dynamics of the SIHUMI. A 2023 study employed the SIHUMI as a model to test bacteriocin-producing strains [52]. In this study, the authors investigated antagonistic interactions between SIHUMI members. It was reported that *R. gnavus* was inhibited by the more dominant members of the mix, including *E. coli*, *E. faecalis* and *L. plantarum*. We hypothesise that there is no change in *R. gnavus* levels due to the strain already being suppressed by others in the population. When we tracked the log copy number of *R. gnavus* JCM6515^T in a monoculture, we could see that LysH1 treatment can effectively inhibit the growth of the strain.

The SIHUMI represents a “normal” human gut. As previously mentioned, the microbiome of IBD patients is different, with certain species like *R. gnavus* being over-represented [20]. The transient blooms of *R. gnavus* correlate with increased inflammatory symptoms in patients. In our simplified intestinal consortium, LysH1 has been shown to not negatively affect population dynamics but can also control a high level of *R. gnavus* growth (as seen in a monoculture). LysH1 could, therefore, be a promising novel antimicrobial that could specifically target the blooms of *R. gnavus* growth observed in IBD patients. Currently, the management of IBD includes methods to control the microbial population. Metronidazole has been shown to improve disease activity in IBD patients, suggesting that the management of microbes through antimicrobials could play a role in therapy [53]. Probiotic therapy was assessed to prevent a flare-up of pouchitis. This is inflammation that occurs in the lining of a pouch

that is created during a total colectomy for the treatment of UC [22]. In a 2003 study, only 10% of patients treated with a probiotic preparation (VSL#3) had an episode of inflammation versus 40% of the group that received a placebo [54]. Studies such as these highlight the potential of microbiome modulation in the long-term management of IBD. Going forward, it would be interesting to investigate the effect of LysH1 on a more complex population. Faecal fermentations using samples from IBD patients would truly test the ability of this endolysin to act on blooms of *R. gnavus* with minimal collateral damage.

When assessing the potential of a new therapeutic, it is important to ensure the drug is specifically targeting the microbe of choice. There is ample evidence documenting how broad-spectrum antimicrobials can have a range of off-target effects and disrupt the balance of the microbial community. Endolysins have been reinvestigated as therapeutics in recent years due to their specificity, lack of resistance mechanisms and fast-acting nature [55]. However, while LysH1 was isolated for the eradication of *E. faecium*, this study suggests it may in fact be more effective as a tool to control *R. gnavus*, another member of the human gut microbiome. While this trait has proven to be advantageous for this study, it does signify the importance of assessing therapeutic candidates against a wide range of microbes before proceeding with drug development.

Materials and Methods

In silico Analysis of LysH1.

Homology searches were performed for LysH1 in HHpred [35] and InterPro Scan [34]. Multiple sequence alignments were generated by MAFFT taking the top 20 hits of the sequence similarity network. This alignment was visualised using the SeaView4 software. For structural analysis of endolysins and cell wall binding domains, AlphaFold v2.3.1 was used to predict the structures with the settings "--enable_gpu_relax=true --model_preset=monomer --use_gpu=true" [37]. Resulting PDB files were visualised using PyMOL.

Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) was used to create a sequence similarity network (SSN) [56]. LysH1 was blasted vs UniProt DB with an E-value of 1e-05 with maximum sequence length of 750. The full network was downloaded and visualised in Cytoscape v3.10.1 with the "perfuse force directed layout" on the %ID value. Taxonomy was appended to each protein using TaxonomizR v0.10.5 and R4.3.0 to pull taxonomic ranks from NCBI.

Bacterial Strains and Growth Conditions.

All strains used in this study were stored at -80 ^{o}C. *Enterococcus* strains were cultured in Tryptic Soy Broth (TSB) (Merck). *R. gnavus* strains were cultured in anaerobe basal broth (ABB) (Oxoid). *Escherichia coli* strains used for cloning were grown in Laura-Bertani (LB) broth (Fischer Scientific), supplemented with 50 µg/mL Kanamycin (Sigma Aldrich) for maintenance of plasmids. *Enterococcus faecium* 374iF1 was isolated in the lab from healthy human faecal samples. The remaining strains were obtained from the APC Culture Collection (Moorepark, Fermoy, Co. Cork, Ireland). Vancomycin-resistant *Enterococcus* strains were isolated and obtained from Cork University Hospital. All aerobic strains were cultivated by incubating at 37 ^{o}C, and shaking at 150 rpm, while anaerobic strains were grown at 37 ^{o}C under strict anaerobic conditions (Type A vinyl anaerobic chamber, Coy Labs).

PCR to assess the presence of *VanA* or *VanB* resistance genes in enterococcal strains was carried out using MyTaq^{TM} Red Mix (Bioline) with primers adapted from Domingo et al, 2005 [57]. The primer sequences are as follows: VanA (F –

AATAGCGCGGACGAATTGGAC, R – AACGCGGCACTGTTTCCCAA) and VanB (F – CTTAACGCTGCGATAGAAGC, R – CTGATGGATGCGGAAGATAC).

The simplified gut consortium was cultured in a modified Brain Heart Infusion (Oxoid) medium with the addition of 0.5% yeast extract, 5 mg/L hemin (Sigma–Aldrich), 1 mg/ml cellobiose (Sigma–Aldrich), 1 mg/ml maltose (Sigma–Aldrich), and 0.5 mg/ml L-cysteine (Sigma–Aldrich) – LYHBHI as described by Soko et al, 2008 [58]. The strains used in this community included *E. coli* LF82, *E. faecalis* OG1RF, *Lactiplantibacillus plantarum* WCFS1, *Faecalibacterium prausnitzii* A2-165, *Phocaeicola vulgatus* DSM1447 and *R. gnavus* JCM6515^{T}.

Cloning of Endolysin and Cell Wall Binding Domain.

Putative lysin genes in the phage genome were identified using BLASTP from NCBI [59]. The amino acid sequence of this putative lysin, LysH1, was further analysed using InterProScan [60] and HHpred [61]. The endolysins isolated in this study can be found at GenBank, accession number UYE92572.1 The whole lysin gene was amplified from a phage lysate of *E. faecium* phage H1. The authors would like to note that, unfortunately, the lysate of this phage was lost in storage, and as a consequence, we are no longer in possession of this phage. Amplification was carried out with KOD Hot Start DNA Polymerase (Merck). The primers used for amplification of LysH1 are as follows with restriction sites underlined: forward primer 5'-GTGCTCGAGttataatctataagttccccacag–3', reverse primer 5'–GGATCCGatgactcgtttttataa – 3' . The PCR products were cleaned and concentrated using Zymo Research Clean and Concentrate DNA Kit (Zymo Research). The PCR product and the plasmid pET28b(+) were digested using restriction enzymes XhoI and BamHI (Thermo Scientific) and ligated with T4 ligase (Rapid Ligation Kit, Thermo Scientific) to create the construct pET28b(+)LysH1. This construct contained the lysin gene LysH1 with 6xHis-tags attached at the N-terminal of the protein. The construct was introduced into One Shot Top10 chemically competent cells (Invitrogen). Transformants were screened for positive clones, which were propagated for plasmid extractions. Extracted plasmids positive for the insert were introduced into One Shot BL21(DE3) chemically competent cells (Invitrogen) for protein expression. Sanger sequencing confirmed the integrity of the construct pET28b(+) LysH1.

Cell wall binding domain GFP constructs were created using NZYEasy Cloning & Expression Kit IX (NZYtech), which created a construct in plasmid pHTP9 that contains GFP as a fusion tag. This construct was introduced into *E. coli* Trans-alpha competent cells. Transformants were screened for positive clones, which were propagated for plasmid extractions. Extracted plasmids positive for the insert were introduced into One Shot BL21(DE3) chemically competent cells (Invitrogen) for protein expression. Sanger sequencing confirmed the integrity of the construct.

Recombinant Expression and Purification .

E. coli BL21(DE3) cells containing each of the constructs were overexpressed in the following manner. Three litres of cells were prepared in LB broth supplemented with 50 µg/mL kanamycin and incubated at 37 ^{o}C at 150 rpm shaking. Cells were grown to an OD 0.6 and expression was induced by the addition of 1 mM final concentration IPTG (Sigma Aldrich). Cultures were left to shake overnight at room temperature and were harvested the following day. Cells were pelleted at 4,000 xg for 30 minutes. Cell pellets were lysed using sonication on ice, pelleted again by centrifugation and the supernatant was extracted for protein purification.

Purification of proteins was facilitated by the Akta Start Purification System (Cytiva), using the Talon Crude Purification 5mL columns (Cytiva). Purification fractions were assessed for quality and purity by SDS-PAGE (4-12% gradient gel) and imaged with iBright500. Fractions containing the protein of interest were pooled, resuspended in a final concentration of 20% glycerol, and stored at -20 ^{o}C for short-term storage, and -80^{o}C for extended storage.

Lytic Activity of LysH1.

The lytic abilities of LysH1 were assessed by host range analysis, turbidity reduction assays, kill curves and counts to assess log reduction of bacteria after exposure to lysin.

To determine the host range of LysH1, 100 µL of an overnight culture was added to 4 mL of growth medium containing 0.4% agarose and poured over an agar plate of 1.5% solid agarose. The overlay was left to set and a 10 µL spot of the purified lysin was spotted on the solidified agar. The plate was incubated overnight at 37 ^{o}C and checked the following day for evidence of lytic activity. An exception to this method

of assessing lytic activity was *F. prausnitzii*, as it does not grow well in semi-solid agar. For this strain 200µL of an overnight culture was spun down at 3,000 x g for 5 minutes and the cell pellet was washed three times with PBS. Activity was assessed by adding 20 µL of purified protein or buffer to 180 µL of cells, and observing a reduction in optical density of the cells versus the control.

Turbidity reduction assays were performed to assess lytic activity of LysH1 against *E. faecium* 374iF1, *E. faecium* ATCC 700221 and *R. gnavus* JCM6515^T. A culture of the test strain was grown in liquid medium to an OD_{600} of ~ 1.0. Aerobic cells were heat-inactivated, and anaerobic cells were inactivated by exposure to oxygen. The cells were pelleted by centrifugation at 3000 x g for 5 minutes. The cell pellet was washed three times in PBS before being resuspended to an OD_{600} of 0.8. A 96-well plate was prepared to 200 µL final volume of cell suspension with the addition of 100 µg/mL purified LysH1, or a negative control of buffer. Changes in optical density were recorded every 5 minutes over the course of 2 hours at 37 ^oC using a microplate reader (Thermo Fischer).

Kill curves of LysH1 were performed for *E. faecium* ATCC 700221 and *R. gnavus* JCM6515T. Cells were sub-cultured from overnight cultures into fresh media at 1% inoculum. In a 96-well plate, a 1% inoculum of cells was treated with purified lysin at various concentrations (100, 50, 25, 12.5 µg/mL) or control buffer. Plates were incubated in a microplate reader (Thermo Fischer) at 37 ^oC aerobically/anaerobically for up to 48 hours. Changes in optical density were recorded every 30 minutes at a wavelength of 595nm. For treatment of log-phase bacteria, a similar method was employed with a culture that was grown to an OD_{600} of 0.6.

Binding Assays and Fluorescent Microscopy.

The binding of LysH1 CBD to a panel of strains was demonstrated through the following methods. Overnight cultures of bacterial candidates were prepared. Cells were harvested by centrifugation at 3,000 xg for 5 minutes. Cells were washed 3 times in PBS (phosphate-buffered saline) and resuspended to an OD_{600} of 0.6. 200 µL cells were combined with a final concentration of 100 µg/mL of purified protein, and shaken at 200 rpm at room temperature for 45 minutes. This cell suspension was also harvested by centrifugation at 3,000 xg for 5 minutes and washed three times in PBS to remove unbound protein. Cells were resuspended to a volume of 200 µL, and 10

µL was added to a microscope slide, dried, and secured with a coverslip. For fluorescent microscopy, cells were viewed with a Zeiss Confocal Microscope at 63X magnification and excited with a green, fluorescent laser at 458 nm.

Analysis of Endolysin Treatment in a Simplified Gut Consortium by Quantitative Real-Time PCR.

The effect of LysH1 targeting *R. gnavus* in a simplified community was assessed by quantitative real-time PCR (qPCR) as previously described [52]. Each bacterial strain was grown in 10 mL of pre-reduced LYHBHI broth from a single colony, and incubated overnight at 37 ^{o}C under strict anaerobic conditions. The next day optical density was recorded by spectrophotometry (Biowave WPA CO8000, Avantor). The OD of each culture was standardized to 1. 10 µL of each normalised suspension was used to inoculate 10 ml of LYHBHI. 1 ml samples were taken at set time points (0, 6, 24, and 48 h post-inoculation). Samples were centrifuged at 8,000 x g for 4 min to separate the cell pellet from the supernatant. The pellets were stored at -20°C until needed for DNA extractions. Thawed cell pellets were used later for total genomic DNA (gDNA) extraction using the GenElute Bacterial Genomic DNA Kit (Sigma Aldrich) as per the kit instructions with a final elution volume of 200 µL per sample.

Selective species-specific primers for each were used for their detection by qPCR within the simplified community **(Table S2)**. Primer specificity was assessed by PCR. qPCR was performed in 96-well plates using the QuantStudio™ 5 Real-Time PCR instrument (Applied Biosystems, Thermo Fisher Scientific). Each reaction was carried out in 20 µL volumes containing 1 × SYBR ChamQ Universal SYBR qPCR Master Mix (Generon), 0.4 µM of each primer (forward and reverse), and 1 µL of gDNA. The thermocycling protocol consisted of initial denaturation at 95°C for 2 min, followed by 45 cycles of 95°C for 15 s, 60°C for 15 s and 72°C for 15 s. Standard curves for each bacterial strain were included in every assay plate, along with negative controls. The mass of each chromosome (one genome copy) was calculated by multiplying each genome size (bp) by 1.096×10^{-12} (average Molecular Weight of 1 bp in ng). Standard curves for each strain were prepared so that the highest concentration is 3×10^{6} copies /µL. Ten-fold serial dilutions were made from this stock to create a series of standards. One µL of each dilution was used as a template in the qPCR reaction to measure the cycle threshold value (Ct). The Ct values were plotted as a

function of the number of copies /µL to obtain the standard curves for each strain. Genome copy numbers of individual strains in the extracted gDNA from the mixed SIHUMI culture at each time point were calculated as copies/µL based on the measured Ct values. The number of copies/ml was estimated by multiplying the copies/µL by 200 (final elution volume of gDNA extracted from 1 ml of culture sample). Results were visualised using GraphPad Prism software.

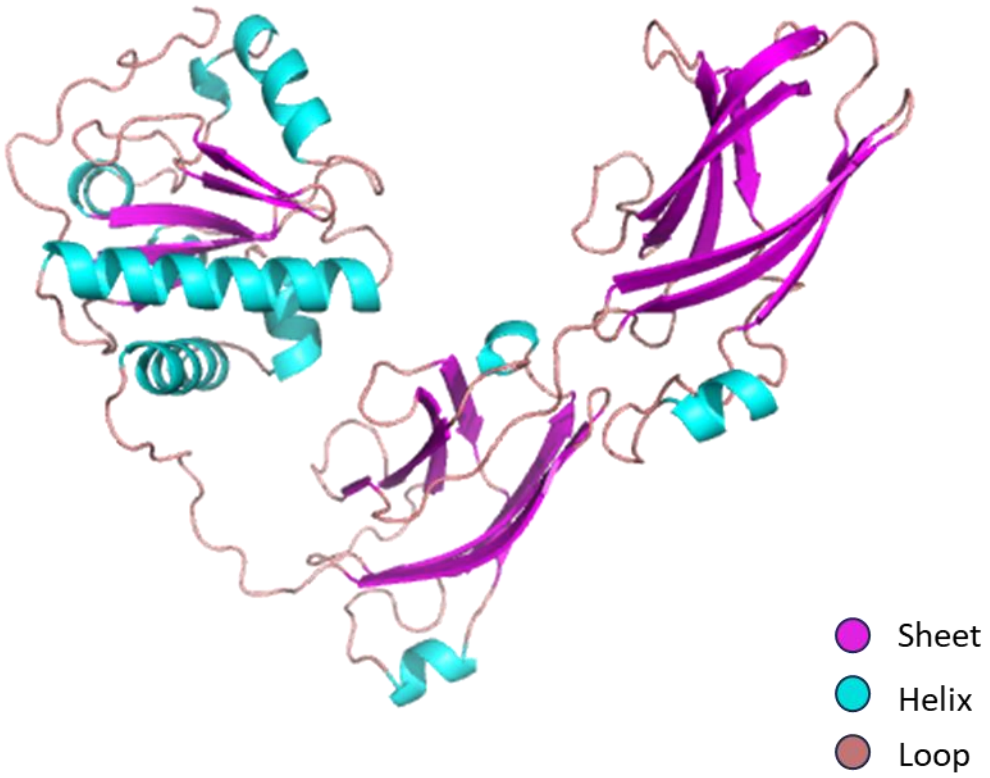


Figure S1. Structural analysis of LysH1 with secondary structure highlighted.

Figure S2. Multiple sequence alignment of pBLAST results for LysH1. Visualised with SeaView4.

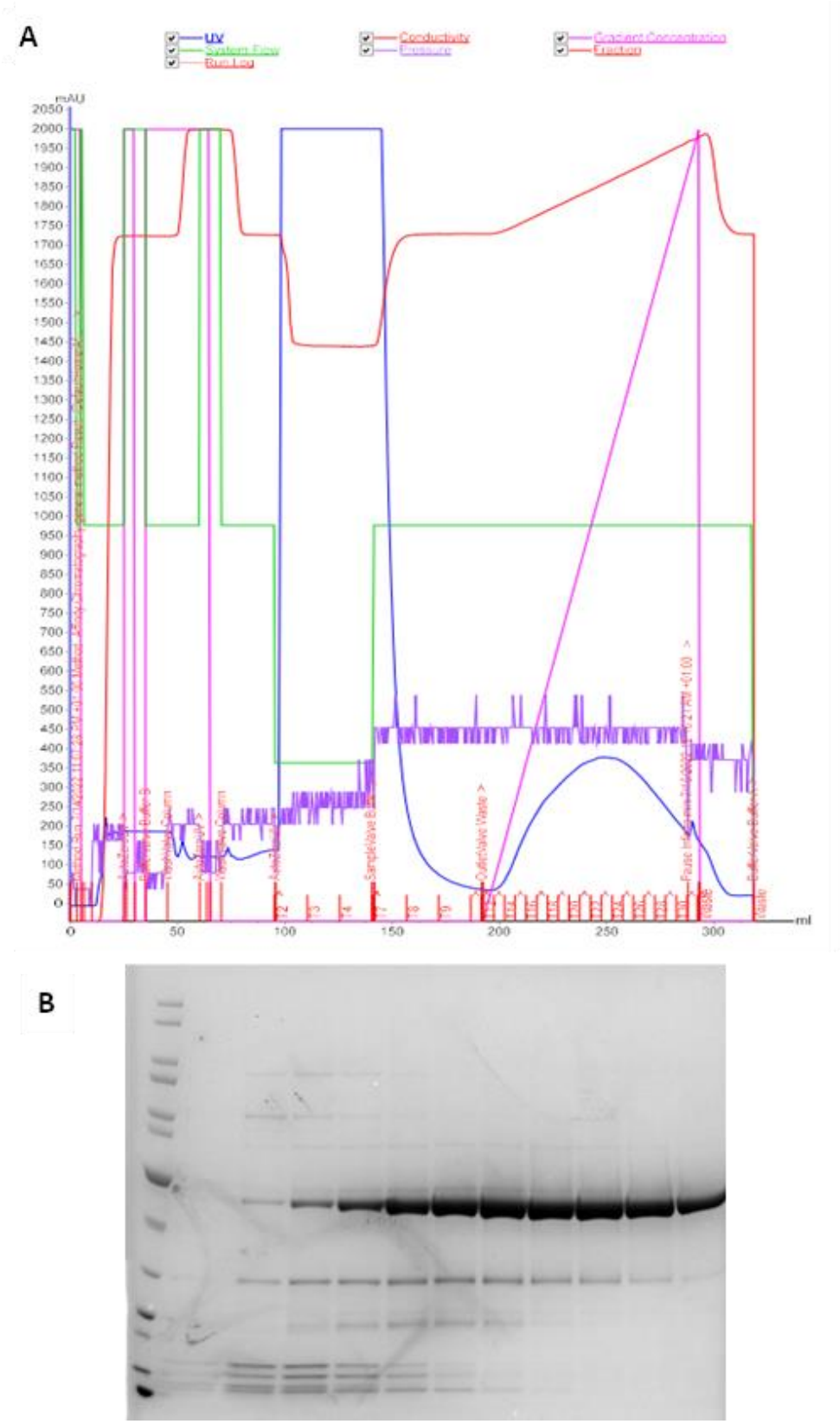


Figure S3. Protein purification with Akta Start affinity chromatography. A) Chromatogram of purification of LysH1. **B)** SDS gel to show fractions of purification process.

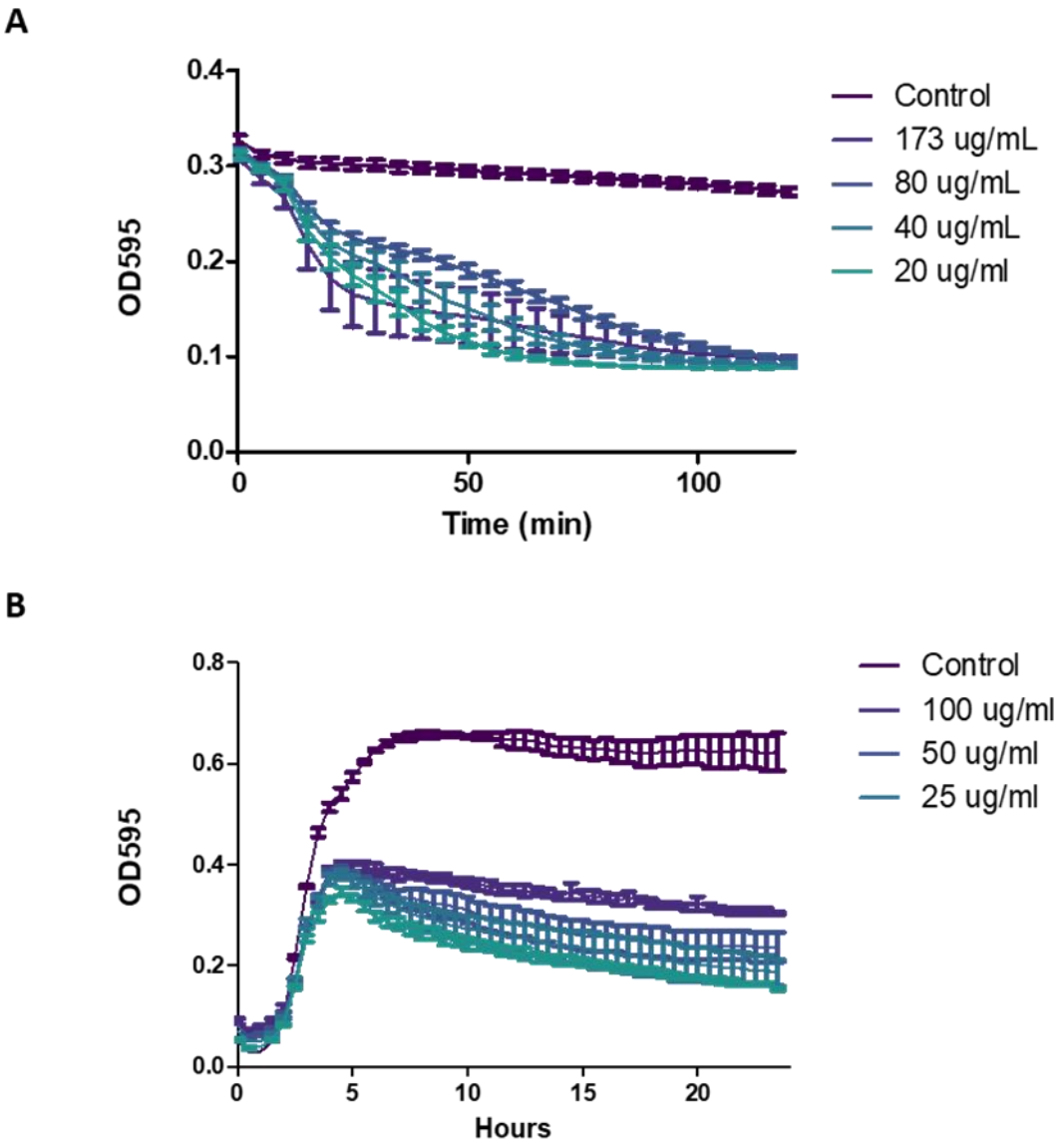


Figure S4. A) Turbidity reduction assay of LysH1 versus *E. faecium* 374IF1. Dead cells treated with LysH1 at varying final concentrations, incubated for 120 minutes at 37 ºC. **B)** *E. faecium* 374IF1 growth in LysH1-containing media at varying concentrations. Cells were inoculated at 1% and incubated for 24 hours at 37 ºC.

Table S1. PCR to check for VanA or VanB genes in a panel of strains of *E. faecium* and *E. faecalis*.

Species	Strain	Source/HPA Number	VanA / VanB Gene Amplification
E. faecium	APC 1025	EC251	VanA
	APC 1026	EC289	VanA
	APC 1027	EC292	VanA
	APC 1028	EC295	VanA
	APC 1029	EC300	VanA
	APC 1030	EC357	VanA
	APC 1031	EC520	VanB
	APC 1032	EC533	VanA
	APC 1033	EC537	VanA
	APC 1035	EC548	VanA
	APC 1037	EC571	Negative
	APC 1038	EC587	VanA
	APC 1044	EC748	VanA
	ATCC700221	Orla-Jensen Type Strain	VanA
	YCFA_942_30	Lab Isolate	Negative
	FAA_942_10	Lab Isolate	Negative
	MRS_932_37	Lab Isolate	VanA
	918H1	Lab Isolate	Negative
	374iF1	Lab Isolate	Negative
	DOTX16	Lab Isolate	Negative
E. faecalis	APC 1039	EC618	VanA
	V583	Lab Isolate	VanB
	APC 1039	EC618	VanA
	APC 1040	EC655	VanA

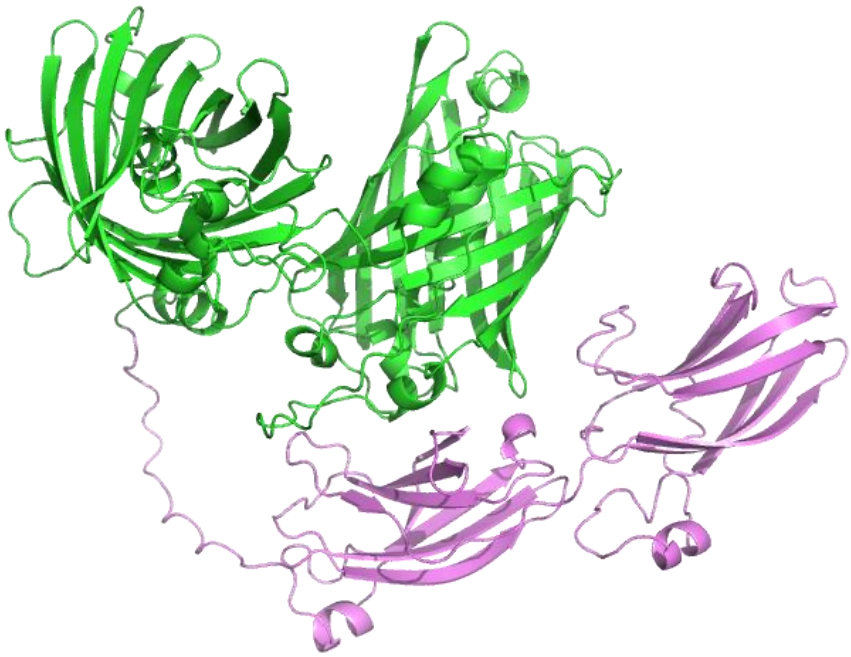


Figure S5. AlphaFold structural prediction for LysH1-CBD fused with green fluorescent protein (GFP). GFP in green, CBD domain in pink.

Table S2. Strain-specific primers used for qPCR.

Target	Strain	Oligonucleotide Sequence (5' to 3')	Reference
E. faecalis	OG1RF	F: ACGGAGATTGTCACGCTTAGT R: TCGGCATTATCTGGGTGGTC	(Buttimer et al. 2022)
E. coli	LF82	F: CGGGTGTTGTCCTAACTGCT R: CGAGTGGTCATTGGCCTCAT	(Buttimer et al. 2022)
R. gnavus	ATCC 29149	F: GCGTGCTTGTATTCCGGATG R: GCCTGAACAGTTGCTTTCGG	(Guerin et al. 2021)
F. prausnitzii	A2-165	F: TATTGCACAATGGGGGAAAC R: CAACAGGAGTTTACAATCCGAAG	(Lengfelder et al. 2019)
P. vulgatus	DSM1447	F: AAGCAGCAGGGAAATGTGGA R: CTTTCCTTACTTGCGCGTCG	(Buttimer et al. 2022)
L. plantarum	WCFS1	F: CGAAGAAGTGCATCGGAAAC R: TCACCGCTACACATGGAGTT	(Lengfelder et al. 2019)
B. longum	ATCC15707	F: GAGGCGATGGTCTGGAAGTT R: CCACATCGCCGAGAAGATTC	(Lawley et al. 2017)
C. difficile	VPI 10463	F: AAGCAGTAACAGTAGCAGTAGAA R: ATTTTCCAACTTCTTCATCACCA	(Kohler et al. 2022)

Author Contributions

EM and CH conceptualised the study. EM cloned and characterised the endolysin. NRC and EM carried out the community analysis. EM and DH carried out *in silico* analysis. ASB isolated the phage. NNA and LAD managed the project. EM, CH, and PR wrote the manuscript.

References

[1] I. Klare, C. Konstabel, D. Badstübner, G. Werner, and W. Witte, “Occurrence and spread of antibiotic resistances in Enterococcus faecium,” Int J Food Microbiol, vol. 88, pp. 269–290, 2003, doi: 10.1016/S0168-1605(03)00190-9.

[2] S. J. van Hal et al., “Evolutionary dynamics of Enterococcus faecium reveals complex genomic relationships between isolates with independent emergence of vancomycin resistance,” Microb Genom, vol. 2, no. 1, p. e000048, Jan. 2016, doi: 10.1099/MGEN.0.000048/CITE/REFWORKS.

[3] T. O’Driscoll and C. W. Crank, “Vancomycin-resistant enterococcal infections: epidemiology, clinical manifestations, and optimal management,” Infect Drug Resist, vol. 8, pp. 217–230, Jul. 2015, doi: 10.2147/IDR.S54125.

[4] B. L. Hollenbeck and L. B. Rice, “Intrinsic and acquired resistance mechanisms in enterococcus,” 2012, doi: 10.4161/viru.21282.

[5] D. Hourigan, E. Stefanovic, C. Hill, and R. P. Ross, “REVIEW Open Access Promiscuous, persistent and problematic: insights into current enterococcal genomics to guide therapeutic strategy”, doi: 10.1186/s12866-024-03243-2.

[6] S. J. van Hal et al., “The global dissemination of hospital clones of Enterococcus faecium,” Genome Med, vol. 13, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/S13073-021-00868-0/FIGURES/6.

[7] J. Top, R. Willems, and M. Bonten, “Emergence of CC17 Enterococcus faecium: from commensal to hospital-adapted pathogen,” FEMS Immunol Med Microbiol, vol. 52, no. 3, pp. 297–308, Apr. 2008, doi: 10.1111/J.1574-695X.2008.00383.X.

[8] B. Janning and P. H. in’t Veld, “Susceptibility of bacterial strains to desiccation: a simple method to test their stability in microbiological reference materials,” Anal Chim Acta, vol. 286, no. 3, pp. 469–476, Feb. 1994, doi: 10.1016/0003-2670(94)85092-5.

[9] J. N. Steenbergen, J. Alder, G. M. Thorne, and F. P. Tally, “Daptomycin: a lipopeptide antibiotic for the treatment of serious Gram-positive infections”, doi: 10.1093/jac/dkh546.

[10] M. A. Misiakou, F. B. Hertz, K. Schønning, S. Häussler, and K. L. Nielsen, “Emergence of linezolid-resistant Enterococcus faecium in a tertiary hospital in Copenhagen,” Microb Genom, vol. 9, no. 7, p. 001055, Jul. 2023, doi: 10.1099/MGEN.0.001055/CITE/REFWORKS.

[11] S. Mollerup et al., “Rapid in vivo development of resistance to daptomycin in vancomycin-resistant Enterococcus faecium due to genomic alterations,” FEMS Microbiol Lett, vol. 369, no. 1, 2022, doi: 10.1093/FEMSLE/FNAC063.

[12] W. Zeng et al., “Acquisition of Daptomycin Resistance by Enterococcus faecium Confers Collateral Sensitivity to Glycopeptides,” Front Microbiol, vol. 13, Apr. 2022, doi: 10.3389/FMICB.2022.815600/FULL.

[13] L. Puchter, I. F. Chaberny, F. Schwab, R.-P. Vonberg, F.-C. Bange, and E. Ebadi, “Economic burden of nosocomial infections caused by vancomycin-resistant enterococci”, doi: 10.1186/s13756-017-0291-z.

[14] N. J. Talley et al., “An evidence-based systematic review on medical therapies for inflammatory bowel disease,” American Journal of Gastroenterology, vol. 106, no. SUPPL. 1, Apr. 2011, doi: 10.1038/AJG.2011.58.

[15] R. J. Xavier and D. K. Podolsky, “Unravelling the pathogenesis of inflammatory bowel disease,” Nature 2007 448:7152, vol. 448, no. 7152, pp. 427–434, Jul. 2007, doi: 10.1038/nature06005.

[16] Y. R. Yu and J. R. Rodriguez, “Clinical presentation of Crohn’s, ulcerative colitis, and indeterminate colitis: Symptoms, extraintestinal manifestations, and disease phenotypes,” Semin Pediatr Surg, vol. 26, no. 6, pp. 349–355, Dec. 2017, doi: 10.1053/J.SEMPEDSURG.2017.10.003.

[17] T. H. Bisgaard, K. H. Allin, L. Keefer, A. N. Ananthakrishnan, and T. Jess, “Depression and anxiety in inflammatory bowel disease: epidemiology, mechanisms and treatment,” Nature Reviews Gastroenterology & Hepatology 2022 19:11, vol. 19, no. 11, pp. 717–726, Jun. 2022, doi: 10.1038/s41575-022-00634-6.

[18] S. H. Lee, J. eun Kwon, and M. La Cho, “Immunological pathogenesis of inflammatory bowel disease,” Intest Res, vol. 16, no. 1, p. 26, 2018, doi: 10.5217/IR.2018.16.1.26.

[19] C. O. Elson, Y. Cong, V. J. McCracken, R. A. Dimmitt, R. G. Lorenz, and C. T. Weaver, “Experimental models of inflammatory bowel disease reveal innate, adaptive, and regulatory mechanisms of host dialogue with the microbiota,” Immunol Rev, vol. 206, no. 1, pp. 260–276, Aug. 2005, doi: 10.1111/J.0105-2896.2005.00291.X.

[20] M. Joossens et al., “Dysbiosis of the faecal microbiota in patients with Crohn’s disease and their unaffected relatives,” Gut, vol. 60, no. 5, pp. 631–637, May 2011, doi: 10.1136/GUT.2010.223263.

[21] B. P. Willing et al., “A Pyrosequencing Study in Twins Shows That Gastrointestinal Microbial Profiles Vary With Inflammatory Bowel Disease Phenotypes,” Gastroenterology, vol. 139, no. 6, pp. 1844-1854.e1, Dec. 2010, doi: 10.1053/J.GASTRO.2010.08.049.

[22] M. T. Henke, D. J. Kenny, C. D. Cassilly, H. Vlamakis, R. J. Xavier, and J. Clardy, “Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn’s disease, produces an inflammatory polysaccharide,” Proc Natl Acad Sci U S A, vol. 116, no. 26, pp. 12672–12677, Jun. 2019, doi: 10.1073/PNAS.1904099116/SUPPL_FILE/PNAS.1904099116.SD03.XLSX.

[23] E. H. Crost, E. Coletto, A. Bell, and N. Juge, “Ruminococcus gnavus: friend or foe for human health,” FEMS Microbiol Rev, vol. 47, no. 2, pp. 1–23, Mar. 2023, doi: 10.1093/FEMSRE/FUAD014.

[24] M. Yamaguchi and K. Yamamoto, “Mucin glycans and their degradation by gut microbiota,” Glycoconj J, vol. 40, no. 4, pp. 493–512, Aug. 2023, doi: 10.1007/S10719-023-10124-9.

[25] O. Nitzan, M. Elias, A. Peretz, and W. Saliba, “Role of antibiotics for treatment of inflammatory bowel disease,” World J Gastroenterol, vol. 22, no. 3, p. 1078, Jan. 2016, doi: 10.3748/WJG.V22.I3.1078.

[26] E. Murray, L. A. Draper, R. P. Ross, and C. Hill, “The Advantages and Challenges of Using Endolysins in a Clinical Setting,” Viruses, vol. 13, no. 4, Apr. 2021, doi: 10.3390/V13040680.

[27] W. Zhang et al., “Characterization of Enterococcus faecalis Phage IME-EF1 and Its Endolysin,” PLoS One, vol. 8, no. 11, p. e80435, Nov. 2013, doi: 10.1371/JOURNAL.PONE.0080435.

[28] H. Zhang, B. A. Buttaro, D. E. Fouts, S. Sanjari, B. S. Evans, and R. H. Stevens, “Bacteriophage φEf11 ORF28 Endolysin, a Multifunctional Lytic Enzyme with Properties Distinct from All Other Identified Enterococcus faecalis Phage Endolysins,” Appl Environ Microbiol, vol. 85, no. 13, pp. 555–574, Jul. 2019, doi: 10.1128/AEM.00555-19.

[29] P. Yoong, R. Schuch, D. Nelson, and V. A. Fischetti, “Identification of a broadly active phage lytic enzyme with lethal activity against antibiotic-resistant Enterococcus faecalis and Enterococcus faecium,” J Bacteriol, vol. 186, no. 14, pp. 4808–4812, Jul. 2004, doi: 10.1128/JB.186.14.4808-4812.2004/ASSET/91BCE7FE-D17F-4201-A450-3CF3EB45B4A6/ASSETS/GRAPHIC/ZJB0140438430003.JPEG.

[30] G. E. Premetis, A. Stathi, A. C. Papageorgiou, and N. E. Labrou, “Structural and functional features of a broad-spectrum prophage-encoded enzybiotic from Enterococcus faecium,” Sci Rep 13, 7450 (2023). https://doi.org/10.1038/s41598-023-34309-2.

[31] P. Gong et al., “Characterization of Enterococcus faecium bacteriophage IME-EFm5 and its endolysin LysEFm5,” Virology, vol. 492, pp. 11–20, May 2016, doi: 10.1016/J.VIROL.2016.02.006.

[32] J. Wang et al., “Bacteriophage endolysin Ply113 as a potent antibacterial agent against polymicrobial biofilms formed by enterococci and Staphylococcus aureus,” Front Microbiol, vol. 14, p. 1304932, Dec. 2023, doi: 10.3389/FMICB.2023.1304932/BIBTEX.

[33] C. Buttimer et al., “Temperate bacteriophages infecting the mucin-degrading bacterium Ruminococcus gnavus from the human gut,” Gut Microbes, vol. 15, no. 1, 2023, doi: 10.1080/19490976.2023.2194794.

[34] P. Jones et al., “InterProScan 5: genome-scale protein function classification,” vol. 30, no. 9, pp. 1236–1240, 2014, doi: 10.1093/bioinformatics/btu031.

[35] J. Söding, A. Biegert, and A. N. Lupas, “The HHpred interactive server for protein homology detection and structure prediction,” Nucleic Acids Res, vol. 33, no. Web Server issue, p. W244, Jul. 2005, doi: 10.1093/NAR/GKI408.

[36] M. van Kempen et al., “Fast and accurate protein structure search with Foldseek,” Nature Biotechnology 2023 42:2, vol. 42, no. 2, pp. 243–246, May 2023, doi: 10.1038/s41587-023-01773-0.

[37] “AlphaFold Protein Structure Database.” Accessed: Jan. 08, 2024. [Online]. Available: https://AlphaFold.ebi.ac.uk/

[38] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” J Mol Biol, vol. 215, no. 3, pp. 403–410, 1990, doi: 10.1016/S0022-2836(05)80360-2.

[39] M. Gouy, S. Guindon, and O. Gascuel, “SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building,” Mol Biol Evol, vol. 27, no. 2, pp. 221–224, Feb. 2010, doi: 10.1093/MOLBEV/MSP259.

[40] F. Sievers et al., “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega,” Mol Syst Biol, vol. 7, p. 539, 2011, doi: 10.1038/MSB.2011.75.

[41] “The Proteomics Protocols Handbook,” The Proteomics Protocols Handbook, 2005, doi: 10.1385/1592598900.

[42] M. T. Henke, D. J. Kenny, C. D. Cassilly, H. Vlamakis, R. J. Xavier, and J. Clardy, “Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn’s disease, produces an inflammatory polysaccharide,” Proc Natl Acad Sci U S A, vol. 116, no. 26, pp. 12672–12677, Jun. 2019, doi: 10.1073/PNAS.1904099116/SUPPL_FILE/PNAS.1904099116.SD03.XLSX.

[43] P. T. McKenney, L. Ling, G. Wang, S. Mane, and E. G. Pamer, “Complete Genome Sequence of Enterococcus faecium ATCC 700221,” Genome Announc, vol. 4, no. 3, pp. 386–402, 2016, doi: 10.1128/GENOMEA.00386-16.

[44] P. Mellroth et al., “LytA, Major Autolysin of Streptococcus pneumoniae, Requires Access to Nascent Peptidoglycan,” J Biol Chem, vol. 287, no. 14, p. 11018, Mar. 2012, doi: 10.1074/JBC.M111.318584.

[45] H. Zhang et al., “Bactericidal synergism between phage endolysin Ply2660 and cathelicidin LL-37 against vancomycin-resistant Enterococcus faecalis biofilms,” npj Biofilms and Microbiomes 2023 9:1, vol. 9, no. 1, pp. 1–14, Apr. 2023, doi: 10.1038/s41522-023-00385-5.

[46] H. N. Wilkinson et al., “Selective Depletion of Staphylococcus aureus Restores the Skin Microbiome and Accelerates Tissue Repair after Injury,” Journal of Investigative Dermatology, Feb. 2024, doi: 10.1016/J.JID.2024.01.018.

[47] H. Kim, T. Park, I. Kwon, and J. Seo, “Specific inhibition of Streptococcus bovis by endolysin LyJH307 supplementation shifts the rumen microbiota and metabolic pathways related to carbohydrate metabolism,” J Anim Sci Biotechnol, vol. 12, no. 1, pp. 1–15, Dec. 2021, doi: 10.1186/S40104-021-00614-X/FIGURES/7.

[48] J. R. Diaz Rodriguez, W. Lu, J. M. Papadopoulos, O. S. Venturelli, and P. A. Romero, “Engineered lysins to modulate human gut microbiome communities”, doi: 10.1101/2024.05.14.594189.

[49] C. S. Eun et al., “Induction of bacterial antigen-specific colitis by a simplified human microbiota consortium in gnotobiotic interleukin-10-/- mice,” Infect Immun, vol. 82, no. 6, pp. 2239–2246, 2014, doi: 10.1128/IAI.01513-13/SUPPL_FILE/ZII999090669SO1.PDF.

[50] C. Buttimer et al., “Impact of a phage cocktail targeting Escherichia coli and Enterococcus faecalis as members of a gut bacterial consortium in vitro and in vivo,” Front Microbiol, vol. 13, p. 936083, Jul. 2022, doi: 10.3389/FMICB.2022.936083/BIBTEX.

[51] B. P. Ganesh, R. Klopfleisch, G. Loh, and M. Blaut, “Commensal Akkermansia muciniphila Exacerbates Gut Inflammation in Salmonella Typhimurium-Infected Gnotobiotic Mice,” PLoS One, vol. 8, no. 9, p. e74963, Sep. 2013, doi: 10.1371/JOURNAL.PONE.0074963.

[52] N. S. Ríos Colombo et al., “Impact of bacteriocin-producing strains on bacterial community composition in a simplified human intestinal microbiota,” Front Microbiol, vol. 14, p. 1290697, Dec. 2023, doi: 10.3389/FMICB.2023.1290697/BIBTEX.

[53] L. Sutherland et al., “Double blind, placebo controlled trial of metronidazole in Crohn’s disease.,” Gut, vol. 32, no. 9, pp. 1071–1075, Sep. 1991, doi: 10.1136/GUT.32.9.1071.

[54] P. Gionchetti et al., “Prophylaxis of pouchitis onset with probiotic therapy: a double-blind, placebo-controlled trial,” Gastroenterology, vol. 124, no. 5, pp. 1202–1209, May 2003, doi: 10.1016/S0016-5085(03)00171-9.

[55] E. Murray, L. A. Draper, R. P. Ross, and C. Hill, “The Advantages and Challenges of Using Endolysins in a Clinical Setting,” Viruses, vol. 13, no. 4, Apr. 2021, doi: 10.3390/V13040680.

[56] J. A. Gerlt et al., “Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks,” Biochim Biophys Acta, vol. 1854, no. 8, p. 1019, Aug. 2015, doi: 10.1016/J.BBAPAP.2015.04.015.

[57] M. C. Domingo et al., “High Prevalence of Glycopeptide Resistance Genes vanB, vanD, and vanG Not Associated with Enterococci in Human Fecal Flora,” Antimicrob Agents Chemother, vol. 49, no. 11, p. 4784, Nov. 2005, doi: 10.1128/AAC.49.11.4784-4786.2005.

[58] H. Sokol et al., “Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients,” Proc Natl Acad Sci U S A, vol. 105, no. 43, pp. 16731–16736, Oct. 2008, doi: 10.1073/PNAS.0804812105.

[59] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” J Mol Biol, vol. 215, no. 3, pp. 403–410, 1990, doi: 10.1016/S0022-2836(05)80360-2.

[60] P. Jones et al., “InterProScan 5: genome-scale protein function classification,” vol. 30, no. 9, pp. 1236–1240, 2014, doi: 10.1093/bioinformatics/btu031.

[61] J. Söding, A. Biegert, and A. N. Lupas, “The HHpred interactive server for protein homology detection and structure prediction,” Nucleic Acids Res, vol. 33, no. Web Server issue, p. W244, Jul. 2005, doi: 10.1093/NAR/GKI408.

Chapter 4. Genomic and Phenotypic Analysis of Endolysin-Resistant Mutants of *Ruminococcus gnavus*.

Ellen Murray, Ekaterina V. Khokhlova, David Hourigan, Neda Nezam Abadi, Lorraine A. Draper, Andrey Shkoporov, Paul R. Ross, Colin Hill.

Abstract

The issue of antimicrobial resistance (AMR) is a pressing concern, necessitating the discovery of new antimicrobials to combat resistant microbes. Endolysins are phage-encoded peptidoglycan hydrolases that are being explored as potential alternatives or adjuncts to antibiotics due to their efficacy, rapid action, and potential for genetic engineering. The reported lack of resistance development is another advantage. Numerous studies have identified viable endolysin candidates for treating various infections. In this study, we isolated endolysins that could target *Ruminococcus gnavus*, a commensal gut bacterium associated with inflammatory bowel disease (IBD). Five novel endolysins-encoding genes were successfully cloned and recombinantly expressed and one was selected for further testing (LysIBDN1). All of the endolysins initially demonstrated effectiveness against *R. gnavus*, but subsequent analysis revealed the emergence of endolysin-resistant mutants. The mutant strains all demonstrated significant fitness costs of resistance. Mutant strains were isolated and characterised, and their genomes were sequenced. This revealed a series of SNPs, deletions and insertions in genes associated with metabolism, transport, and the stringent stress response. The use of green fluorescent protein (GFP) fused to the endolysin cell wall binding domain demonstrated a reduction in the binding efficiency of LysIBDN1 to the mutants, providing insights into the mechanisms of endolysin resistance.

Introduction

Ruminococcus gnavus is a gram-positive, strictly anaerobic bacterium of the family *Lachnospiraceae* in the phylum *Bacillota*. In 2018, the species was reassigned as *Mediterraneibacter gnavus* was introduced to avoid confusion with a different genus *Ruminococcus*, now part of the family *Ruminococcaceae*. But since the old name remains a validly published homotypic synonym, for the purpose of this study, we will continue to refer to the bacterium as *R. gnavus* [1]. It is a gut commensal that is found in the digestive tract of over 90% of humans [2]. In recent years, several studies have associated increased levels of *R. gnavus* with gut diseases such as inflammatory bowel disease (IBD), particularly Crohn's disease.

IBD is an umbrella term for Crohn's disease and Ulcerative Colitis (UC). IBD is characterised by chronic digestive tract inflammation with periods of remission. Symptoms of Crohn's disease include abdominal pain, diarrhoea, weight loss, fever, the passing of blood or mucus, and clinical obstruction of the bowels [3]. As of 2021, 1.3 million people in Europe are estimated to be affected by IBD, which equates to about 0.2% of the total population [4]. IBD is not currently curable, and treatment instead aims to manage symptoms and flare-ups through the use of corticosteroids, aminosalicylates, antibiotics, and immunosuppressive drugs [5]. Management of inflammation includes using proinflammatory cytokine tumour necrosis factor-α (TNF-α) inhibitors [6]. If drug treatments do not improve the quality of life for patients, surgery can be implemented to remove affected portions of the digestive tract. The NHS has reported that up to 75% of patients will need surgery to repair damage to their gastrointestinal tract and treat complications of Crohn's disease [7].

The exact cause of IBD is not fully established, but it is thought to be a combination of genetic factors, and environmental triggers, and a role has been suggested for commensal gut bacteria. Risk factors for the development of IBD include antibiotic exposure, method of birth, breastfeeding, and diet. [8]. A healthy human gut microbiome is a complex community of microorganisms that interact with each other and with their host. When the microbiome is disrupted, it can result in a disease state such as IBD. A study investigating the composition of the microbiota of Crohn's disease patients and their unaffected relatives concluded that patients have a lower abundance of *Bifidobacterium*, a specific *Clostridium* cluster XIVa, *Dialister invisus,*

and *Faecalibacterium prausnitzii*, together with a significant increase in *R. gnavus* [9]. This disrupted state existed in IBD patients alone, despite sharing a genetic background and dietary habits with their unaffected relatives. A more recent study has carried out metagenomic sequencing of stool samples from IBD patients and also reported a significantly higher abundance of *R. gnavus* [10]. Interestingly, this study also detected substantial blooms in *R. gnavus* levels that coincided with disease flare-ups in patients. Certain strains of *R. gnavus* (including strain JM6515^T that was used as a target strain for this study) are capable of using host mucin glycans as their energy source, and so these blooms could potentially result in damage to the gut epithelial mucous barrier [11]. *R. gnavus* has also been characterised as producing an inflammatory polysaccharide. This consists of a complex glucorhamnan polysaccharide with a rhamnose backbone and glucose side chains. A 2019 study has suggested that this polysaccharide induces an immune reaction that produces inflammatory cytokine tumour necrosis factor-alpha (TNF-α), often associated with disease severity in IBD [12]. This further supports a link between *R. gnavus* and IBD.

Due to these associations, targeting the bacterium with antimicrobials may be a potential treatment option for Crohn's disease. Antibiotics such as metronidazole and ciprofloxacin are commonly used for the management of Crohn's disease. However, it is well-documented that antimicrobial resistance (AMR) is a global health emergency: the Centre for Disease Control (CDC) states that if no new antibiotics are made available by the year 2050, it is estimated that 10 million deaths will occur per year by bacterial infections [13]. Using antibiotics for the management of lifelong diseases, such as IBD, could exacerbate resistance levels and increase the threat of AMR. There has been a rise in the number of studies investigating alternative approaches to antibiotics such as bacteriophage (phage) therapy. For example, a 2022 publication documented the use of a phage consortium to suppress intestinal inflammation caused by IBD-associated bacteria [14]. This study focused on *Klebsiella pneumoniae* and used a cocktail of five phages to achieve immune suppression.

Phage endolysins have gained much traction recently as a promising alternative to traditional antibiotics. They are classed as phage-encoded peptidoglycan hydrolases and are generally comprised of a two-domain modular structure consisting of an enzymatically active domain (EAD) and a cell wall binding domain (CBD). They are encoded in the phage genome and are involved in the phage

replication cycle. Endolysins, paired with a second protein known as a holin, are produced within the host bacterial cell during phage replication and assembly until they reach a specific level that allows them to degrade the cell wall of the bacteria. This ultimately disrupts the integrity of the cell wall, leading to lysis and release of phage progeny into the surrounding environment. This “lysis from within” can be harnessed as a therapeutic by adding endolysins exogenously to gram-positive cells and achieving the same cell lysis event through “lysis from without”. Endolysins have been reported to be desirable next-generation antimicrobials due to their specificity, mode of action, and a reported lack of resistance mechanisms [15].

In this study, we present the isolation and characterisation of five novel endolysins targeting *R. gnavus*. These endolysins were identified in the genomes of novel *R. gnavus* phages, which have been reported by Buttimer et al. [16]. These phages were isolated from human faecal and environmental samples, and genome analysis revealed they had a temperate lifestyle. The phages were reported to lyse their bacterial targets in classical broth culture assays; however, when applied to an *in vivo* setting in a mouse trial, the phages co-existed with their target bacterium and did not cause a significant reduction in levels of *R. gnavus*. These findings confirm that the phages isolated in this study were not suitable candidates to be brought forward for phage therapy of IBD patients. Despite this setback, we aimed to target *R. gnavus* through the recombinant expression of their endolysins.

While the endolysins investigated here were lytic for the bacterium, we observed the regrowth of *R. gnavus* in the presence of endolysin. It was determined that these were endolysin-resistant mutants of *R. gnavus* JCM6515^{T}. Genomic analysis of these strains revealed a number of single nucleotide polymorphisms (SNPs), insertions and deletions in the genomes of these mutant strains. To our knowledge, this is the first instance of endolysins targeting *R. gnavus* to be reported in the literature and the first characterisation of endolysin-resistant mutants arising from a single exposure.

Results

We followed a set of experimental steps to isolate biologically active endolysins from *R. gnavus* phages, characterise their antimicrobial activity and investigate the resistant mutants of *R. gnavus* that emerged during this characterisation (**Figure 1**). Endolysin genes were bioinformatically identified in the genomes of the previously mentioned *R. gnavus* phages. All five endolysin genes were cloned into expression vectors for recombinant protein expression. In addition, the CBD domain of LysIBDN1 was expressed as a GFP-fusion protein to allow us to functionally characterise this novel domain. Lytic activity tests with all five endolysins gave rise to an initial kill, followed by a phenomenon of re-growth of *R. gnavus* in the presence of endolysin. These apparently resistant colonies were picked and purified. The purified colonies were sent for whole genome sequencing (WGS), along with the wild-type strain. The sequencing results identify single nucleotide polymorphisms (SNPs), insertions and deletions in several genes.

In Silico and Structural Analysis of Endolysin Genes.

The genomes of five lysogenic *R. gnavus* phages φRg507T2/2, φRg507T2/3, φRg519T2, φRgPS6, φRgIBDN1 (accession numbers MT980836, MT980837, MT980838, MT980839, MT980840, respectively) were compared by employing BLASTp and visualised with GGgenomes **(Figure 2A).** All five phage genomes encode an endolysin, all with an N-acetylmuramoyl-L-alanine amidase active domain as predicted by InterPro. The exact classification of these N-acetylmuramoyl-L-alanine amidases splits them into two protein families. The endolysins from φRg507T2/2, φRg519T2, φRgPS6, and φRgIBDN1 belong to IPR002508 – Amidase Type 2, whereas the endolysin from φRg507T2/3 belongs to IPR002502 – Amidase Type 3. InterPro was also used to assign functional domains to the endolysins. All endolysins possess a Zinc-dependent N-terminal catalytic domain. HHpred analysis was also employed to assign domain functions [17]. The endolysin from φRg507T2/3 displayed a C-terminal domain comprising three ChW Clostridial hydrophobic repeats (SMART accession number SM00728). The four remaining endolysins showed C-terminal domains of endolysins of *Enterococcus* phages IMEEF1 (PDB 6IST) and M1EF22 (PDB 7D55).

The similarity between the five endolysins was further assessed by amino acid multiple sequence alignments. Alignments showed that enzymes in family IPR002508 had significant similarity (>95%) to each other but were dissimilar to the endolysin from φRg507T2/3 **(Figure 2B).**

AlphaFold2 was used to predict the tertiary structure of the endolysins [18]. Structural analysis was carried out on the endolysin from φRgIBDN1 (LysIBDN1) as a representative of protein family IPR002508 and φRg507T2/3 (Lys3-1) as a representative of protein family IPR002502, using AlphaFold v2.3 and visualised with PyMOL [19]. Each model produced a per-residue confidence score (pLDDT) (**Figure 2C and 2D**)**.** Both proteins displayed a 2-domain structure of high confidence, with a linker region connecting them. The N-terminal enzymatically active domain of Lys3-1 consists of five α-helices whereas the C-terminal binding domain is comprised of three anti-parallel β-sheets. The N-terminal residue of LysIBDN1 comprises five α-helices and its C-terminal comprises two antiparallel β-sheets flanked by an α-helix at both sides. Secondary structure predictions are visualised in supplementary material **Figure S1**.

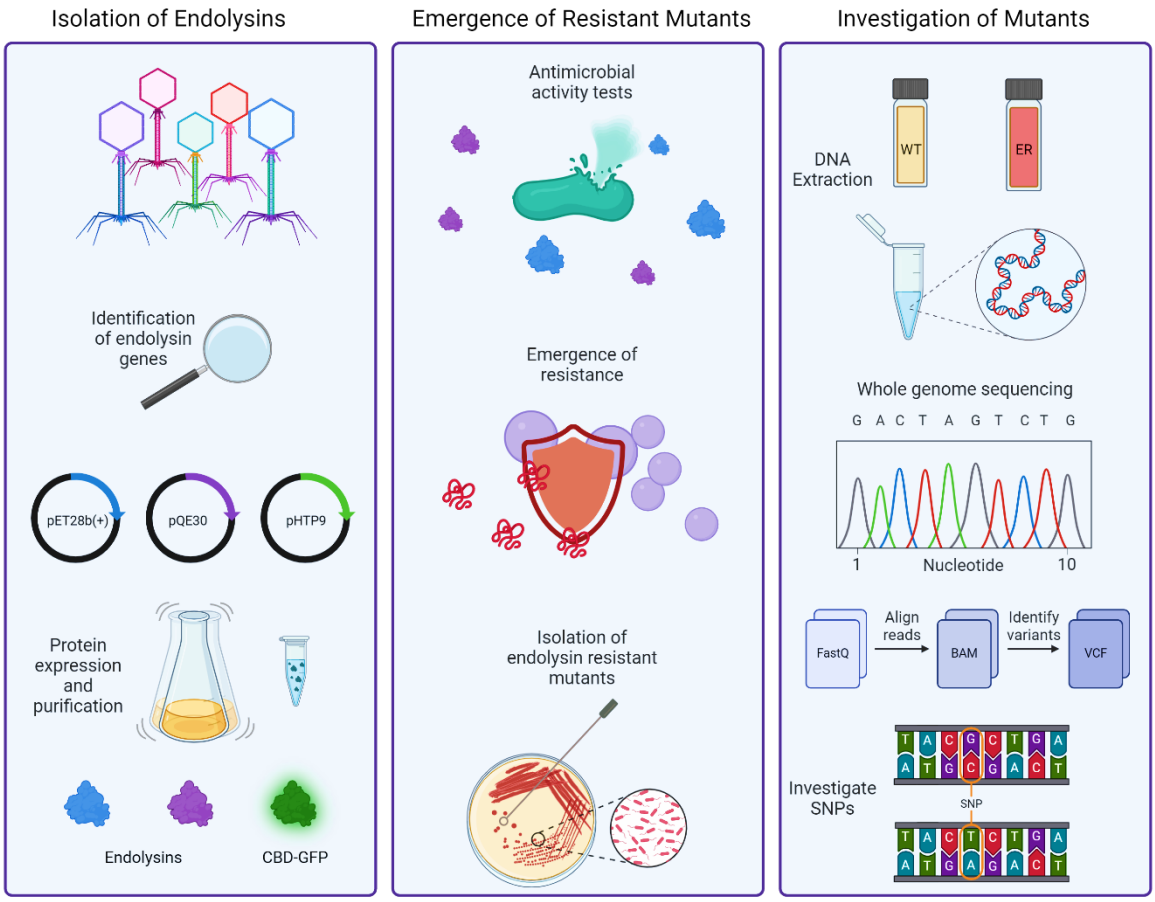


Figure 1. Schematic overview of the experimental steps to investigate the emergence of resistant mutants to *R. gnavus* in response to endolysin treatment.

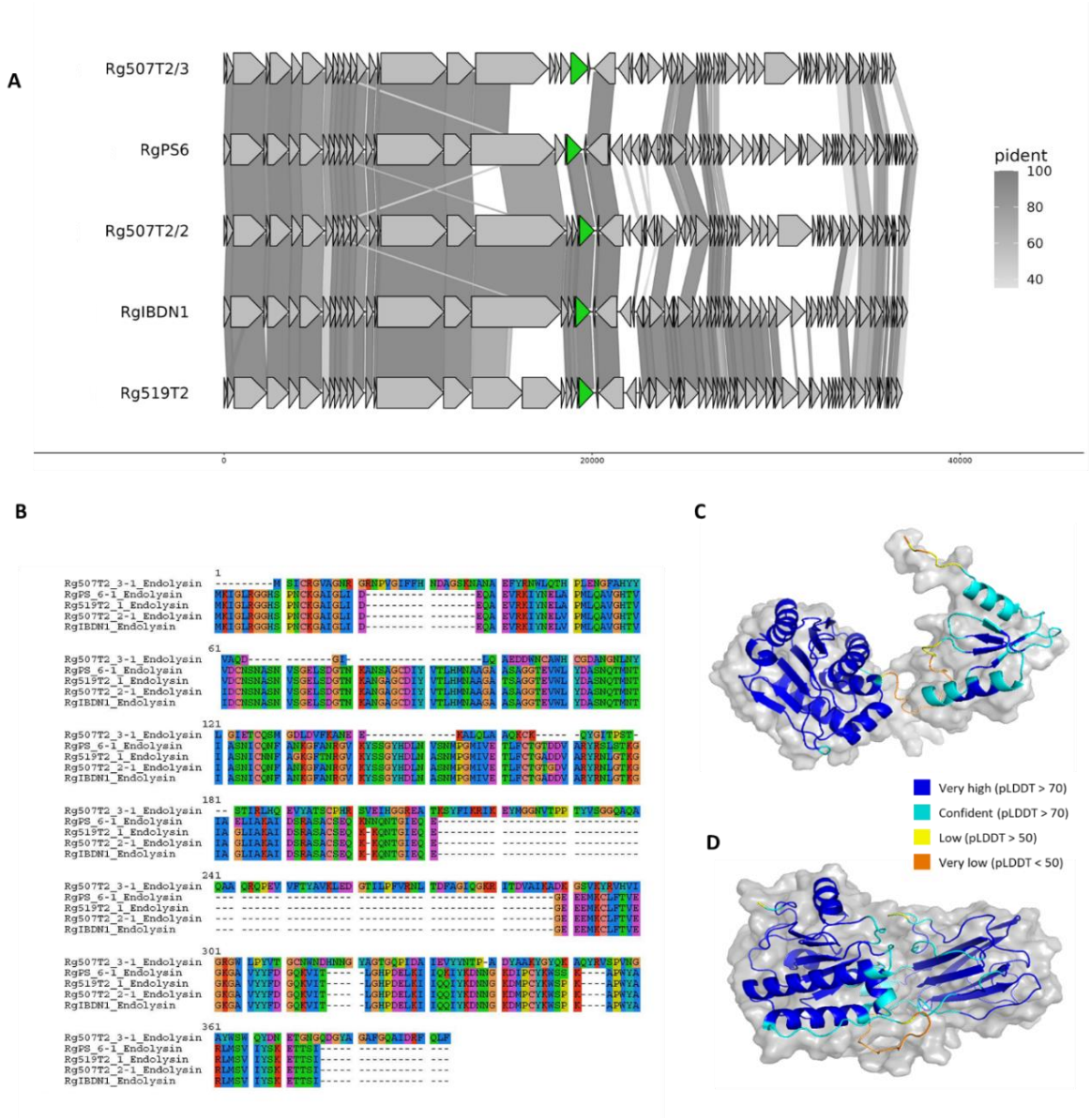


Figure 2. ***In silico*** **analysis of endolysin genes**. **A)** Genomes of *R. gnavus* lysogenic phages compared by BLASTp and visualised using gggenomes. **B)** Multiple sequence alignment of *R. gnavus* endolysin genes, amino acid sequence alignments produced with Clustal Omega and visualised using SeaView4. **C)** Structural analysis of LysIBDN1 carried out with AlphaFold and visualised with PyMOL. **D)** Structural analysis of Lys3-1 carried out with AlphaFold2 and visualised with PyMOL.

Cloning, Expression, and Purification of Endolysins.

Endolysin genes were PCR-amplified and cloned into *E. coli* expression vectors (**Table S1**). The complete endolysins ranged from 267 – 342 amino acids in size, with weights ranging from 28.81 – 36.03 kDa, including 6xHis-tags for subsequent purification. The amplicons were cloned into the pQE30 expression vector with the exception of Lys3-1. This was because Lys3-1 displayed solubility issues, resulting in low yields when expressed in pQE30, but this was improved with a change of cell line. Endolysin genes cloned in the pQE30 vector were introduced into E. coli M15 [pRep4], whereas Lys3-1 in pET28b(+) was introduced into E. coli BL21(DE3). After transformation the 6xHis-tagged endolysins were purified by metal-affinity chromatography and elution fractions were verified with SDS-PAGE (4-12% polyacrylamide gradient) to confirm the expression of soluble endolysins (Figure 3A). Large-scale expression (3L preparations) of endolysins Lys3-1 and LysIBDN1 were carried out, and recombinant 6xHis-tagged proteins were purified using the ÄKTA start protein purification system. The supplementary material shows an example of a resulting chromatogram from the purification of LysIBDN1 (**Figure S3**).

Lytic Activity of Endolysins.

To demonstrate that the recombinantly produced, purified endolysins retained their predicted activity after the purification process, spots of each lysin were applied to a lawn of *R. gnavus* (**Figure 3B**), where each gave a distinct zone in the target overlay.

Antimicrobial activity was tested against a panel of strains to determine the spectrum of activity of the *R. gnavus* endolysins **(Table 1)**. The lysins were primarily specific to strains of *R. gnavus*, but with some exceptions. Lys3-1 had the highest specificity, with the only off-target effect being observed as weak activity against *Clostridium scindens*. The remaining lysins gave stronger activity against *C. scindens*. LysPS6 had reduced activity against the strain *R. gnavus* PS/160, which, interestingly, is the strain from which its parental phage was isolated. Lys519 and LysIBDN1 had a second off-target effect, with activity against *Faecalibacterium prausnitzii* 918/95B.

For further activity assays, LysIBDN1 was chosen to be characterised as a potential antimicrobial for *R. gnavus.* This was due to its efficiency of recombinant production and its strong lytic activity on plates. To quantify the lytic activity of LysIBDN1, turbidity reduction assays (TRA) and killing curves were performed on *R.*

gnavus JCM6515^{T}. For the TRA against inactivated cells, LysIBDN1 caused a decrease in optical density (OD_{595}), with the reduction occurring in the first ten minutes (**Figure 4A**). This confirmed the rapid action of the endolysin against the target strain. However, an unusual trend was observed when the antimicrobial activity was tested against live cells in an actively growing culture. Up to 10 – 20 hours following exposure (depending on the concentration of lysin used), the growth of *R. gnavus* was suppressed. From this point onwards, a recurrence of growth was observed (**Figure 4B**).

To investigate this observation further, we isolated individual colonies that could grow in the presence of endolysin. This was achieved by inoculating semi-solid anaerobe basal agar containing 50 µg/ml endolysin (0.4% agarose) with *R. gnavus* JCM6515^{T} and LysIBDN1 at concentrations ranging from 12.5 – 100 µg/mL. These plates were stored anaerobically for up to 72 hours at 37 ^{o}C. As individual colonies appeared in the agar, they were picked and purified in the continued presence of endolysin. The frequency of endolysin-resistant mutants occurs at 10^{-5}. The identity of these endolysin-resistant colonies was confirmed by PCR with *R. gnavus*-specific primers and 16S rRNA Sanger Sequencing. The specific primers **(Table S3)** were designed by Guerin et al. and have an amplicon size of 115 bp [20]. The results are included in the supplementary materials **(Figure S4)**.

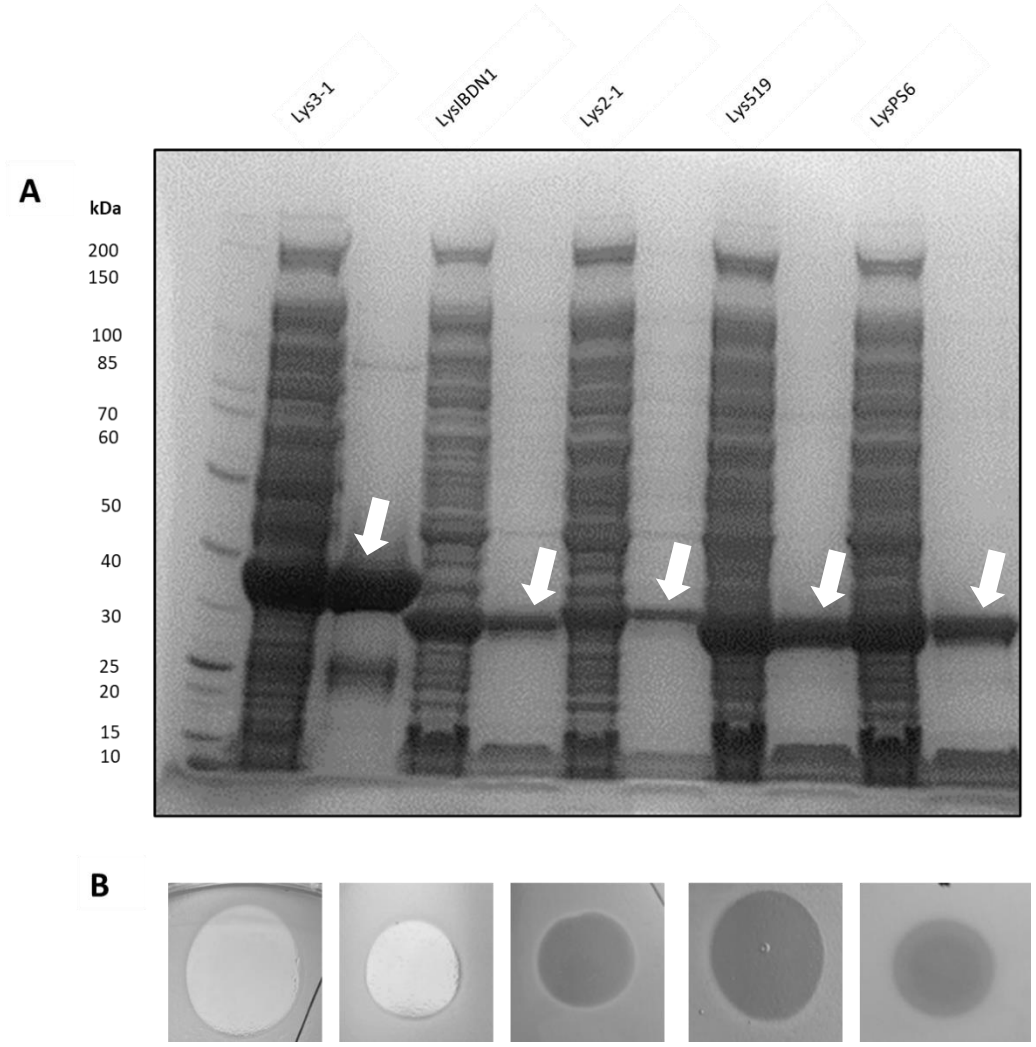


Figure 3. Expression and purification of endolysins. **A)** Protein expression, crude *E. coli* lysates of endolysins and their respective purification products, visualised by SDS-PAGE. Purified endolysins are indicated by the white arrows. **B)** Spot assays (10 µL, 600 µg/mL) of each purified protein against *R. gnavus* JCM6515^{T}. Proteins are displayed the following order: Lys3-1, LysIBDN1, Lys2-1, Lys519, LysPS6. The spot assays are also in the following order: Lys3-1, LysIBDN1, Lys2-1, Lys519, LysPS6.

Table 1. The spectrum of activity of *R. gnavus* endolysins targeting a panel of relevant strains. "+" indicates the observation of lysis, "+/-" indicates the observation of weak lysis or hazy zones, and "-" indicates no lysis.

Species	Strain	Lys3-1	Lys2-1	LysPS6	Lys519	LysIBDN1
R. gnavus	CCUG 54531	+	+	+	+	+
	CCUG 57208	+	+	+	+	+
	CCUG 57137	+	+	+	+	+
	CCUG 52279	+	+	+	+	+
	JCM6515^T	+	+	+	+	+
	PS/160	+	+	+/-	+	+
Faecalibacterium prausnitzii	918/95B	-	-	-	+/-	+/-
	922/41-1	-	-	-	-	-
	DSM 17677	-	-	-	-	-
Anaerostipes hadrus	942/1	-	-	-	-	-
Agathobacter rectalis	17629	-	-	-	-	-
Parabacteroides distasonis	DSM 20701	-	-	-	-	-
Bacteroides intestinalis	919/174	-	-	-	-	-
Clostridium scindens	DSM 5676	+/-	+	+	+	+
Bifidobacterium breve	UCC2003	-	-	-	-	-
Bifidobacterium longum	44B	-	-	-	-	-
Enterococcus faecalis	OG1RF	-	-	-	-	-
Enterococcus faecium	ATCC 700221	-	-	-	-	-
Ruminococcus gauvreauii	DSM 19829	-	-	-	-	-
Ruminococcus champanellensis	DSM 18848	-	-	-	-	-
Blautia producta	DSM 2950	-	-	-	-	-
Lactiplantibacillus plantarum	WCSF1	-	-	-	-	-

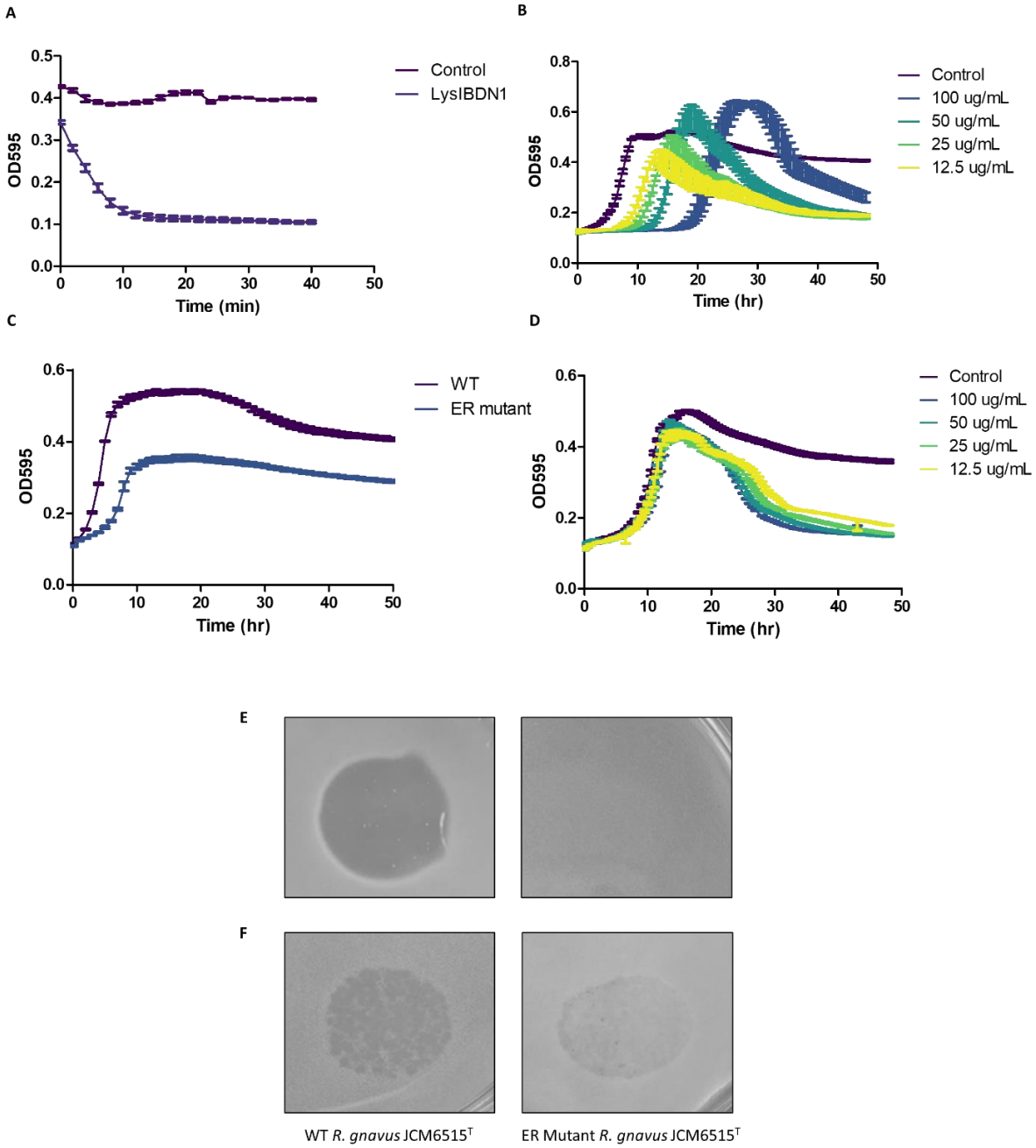


Figure 4. *In vitro* analysis of the effect of LysIBDN1 on *R. gnavus* JCM6515^T. A) Turbidity reduction assay of LysIBDN1 vs *R. gnavus*, with endolysin added to a final concentration of 100 µg/mL. **B)** Killing curve of LysIBDN1 at varying concentrations. **C)** Growth curve of isolated mutant ER5 from LysIBDN1 versus the growth of wild-type *R. gnavus*. **D)** Kill curve of mutant *R. gnavus* ER5 versus LysIBDN1 at varying final concentrations. **E)** Spot assays of LysIBDN1 on *R. gnavus* WT and resistant mutant ER5, respectively. **F)** Spot assays of phiRgIBDN1 on *R. gnavus* WT and a resistant mutant, respectively.

Characterisation of R. gnavus Mutants.

Once it was established that the regrowth of bacteria observed in the endolysin kill curves was not due to contamination but to the emergence of resistant *R. gnavus* (ER), we endeavoured to characterise these mutants and understand the underlying mechanism(s).

The growth of wild-type (WT) *R. gnavus* JCM6515^T was compared against endolysin resistant mutant 5 (ER5) that arose during treatment with LysIBDN1 (**Figure 4C**). In the absence of endolysin, the WT strain grew normally, reaching an OD_{595} of 0.4 to 0.6 after approximately 12 hours of incubation. Also, in the absence of endolysin the endolysin-resistant mutant grew at a slower rate and peaked at a lower OD_{595} of 0.2 to 0.3. The mutant strain also showed other defects in that it did not survive on average four subcultures and was also difficult to revive from frozen stocks. This growth and survival defect was observed in all mutants. A 1% inoculum of endolysin-resistant *R. gnavus* ER5 was supplemented with LysIBDN1 at a range of concentrations (**Figure 4D**). *R. gnavus* ER5 grew at a similar rate in the presence of endolysin, regardless of concentration. However, in the presence of endolysin the OD_{595} of the ER5 culture decreased rapidly in stationary phase. We also performed spot tests of endolysin LysIBDN1 (10 µL spot of neat, pure protein preparation at a concentration of 600 µg/mL) and high titre phage lysate on the WT and ER strains (**Figure 4E and 4F**). The WT strain gave a clear zone of lysis for the endolysin; however, in the case of strain ER5, no lytic activity was observed. For the phage spots, the phage plaqued normally on the WT strain. There was reduced sensitivity to the phage in the ER5 strain. To assess the ability of the mutants to withstand oxidative stress, lawns of culture were treated with discs containing 3% hydrogen peroxide. All mutant strains of *R. gnavus* had increased sensitivity to hydrogen peroxide versus the WT (Supplementary **Figure S5**).

To investigate any potential cell wall changes in the ER strains of *R. gnavus* JCM6515^T, we employed CBD binding assays. Green fluorescent protein (GFP)-CBD hybrids of endolysins were used to assess the ability of the lysins to bind parent and resistant strains of *R. gnavus* JCM6515^T (**Figure 5**). In this assay, we used GFP-CBD fusions of two endolysins, LysIBDN1 and LysH1. LysH1 is an enterococcal phage lysin with lytic activity against *R. gnavus*, as described in Chapter 3 of this thesis. Binding

assays were performed against *R. gnavus* JCM6515^T with both CBD-GFP hybrid proteins. Both WT and ER strains remained sensitive to LysH1, and this activity is reflected by the fact that the LysH1 CBD-GFP hybrid continued to bind to both parent and mutant strains. However, when we visualised the binding of the LysIBDN1 CBD-GFP fusion protein we could clearly see that the signal of fluorescence was weaker for the mutant strain than that of the WT, indicating that the *R. gnavus* mutants have a lower binding efficiency, which could account for, at least in part, their insensitivity to endolysin treatment.

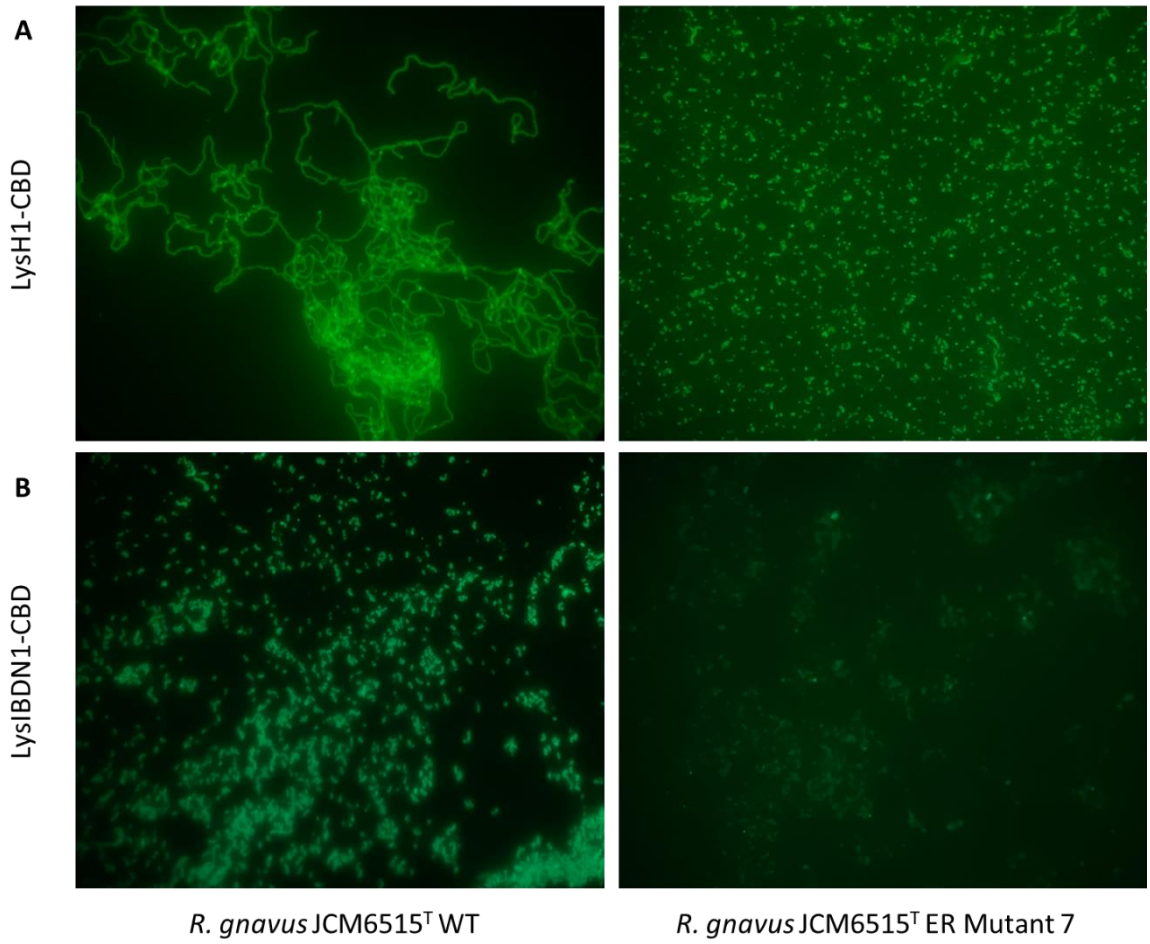


Figure 5. Fluorescence microscopy analysis of the wildtype strain and a resistant mutant of *R. gnavus* JCM6515^{T}. Fluorescence microscopy was carried out on the binding assays of WT and a representative ER *R. gnavus* strain with LysH1-CBD GFP hybrid protein **(A)** and LysIBDN1-CBD-GFP hybrid protein **(B)**.

Whole Genome Sequencing of R. gnavus ER Mutants.

Whole genome sequencing of the wildtype and endolysin-resistant strains was performed to try to establish the nature of the mutations leading to an endolysin-resistant phenotype. Both the wildtype *R. gnavus* JCM6515^T and seven mutant strains were sequenced. Their sequences were mapped against the wild-type strain, and a number of mutations were identified. These included non-synonymous SNPs, insertions, and deletions. To identify these changes, the computational pipeline Breseq was employed. Breseq identifies and annotates genetic differences relevant to a reference sequence [21]. Mutations were further assessed by multiple sequence alignments with Clustal Omega and structural analysis with AlphaFold2.3. A map of each mutant and its individual mutation(s) respective to the WT stain was prepared **(Figure 6)**. Mutations were identified in eight genes and two intergenic regions across the seven mutants. Three mutations in three separate mutants are located in the gene encoding (p)ppGpp synthetase, but each is a unique change that has occurred at different positions in the gene. The details of each mutant and its changes are shown in **Table 2**. Positions of SNPs with respect to neighbouring genes and gene clusters are illustrated in **Figure S10.** Properties of amino acids were acquired from resources at Thermo Fisher [22].

Endolysin-resistant mutant 1 (ER1) has a non-synonymous SNP in the gene encoding Inosine-5′-monophosphate (IMP) dehydrogenase, also known as *GuaB*. This is a nucleotide polymorphism that changes an A to a T and results in an amino acid change of lysine (K) to isoleucine (I) at residue 249 of a 485-residue protein. Lysine is hydrophilic and positively charged, whereas isoleucine is hydrophobic with no net charge. AlphaFold was performed on *GuaB* and the mutant protein (**Figure S8**). This allows us to visualise the site of the polymorphism within both the WT and mutant proteins. AlphaFold suggests that no change in overall protein structure would result from the mutation.

Endolysin-resistant mutants 2, 3 and 6 (ER2, ER3 and ER6) all have a unique mutation in the gene that encodes Bifunctional (p)ppGpp synthetase/guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase. From here on, we will refer to this protein as the (p)ppGpp synthetase. This is the enzyme that is responsible for producing the molecule (p)ppGpp. (p)ppGpp is a regulator of many aspects of bacterial survival,

including growth rate control, motility, gene expression, biofilm formation and antibiotic tolerance [23]. The mutations in this gene are independent deletions in both ER2 and ER3 and a non-synonymous SNP in ER6. The deletions in ER2 and ER3 occur at different positions and have different consequences for the protein structure. AlphaFold structural predictions of the impact of these deletions (**Figure S6**) illustrate dramatic changes to the integrity of the structure, particularly in the case of ER2. The deletion to ER2 results in a protein with just the first 109 amino acids (out of 776) of the C-terminal. In the case of ER3, there is a deletion at the N-terminal of the mutant (at residue 656 of 776). The SNP in ER6 is T to A, which results in an amino acid change of histidine (H) to leucine (L). The physical properties of these amino acids represent a change of a moderate hydropathy (H) to a hydrophobic residue (L). Structural analysis with AlphaFold suggests that this polymorphism has no effect on the structure of the protein **(Figure S6E)**.

Endolysin-resistant mutant 4 (ER4) had four mutations located in genes encoding an extracellular solute-binding protein, a response regulator transcription factor, a sugar ABC transporter permease, and a TetR family transcriptional regulator. These were all non-synonymous SNPs, with the exception of an insertion in the gene encoding the response regulator transcription factor. The insertion in the response regulator was assessed by structural analysis (**Figure S9**). This insertion consisted of an addition of 23 nucleotides at the N-terminal of the gene in a region that encodes a β-sheet. The response regulator is clustered with uncharacterised proteins, HAMP domain-containing histidine kinase, GerMN-containing protein and SymE **(Figure S10A)**. The mutation in the transcriptional regulator TetR resulted in a change at residue 181 of an 188 residue protein, changing threonine (T) to asparagine (N) – these amino acids have similar physical properties. This regulator is neighbouring oleate hydratase and a helix-turn-helix transcription factor **(Figure S10B).** The SNP in the gene encoding the extracellular solute binding protein of ER4 caused a change at residue 349 of an 410 residue protein, causing an amino acid change of phenylalanine (F) to valine (V) – these also carry similar physical properties. The gene encoding the ABC transporter permease had a SNP that would substitute proline (P) for serine (S) at residue 270 of a total of 459 residues. These last two genes are positioned next to each other in the genome of *R. gnavus* **(Figure S10C and D)** at genome positions 136,471 and 137,418, respectively. Extracellular solute binding domain proteins

function as recognition molecules of bacterial transport systems and provide substrates to transporters. They are also reported to act as chemoreceptors [24]. The ABC transporter ATP-binding superfamily is one of the most prominent protein families and has a range of biological functions, but their key role involves the translocation of compounds across cellular membranes [25]. These compounds can include lipids, sugars, peptides, and drugs or toxins. The ABC transporters function in both the efflux and outflux of molecules in bacteria [26]. It has been reported that bacterial ABC transporters can mediate multidrug resistance, acquisition of essential nutrients and toxin secretion [27].

In endolysin-resistant mutant 5 (ER5) there was an insertion of 1nt (C) upstream of a gene encoding a Gcn5-related N-acetyltransferases (GNAT) family N-acetyltransferase. This insertion occurred at the start of the contig, ahead of the genes encoding a GNAT family N-acetyltransferase and a hypothetical protein.

Together with the already described change to the gene encoding the (p)ppGpp synthetase, endolysin resistant mutant 6 (ER6) also had a deletion in a gene encoding a HIRAN domain-containing protein. To further annotate this gene, we searched with NCBI Blast and Foldseek [28]. Both searches resulted in a hypothetical/uncharacterised protein. AlphaFold analysis predicts a missing domain at the N-terminus of the protein (**Figure S7**).

Endolysin-resistant mutant 7 (ER7) has an intergenic insertion between genes encoding a Trypsin-like serine protease and a hypothetical protein.

Endolysin-resistant mutant 8 (ER8) has a SNP in the gene encoding a pyruvate, phosphate dikinase, with a T to an A resulting in an amino acid change of methionine (M) to lysine (K) at position 241 – a change to a hydrophilic amino acid.

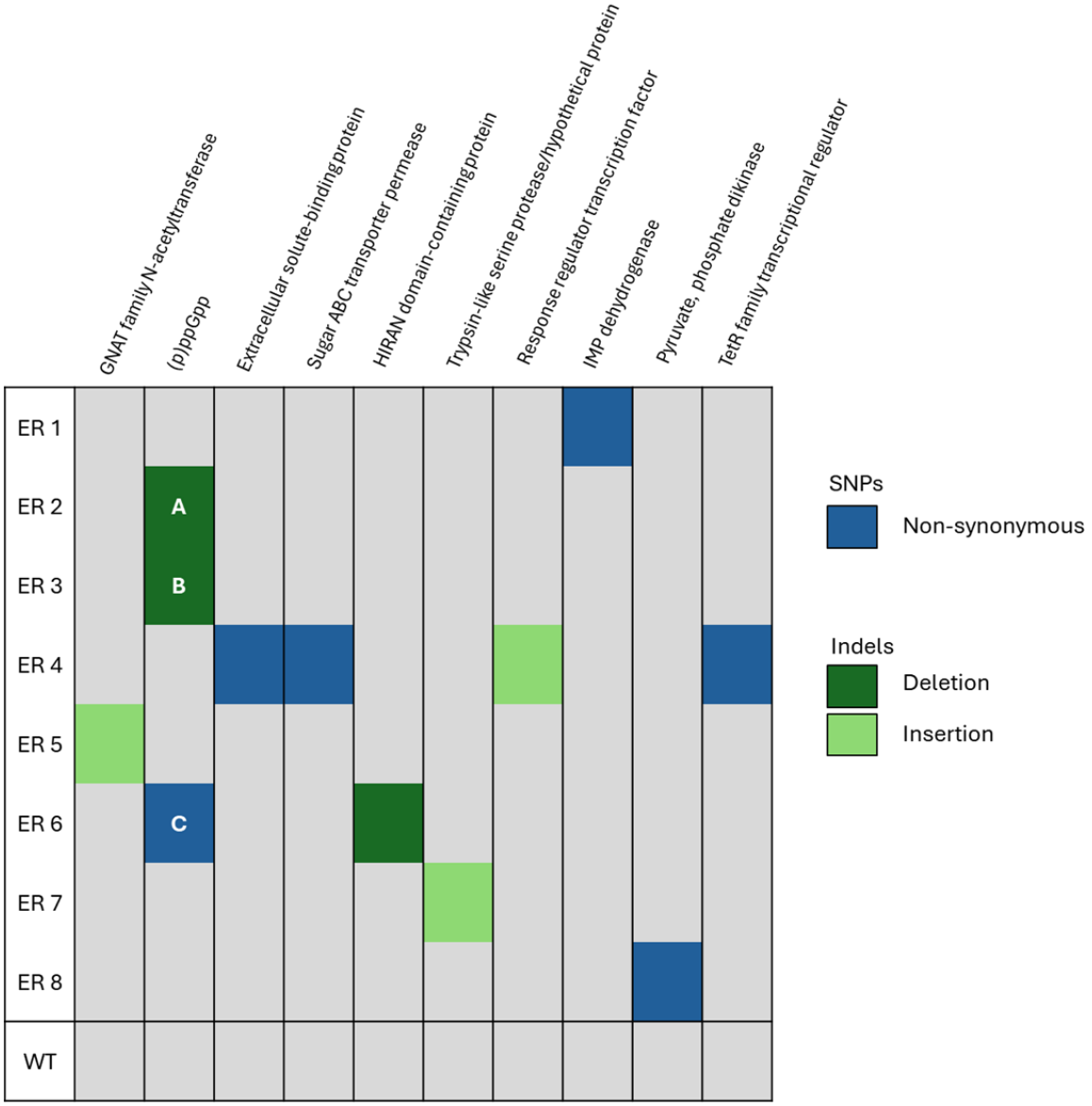


Figure 6. Map of mutations in endolysin-resistant (ER) strains of *R. gnavus* versus the WT. Non-synonymous SNPs, deletions and insertions are coloured as per the legend. A, B and C denote different changes in the same gene.

Table 2. Mutations occurring in the genomes endolysin-resistant mutants of *R. gnavus* JCM6515^T. This table outlines the unique changes that have been reported by Breseq when mutant genomes were mapped against the WT.

Mutant	Gene Product	Type	Gene Position	Ref Sequence	New Sequence	Gene Name
ER1	IMP dehydrogenase	SNP	881	A	T	*guaB* DAGEDA_09045
ER2	(p)ppGpp synthetase	DEL	coding (137/2307 nt)	A		DAGEDA_00530
ER3	(p)ppGpp synthetase	DEL	coding (1970/2307 nt)	G		DAGEDA_00530
ER4	Extracellular solute-binding protein	SNP	1048	T	G	DAGEDA_08520
	Response regulator transcription factor	INS	coding (670/693 nt)	G	G	DAGEDA_13540
	Sugar ABC transporter permease	SNP	808	C	T	DAGEDA_08525
	TetR family transcriptional regulator	SNP	542	G	T	*tetR* DAGEDA_07110
ER5	GNAT family N-acetyltransferase	INS	intergenic (/-81)	C	C	DAGEDA_17210
ER6	(p)ppGpp synthetase	SNP	1181	T	A	DAGEDA_00530
	HIRAN domain-containing protein	DEL	coding (476/507 nt)	G		DAGEDA_09010
ER7	Trypsin-like serine protease/hypothetical protein	INS	intergenic (+205/-9)	G	G	DAGEDA_11295 DAGEDA_11300
ER8	Pyruvate, phosphate dikinase	SNP	722	T	A	*ppdK* DAGEDA_00415

Discussion

In this study, we cloned genes encoding five novel endolysins from the genomes of temperate *R. gnavus* phages. These endolysins are all predicted to have an amidase mode of action. The resulting purified proteins have all been shown to have lytic activity against their target *R. gnavus* and have limited off-target effects against other strains tested. LysIBDN1 was further characterised with the aim to be tested as a novel therapeutic for IBD. The ability of this endolysin to effectively lyse and kill *R. gnavus* could make LysIBDN1 a promising candidate for the management of *R. gnavus* blooms associated with IBD. To our surprise, we observed the emergence of endolysin-resistant mutants of *R. gnavus* JCM6515^{T} against all five endolysins. The addition of endolysin to the wildtype strain resulted in an initial suppression of the bacterium, followed by the outgrowth of resistant mutants. These mutants were biologically characterised as being tolerant to endolysin treatment on plates and in liquid media, have reduced sensitivity to the parental phage, and we also showed that the endolysin has a reduced ability to bind these cells. These traits are all consistent with an altered cell wall structure. Whole genome sequencing revealed a number of polymorphisms across the genomes of resistant mutants, some of which occurred in genes associated with the bacterial stress response, transport, and metabolism. To our knowledge, this is the first reported case of endolysin-resistant mutants arising from a single dose of endolysin treatment. It is noteworthy that all of the endolysin-resistant mutants have significant growth defects and would be unlikely to thrive in a competitive situation such as that found in the gut microbiome.

Bioinformatic analysis of the endolysin genes and their products was performed to assess the similarity and structure of the enzymes. Multiple sequence alignments and a percentage identity matrix confirmed that there are two clusters of *R. gnavus* endolysins. While both groups possess an amidase mode of action, they belong to different amidase families. Lys3-1 encoded by phiRg507T2/3 is dissimilar to the other lysins, with less than 20% amino acid similarity to the others. Multiple sequence alignments highlighted conserved regions at the N-terminal of the proteins. The majority of differences occurred in their C-terminal regions, suggesting differences in their binding domains. This is supported by HHpred analysis and structural predictions of Lys3-1 and LysIBDN1.

While the CBD-GFP fusion protein of LysIBDN1 allowed for the characterisation of resistant mutants, it also served as a means to functionally characterise the CBD domain of LysIBDN1. As these are the first characterised endolysins from *R. gnavus* phages, this CBD is also the first reported CBD from an *R. gnavus* lysin. Homology searches revealed that this C-terminal domain had similarities to those of *Enterococcus* phage endolysins. However, no lysis was observed against *E. faecium* and *E. faecalis* strains. Interestingly, as outlined in **Chapter 3** of this thesis, LysH1 is an endolysin derived from an enterococcal phage that has activity against both *E. faecium* and *R. gnavus*, and its CBD-GFP hybrid protein shows efficient binding to both species. The bioinformatic assessment also revealed that LysH1 has homology to *R. gnavus* endolysins. It is likely that these bacteria have similarities in their cell wall composition that allow for this shared recognition.

Analysis of the antibacterial activity of the recombinant endolysins indicated a narrow host range. The endolysin encoded by φIBDN1 was selected as the most promising candidate according to its performance in agar-based tests, its spectrum of activity, tests in liquid culture and its expression efficiency. Resistance seemed to occur in a dose-dependent manner, with mutants arising earlier at 25 µg/ml and later at 100 µg/mL. The mutation frequency was determined to be ~$1x10^{-5}$. Mutation frequency can be influenced by a number of factors ranging from oxidative stress and selective pressure to intrinsic mutability – some regions of a bacterial genome may be more prone to mutations than others. The mutation frequency in the context of antimicrobial resistance can be defined as the *in vitro* frequency at which detectable mutants arise in the bacterial population in response to treatment with an antibiotic or, in this case, an endolysin [29]. As already mentioned, the isolated mutants had significant fitness costs associated with their endolysin resistance. The resistant strains were severely affected in terms of cell growth and proliferation; they were unable to undergo multiple passages, had a reduced ability to grow on solid agar and did not always withstand storage at -80 ^{o}C.

The emergence of resistance to endolysins is not a commonly reported observation. Endolysins are generally renowned for a lack of associated resistance mechanisms. This is largely attributed to their rapid action and the hypothesis that because phages and bacteria have evolved in parallel, to avoid the phage being trapped inside the host endolysins target essential molecules in the cell wall, making

resistance unlikely [30] [31]. Attempts have been made to generate resistance to several endolysins, but none have displayed high levels of resistance to date. Endolysin Pal was isolated in 2001 from the streptococcal phage Dp-1. The authors exposed bacteria to Pal at increasing concentrations in liquid culture, but no resistant strains were recovered [32]. LysH5 is an endolysin targeting *Staphylococcus aureus*. A 2013 study assessed the risk of staphylococcal resistance emerging from treatment with LysH5 [33]. In this study, *S. aureus* was exposed to 10 rounds of endolysin in liquid and plate cultures, but no resistance was observed. *Streptococcus pyogenes* endolysin PlySs2 was subjected to similar tests [34]. *S. pyogenes* and *S. aureus* strains were treated with increasing concentrations of PlySs2, but none of the strains tested displayed more than a two-fold increase in the MIC.

Through whole genome sequencing, we identified a number of SNPs, insertions, and deletions across seven endolysin-resistant strains of *R. gnavus*. The genes associated with these changes are linked to the bacterial stress response, transport systems and gene regulation. ER2, ER3 and ER6 had mutations in the gene encoding (p)ppGpp synthetase. ER1 had a SNP in the gene encoding inosine monophosphate (IMP) dehydrogenase, which is also called *GuaB* and is associated with the (p)ppGpp pathway [35]. (p)ppGpp is a mediator of the stringent response. We observed changes in this gene at three unique positions in three different mutants. In gram-positive bacteria, (p)ppGpp synthetases are usually homologues of RelA/SpoT. Mutations to these enzymes could cause a range of downstream changes to the (p)ppGpp pathway by either upregulating or disrupting synthesis of (p)ppGpp. A number of studies have created (p)ppGpp deficient mutants to observe the physiological challenges of bacteria in the absence of (p)ppGpp. A 2014 study showed that *B. subtilis* lacking (p)ppGpp as a result of mutations to synthetase genes demonstrated growth defects and amino acid auxotrophy [36]. A comprehensive review published in 2021 documents how (p)ppGpp-deficient bacteria respond to stress and can continue to grow for a finite amount of time even when carrying this defect [37]. It is reported from a survey of publications in this review that deficient strains can have loss of membrane integrity and an increase in sensitivity to oxidative stress. While these cells can continue to proliferate, the absence of (p)ppGpp to regulate pathways results in an accumulation of nonfunctional proteins and dysregulation of purine synthesis. (p)ppGpp deficient cells have poor long-term

viability which ultimately ends in cell death. The mutant strains isolated in our study present a similar phenotype, by which they can proliferate for a small number of passages with impaired growth rate and poor survival characteristics. Despite attempts to continue passages, the cells are unable to be recovered.

GuaB is responsible for the conversion of IMP to xanthosine monophosphate (XMP) in the GTP biosynthesis pathway (**Figure S11**). This pathway is regulated by (p)ppGpp and is essential for growth and amino acid starvation survival [38]. The SNP in the *GuaB* gene could be disrupting or upregulating the GTP biosynthesis pathway of *R. gnavus*. In structural analysis, we can see that the SNP does not confer major changes in protein structure, but the site of the SNP is nested within the protein and could be involved in the enzymatic activity of *GuaB*. In *B. subtilis*, suppressed GTP levels result in slowed growth and stress resistance, however an increase in GTP levels above a certain threshold can result in a cytotoxic effect. A study characterised the relationship between stress survival and GTP in *B. subtilis* and concluded that (p)ppGpp defective mutants could survive a stress event but at the expense of normal cell proliferation [39].

In recent years (p)ppGpp has become increasingly associated with antimicrobial resistance. Studies have reported that point mutations to the synthetase enzyme Rel can impair its hydrolase activity which leads to an overproduction of (p)ppGpp and results in antibiotic resistance [40]. Once produced, (p)ppGpp results in the downregulation of active growth and upregulation of genes involved in stress response [41]. In the *Bacillota* (formerly *Firmicutes*), the stringent response is mediated by (p)ppGpp and supports the development of antibiotic resistance, tolerance, and survival under challenging conditions [42]. Resistance to antibiotics is generally associated with an increase in levels of (p)ppGpp. These can occur as a result of upregulation of expression of (p)ppGpp synthetase, deletions to the Rel genes, or point mutations to Rel. Examples of such include *S. aureus* resistance to β-lactam antibiotics, *B. subtilis* decreased MICs of a number of drugs, and *E. faecalis* decreased MIC for vancomycin [43] [44] [45]. In one example, a missense mutation in *relA* in vancomycin-resistant *E. faecium* (VRE) resulted in the activation of the stringent response and elevated levels of (p)ppGpp. While the mutation resulted in a tolerance to high doses of antibiotics, this tolerance came with a fitness cost that resulted in smaller colony morphology and less well-populated biofilms [46]. Due to

structural differences in WT and ER (p)ppGpp synthetase in mutant strains that had deletions, it is likely that ER2 and ER3 are no longer producing (p)ppGpp. In mutant strains, ER6 and ER1, we see SNPs in the synthetase and *GuaB* respectively, however, it is yet unclear whether these SNPs result in defective enzymes or an increase in their function. ER6 also had a structural change to a HIRAN domain-containing protein, but we propose that the mutation leading to the endolysin-resistant phenotype is more likely to be the SNP in (p)ppGpp synthetase.

Endolysin resistant mutant ER4 had a number of changes, two of which occurred in genes encoding regulators (a response regulator transcription factor and TetR family transcriptional regulator) and the remaining two in genes encoding a Sugar ABC transporter permease and Extracellular solute-binding protein. The gene encoding the TetR family transcriptional regulator is clustered with a gene encoding oleate hydratase. This is a fatty acid hydratase that can facilitate detoxification of a bacterial environment [47]. An insertion to the response regulator transcription factor could have an effect on the gene cluster this is regulating. This regulator is clustered with a HAMP domain-containing protein. HAMP domain-containing proteins are reported to be found in bacterial sensors and chemoreceptors [48]. In ER4, SNPs are found in the same gene cluster encoding a Sugar ABC transporter permease and an Extracellular solute-binding protein. ABC transporters are documented to be associated with antimicrobial resistance [49] [50] [51]. The ABC family transporters can protect bacteria from antimicrobials by the use of efflux pumps to maintain homeostasis and expel toxic molecules. Extracellular solute-binding proteins are also chemoreceptors that can recognise bacterial transport systems [52]. Their main function is to provide substrates to the transporter systems [53]. It is possible that the SNPs in these two genes are conferring a protective effect to the mutant by maintaining homeostasis within the cell during treatment with endolysin.

Mutant strain ER5 only had a detectable change in the form of a 1nt insertion in an intergenic region positioned ahead of a gene encoding a GNAT family N-acetyltransferase. Mutations at intergenic regions could have an impact on cell physiology because non-coding DNA may still be involved in the regulation of nearby genes. These non-coding regions can act as regulatory elements such as enhancers, silencers, and promoters that control gene expression. It is possible that the changes

occurring in these regions in our strains could potentially alter the expression levels of nearby genes.

Resistant mutants ER7 and ER8 had mutations to genes encoding a Trypsin-like serine protease and Pyruvate phosphate dikinase, respectively. Trypsin-like serine protease is a proteolytic enzyme. While these kinds of enzymes are widespread in nature, their functional associations are largely unknown [54]. Pyruvate, phosphate dikinase (PPDK) is responsible for the catalysis of ATP and a pyruvate to AMP and phosphoenolpyruvate [55].

The genomic analysis of endolysin resistant mutants of *R. gnavus* has shed some light on the possible mechanisms of resistance to LysIBDN1, while also leaving some mysteries. SNP analysis with Breseq highlighted some key pathways such as (p)ppGpp and transporter systems. While not every mutation could be connected with the observed phenotype, we believe that the associations made here are a promising start to investigate the means by which this phenomenon occurs. It appears beyond doubt that mutations affecting the stringent response are involved in endolysin resistance. It is also possible that events such as phase variation as a result of inversions may have been overlooked by the programs used in this analysis.

While the development of resistance to endolysins is a concerning development for endolysin therapy, the associated fitness costs make it unlikely that they would proliferate at high enough levels to maintain a disease state. There are potentially a number of avenues we could explore to overcome this resistance. These could include the creation of chimeric lysins or the application of a cocktail of lysins that have varying mechanisms of action, as well as using endolysins in synergy with other antimicrobials [56]. Bacterial resistance is less likely to evolve in response to endolysins with multiple catalytic domains as they would have to evolve resistance to two distinct cleavage sites. A number of chimeric lysins have been developed in recent years that have high antimicrobial activity against multidrug-resistant strains [57] [58]. Future work on our endolysins could include using their existing cell wall binding domains paired with new novel catalytic domains to create a next-generation endolysin that can overpower resistance mechanisms in *R. gnavus*. Endolysins are in clinical trials and could soon enter clinical use; therefore, understanding the risk of resistance development is of the utmost importance. While endolysins are generally promoted as

potential alternatives to antibiotics, mainly targeting multi-drug resistant pathogens, and praised for their lack of resistance mechanisms, it is clear that drug candidates such as these should be rigorously checked for the emergence of resistance.

In conclusion, we have isolated five novel endolysins targeting *R. gnavus* JCM6515^T with good lytic activity. Further characterisation revealed the emergence of endolysin-resistant strains of *R. gnavus* JCM6515^T. Genomic analysis revealed a number of mutations in the genomes of these resistant strains that were associated with the stringent stress response, regulation of growth and transport systems. Phenotypically, these mutants displayed resistance to varying doses of endolysin treatment and reduced sensitivity to the parental phage. However, these survival mechanisms came at a significant cost to the fitness of the bacteria. The emergence of resistance, regardless of fitness costs, must be taken into account when evaluating the potential of endolysins as therapeutics, but in terms of basic science, the mutants have provided us valuable insights into the potential underlying mechanisms.

Materials and Methods

Bacterial Strains and Growth Conditions.

Strains used in this study were stored at - 80 ^{o}C and cultivated at 37 ^{o}C. *R. gnavus* strain JCM6515^{T} was obtained from the Japan Collection of Microorganisms (alternative name ATCC 29149) and was cultured in Anaerobe Basal Broth (ABB) (Oxoid). Recombinant *E. coli* strains were grown in the same medium but with differing antibiotics due to the requirements of the plasmids. Lys507/3-1/BL21(DE3) was maintained in Lysogeny Broth (LB) broth supplemented with 50 µg/mL Kanamycin Sulphate. LysIBDN1/M15[pREP4], Lys507/2-1/ M15[pREP4], Lys519/ M15[pREP4], LysPS6-1/ M15[pREP4] were cultured in Lysogeny Broth (LB) broth supplemented with 50 µg/mL Kanamycin Sulphate and 100 µg/mL Ampicillin (Sigma-Aldrich).

R. gnavus CCUG 54531, CCUG 33437, CCUG 57161, CCUG 49994, CCUG 57208, CCUG 51289, CCUG 57137, CCUG 52279, CCUG 52279, PS/160 were cultured in Anaerobe Basal Broth (Oxoid). *Faecalibacterium prausnitzii* 918/95B, 922/41-1, DSM 17677 and *Agathobacter rectalis* 17629 were cultivated in Yeast Extract, Casitone and Fatty Acid Agar (YCFA-GSCM). *Clostridium scindens* DSM 5676, *Bacteroides intestinalis* 919/174, *Bacteroides distasonis* DSM 20701 and *Anaerostipes hadrus* 942/1 were cultured in Fastidious Anaerobe Broth (Oxoid). These strains were all grown under strict anaerobic conditions (Type A vinyl anaerobic chamber, Coy Labs).

Bioinformatic Analysis.

Phage genomes were annotated with Pharokka v1.7.1. with the default settings. BLASTp (2.11.0+) was used to blast all-vs-all proteins, and the synteny plot with blast links was plotted using Gggenomes (v0.9.12.9) and R v4.3.0 [59]. Highlighted arrows in **Figure 2 (A)** show the location of endolysin genes that were cloned and recombinantly expressed. The genomes of *R. gnavus* phages can be found at the following GenBank accession numbers - phiRg507T2/2: MT980836, phiRg507T2/3: MT980837, phiRg519T2: MT980838, phiRgPS6: MT980839, phiRgIBDN1: MT980840.

Homology searches were conducted for *R. gnavus* endolysins in HHpred [60]and InterPro Scan [61]. Comparative analysis was performed on the endolysins to

assess their similarity. Multiple sequence alignments were generated using Clustal Omega [62]and visualised using SeaView4 [63]. Percentage Identity Matrix of endolysins was also generated with Clustal Omega on an amino acid level, and the resulting matrix was visualised with Heatmap2 on the Galaxy EU platform. For structural analysis of endolysins and cell wall binding domains, AlphaFold v2.3.1 was used to predict the structures with the settings "--enable_gpu_relax=true --model_preset=monomer --use_gpu=true" [18].

Cloning.

Endolysin genes were amplified from φRgIBDN1, φRg507T2 2-1, φRg507T2 3-1, φRg519T2 and φPS6-1 directly from phage lysates by PCR amplification using the primers listed in the supplementary materials in **Table S2**. The PCR products were purified using the GeneJET PCR Purification Kit. For the endolysin 507/3-1, derived from φRg507T2 3-1, the pET28b(+) expression vector was selected. The remaining endolysins were cloned into the pQE-30 expression vector. To construct the expression plasmid, PCR products and expression vectors were digested by Thermo Fast Digest restriction enzymes BamHI and HindIII. The fragments were again purified by GeneJET PCR Purification kit before carrying out ligation using Fast Digest T4 Ligase (Thermo Fischer). The ligation mixes were introduced into TOP10 *E. coli* competent cells (Invitrogen). Screening for positive clones was carried out by PCR. The clones containing the expected plasmid with the endolysin gene were extracted using GeneJET Plasmid Miniprep Kit (Thermo) and sent for Sanger Sequencing. The viable plasmids were introduced into competent cells for recombinant expression. The plasmid containing lysin 507/3-1 was introduced into *E. coli* BL21(DE3) and the remaining plasmids were introduced into M15[pREP4] *E. coli*.

The putative cell wall binding domain of LysIBDN1 with a GFP fusion construct was created using NZYEasy Cloning & Expression Kit IX (NZYtech), which created a construct in plasmid pHTP9 that contains GFP as a fusion tag. This construct was introduced into *E. coli* Trans-alpha competent cells. Transformants were screened for positive clones, which were propagated for plasmid extractions. Extracted plasmids positive for the insert were introduced into One Shot BL21(DE3) chemically competent cells (Invitrogen) for protein expression. Sanger sequencing confirmed the integrity of the construct. The CBD-GFP fusion construct of LysH1-CBD-GFP used for the

characterisation of resistant mutants has been isolated, as described in **Chapter 4** of this thesis.

Overexpression in Recombinant E. coli and Protein Purification.

Recombinant *E. coli* cells were grown in LB media containing 50 µg/mL Kanamycin Sulfate for Lys507/3-1/BL21(DE3) and 50 µg/mL Ampicillin (Sigma Aldrich) for the remaining strains. Cultures were grown to OD_{600} 0.5 – 0.8. Cultures were induced with isopropyl-β-D-thiogalactopyranoside (IPTG) (Sigma Aldrich) at a final concentration of 100 mM. Cultures were left to incubate overnight, shaking, at room temperature. Cell pellets were harvested (4000 rpm, 30 minutes at 4^{o}C), and the resulting pellets were resuspended in Binding Buffer, 1 mL/g. Cells were sonicated on ice (80% power, 50% duty factor, 10-minute duration). Protein purification was carried out using affinity tag purification. His GraviTrap columns (GE Healthcare) were used with the buffer compositions in supplementary **Table S3**. All fractions of the purification process were analysed by SDS-PAGE using Bolt 4–12% Bis-Tris gels (Invitrogen). Imidazole was removed from the eluate using Zeba™ Spin Desalting Columns (Thermo Fischer). Protein concentrations were quantified using a Qubit Protein Assay Kit. Aliquots of purified protein were stored in a final concentration of 20% glycerol at – 80 ^{o}C for extended storage.

Lytic Activity of Endolysins.

The lytic activity of Lys507T/3-1 against *R. gnavus* JCM6515^{T} was tested by turbidity reduction assays. An overnight culture of *R. gnavus* JCM6515^{T} was prepared. The cells were exposed to oxygen to inactivate the culture and were harvested by centrifugation (4,000 rpm, 15 min) before washing with PBS and resuspended to a final volume OD of 1.0 at a 1 cm path length. Turbidity reduction was measured by mixing 100 µL of cell suspension with 20 µL of Lys507T/3-1 at varying concentrations in a 96-well plate. The assay was also carried out with the addition of 10 mM $CaCl_2$ and 10 mM $MgSO_4$. Changes in optical density (OD_{595}) were measured at 3 min intervals using a microtiter plate reader (Thermo Fisher) at 37 ^{o}C for 1 hour.

Spot assays were performed on several bacterial strains to determine the spectrum of activity of purified endolysins. These included various strains of *R. gnavus* and other relevant gut-associated bacteria. 100 µL of the test culture was inoculated into 4 mL of its growth media with 0.4% agarose. Depending on the strain

requirements, these lawns were incubated at 37 ^{o}C for 24 hours aerobically or anaerobically. 5 µL of purified protein sample was spotted on the preformed bacterial lawn. Plates were incubated and observed for lysis the following day.

Generation of Endolysin-Resistant Mutants.

Mutants of endolysin-resistant mutants were generated from treatment with endolysin LysIBDN1. An overnight culture of *R. gnavus* JCM6515^{T} was prepared. In a 24-well plate, 1mL semi-solid agar (0.4% agar) was inoculated with 1% *R. gnavus* JCM6515^{T}, and endolysin was added in a final concentration of 100-12.5 µg/mL. Each well received a set concentration of endolysin. Control wells with buffer only and media only were also prepared. The 24-well plate was left to incubate for up to 72 hours at 37 ^{o}C under anaerobic conditions. Each day, single colonies that appeared in the agar were picked and purified three times in the presence of endolysin. These pure strains were brought forward for spot assays (as previously described) to test sensitivity to the endolysin. Strains that were insensitive to endolysin treatment were sent for 16S rRNA sequencing, and their identity was also confirmed by PCR with *R. gnavus*-specific primers (these are listed in supplementary **Table S2**). Once it was confirmed that they were indeed endolysin-resistant strains of *R. gnavus,* they were sent for whole-genome sequencing.

Sequencing of Mutants and Analysis of Genomes.

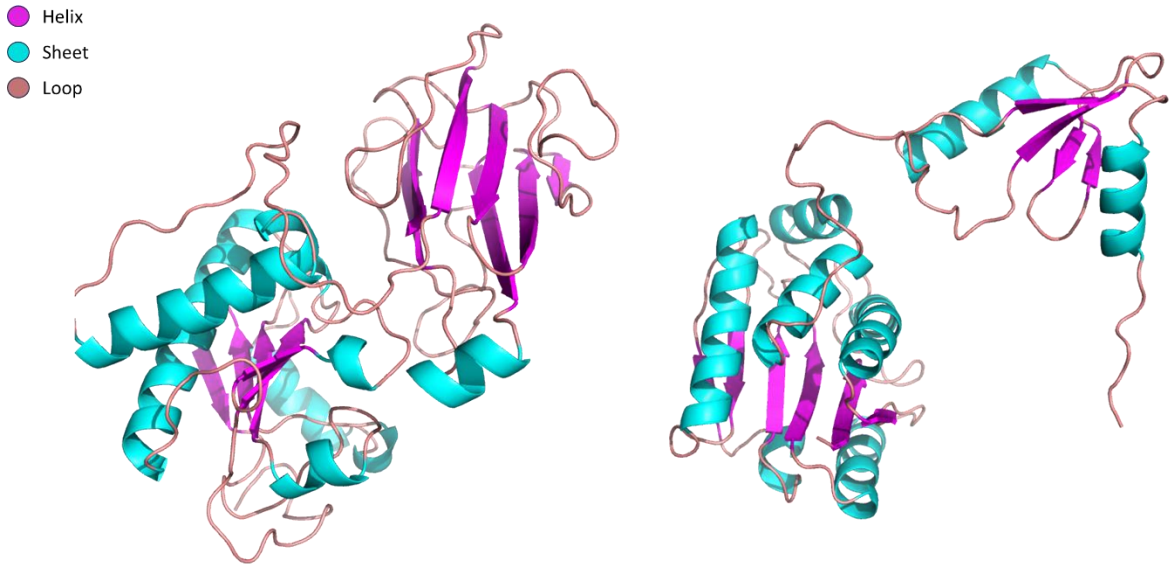
Overnight cultures of wild-type and mutant strains of *R. gnavus* were prepared, and 1 mL of culture was harvested the following day. Cell pellets were collected by centrifugation at 4,000 xg for 5 minutes. DNA was extracted from the samples with the GenElute™ Bacterial Genomic DNA Kit (Sigma Aldrich), following the manual for gram-positive extractions. Extracted samples were quantified by Qubit dsDNA quantification kit (Invitrogen) and normalised to a final concentration of 20 ng/µL for sequencing. Sequencing was conducted by Novogene (UK). The genomic DNA of each sample was randomly sheared into short fragments of ~350 bp. The obtained fragments were subjected to library construction using the NEBNext DNA Library Prep Kit. Sequencing was carried out on their Illumina NovaSeq X Plus, using a 10B flow cell and a 300-cycle reagent kit (cat. 20085594).

Quality control and trimming were performed on raw reads using fastp v0.23.2 [64]. Genomes were assembled from trimmed reads using spades v3.15.5 with the

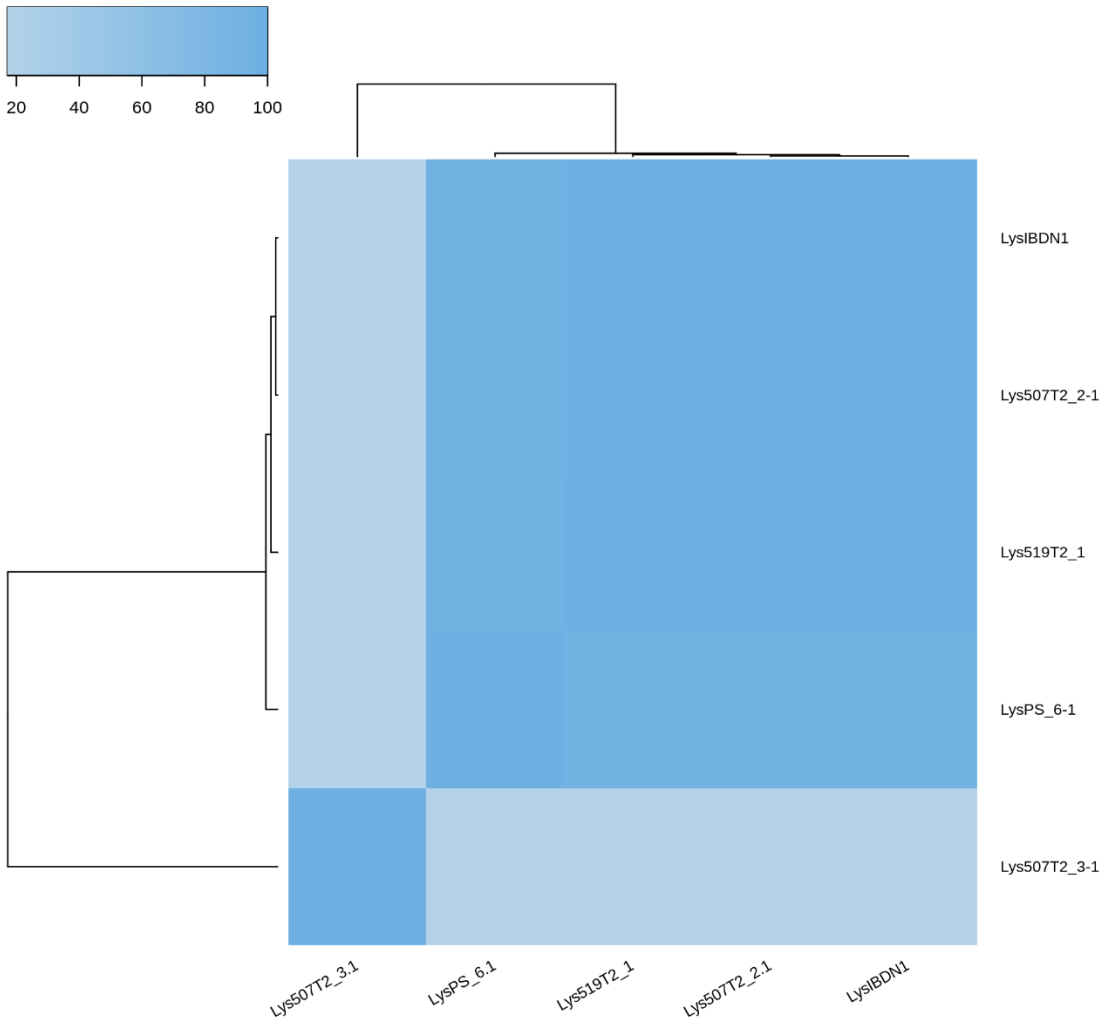
parameter “--isolate” [65]. Genomes were annotated using bakta v1.9.3 using database v5 [66]. SNPs, inserts and deletions were detected using breseq v0.37.1 [67]. The breseq command gdtools was used to tabulate the data with the options “ANNOTATE -f TSV” and this was imported into R v4.3.0, tabulated and exported. Trimmed reads were also aligned to the WT genome using bowtie v2.5.2 with the options “--phred33 --end-to-end”. Samtools v1.13 was used to create and sort bamfiles using the commands view, sort, index, flagstat and faidx. The indexed bam file was manually inspected using Interactive Genome Viewer (IGV) v2.17.4 [67].

Fluorescence microscopy.

The binding of LysIBDN1 CBD to *R. gnavus* wild-type and endolysin-resistant mutant strains was demonstrated through the following methods. Bacterial strains were grown to the mid-log phase. Cells were harvested by centrifugation at 3,000 xg for 5 minutes. Cells were washed three times in PBS (phosphate-buffered saline) and resuspended to an OD_{600} of 0.6. From this suspension, 200 µL cells were combined with a final concentration of 50 µg/mL of purified protein and shaken at 200 rpm at room temperature for 45 minutes. This cell suspension was also harvested by centrifugation at 3,000 xg for 5 minutes and washed thrice in PBS to remove unbound protein. Cells were resuspended to a volume of 200 µL, and 10 µL was added to a microscope slide, left to dry and secured with a coverslip. For fluorescence microscopy, cells were viewed with a Zeiss Confocal Microscope at 63X magnification and excited with a green fluorescent laser at 458 nm.



Supplementary Figure S1. Secondary structure prediction of Lys3-1 and LysIBDN1, respectively. These structures were predicted with AlphaFold and visualised with PyMOL. The structures are coloured based on their secondary structure predictions, as shown in the figure legend.



Supplementary Figure S2. Similarity of Endolysins. Percentage identity matrix of *R. gnavus* endolysins on an amino acid level. Percentage identity matrix generated with Clustal Omega and visualised with Heatmap2 on the Galaxy Europe platform.

Table S1. Details of Recombinantly Expressed Endolysins.

Endolysin	Parental φ	Mode of Action	Vector	Host *E. coli*	Size (AA)	Weight (kDa)
3-1	Rg507T2/3	Amidase 3	pET28b+	BL21(DE3)	342	36.03
IBDN1	RgIBDN1	Amidase 2	pQE30	M15 [pRep4]	267	28.83
2-1	Rg507T2/2	Amidase 2	pQE30	M15 [pRep4]	267	28.81
519	Rg519T2	Amidase 2	pQE30	M15 [pRep4]	267	28.78
PS6	RgPS6	Amidase 2	pQE30	M15 [pRep4]	268	29.03

Table S2. Primer sequences for the amplification of Endolysin genes and colony PCR.

Primer Name	Forward Sequence	Reverse Sequence
LysIBDN1	5'-GTCGGATCCATGAAAATTGGCTTAAGGG-3'	5'-ATTAAGCTTCACCCCTCAGAGAGG-3'
Lys2-1	5'-GTCGGATCCATGAAAATTGGCTTAAGGG-3'	5'-ATTAAGCTTCACCCTCAGAGAACTGATC-3'
Lys3-1 (for PQE30)	5'-TTAGGATCCCATGAGTATTTGTCG-3'	5'-ATTAAGCTTAGAACAACTGGAACCGATCG-3'
Lys3-1 (for pET28b(+))	5'-TTAGGATCCGATGAGTATTTGTC-3'	5'-ATTAAGCTTAGAACAACTGGAACCGATCG-3'
Lys519	5'-GTCGGATCCATGAAAATTGGCTTAAGGG-3'	5'-ATTAAGCTTCACCCTCAGAGAACTGATC-3'
LysPS6	5'-TTAGGATCCATGAAAATCGGCTTAAGGG-3'	5'-ATTAAGCTTCACCCCTCAGAGAACCGAT-3'
***R. gnavus* specific primers**	5'-GCGTGCTTGTATTCCGGATG-3'	5'-GCCTGAACAGTTGCTTTCGG-3'

Table S3. Protein purification buffer compositions for gravity column purification.

Endolysin	Binding Buffer	Wash Buffer	Elution Buffer
Lys3-1, LysIBDN1, Lys519, LysPS6.	50 mM Sodium Phosphate pH 7.5, 300 mM sodium chloride, 10 mM Imidazole, 0.03% Triton X-100.	50 mM Sodium Phosphate pH 7.5, 300 mM sodium chloride, 50 mM Imidazole, 0.03% Triton X-100.	50 mM Sodium Phosphate pH 7.5, 300 mM sodium chloride, 250 mM Imidazole.
Lys2-1.	50 mM MES Solution, 300 mM sodium chloride, 10 mM Imidazole, 0.03% Triton X-100.	50 mM MES Solution, 300 mM sodium chloride, 50 mM Imidazole, 0.03% Triton X-100.	50 mM MES Solution, 300 mM sodium chloride, 250 mM Imidazole.

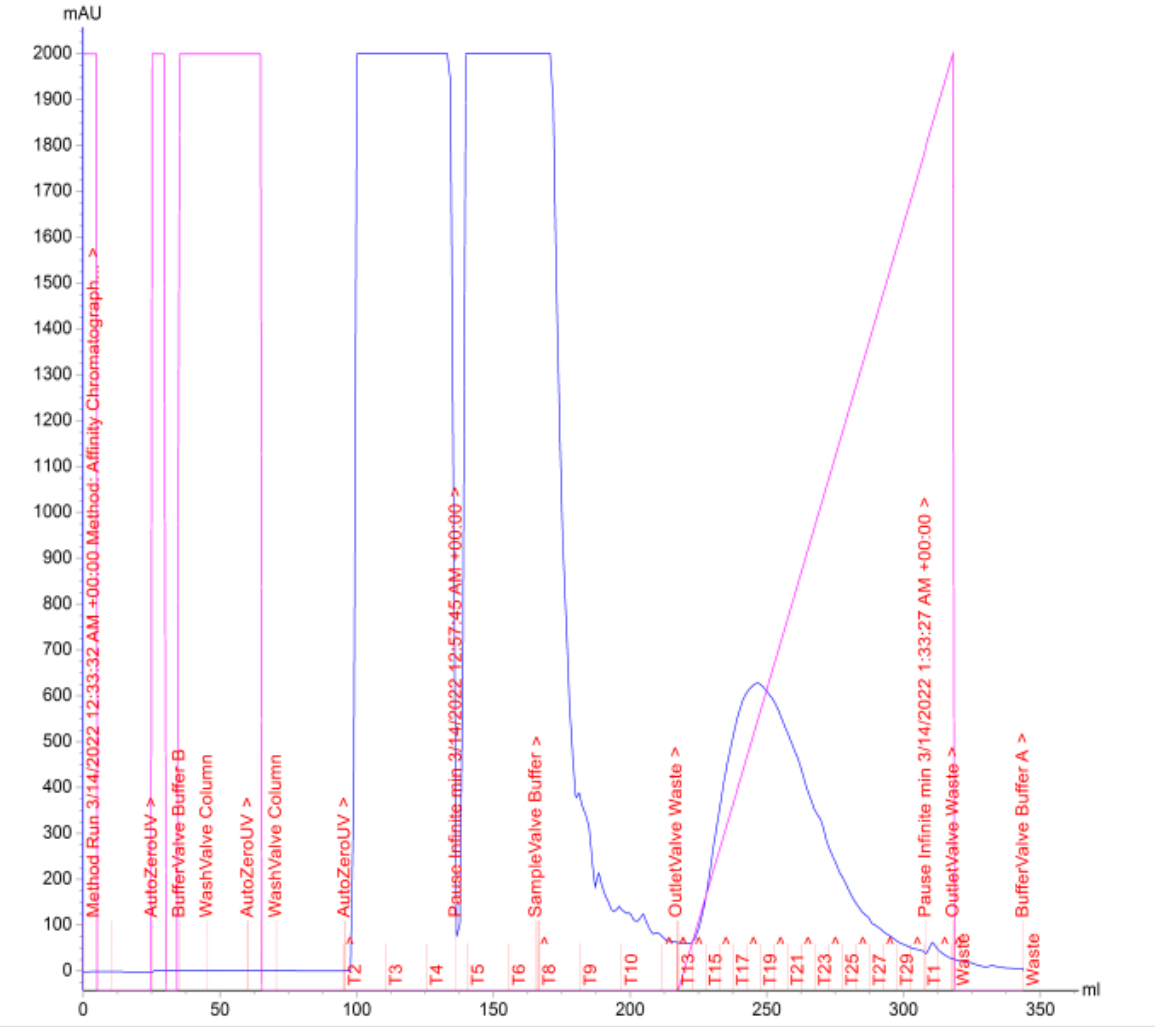


Figure S3. Chromatogram of the affinity chromatography purification of endolysin LysIBDN1. This chromatogram was generated with the Unicorn software and shows a peak of the endolysin coming off the Cytiva Talon Crude column between 225 mL and 300 mL, with a peak UV absorbance (mAU) of 600.

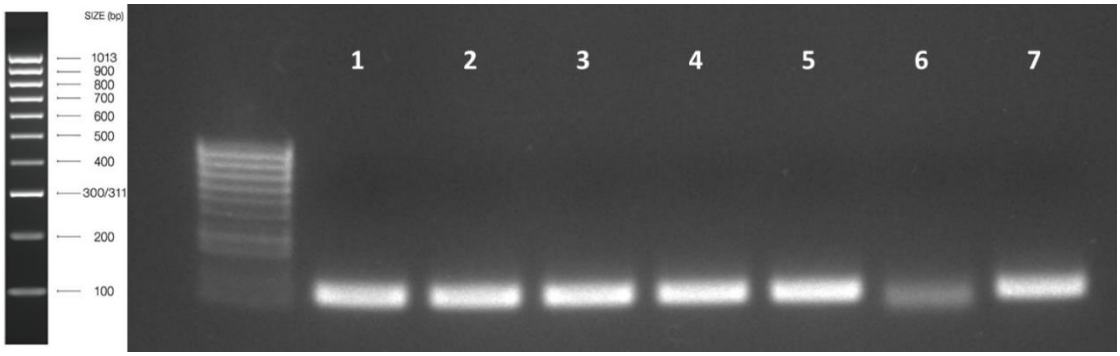
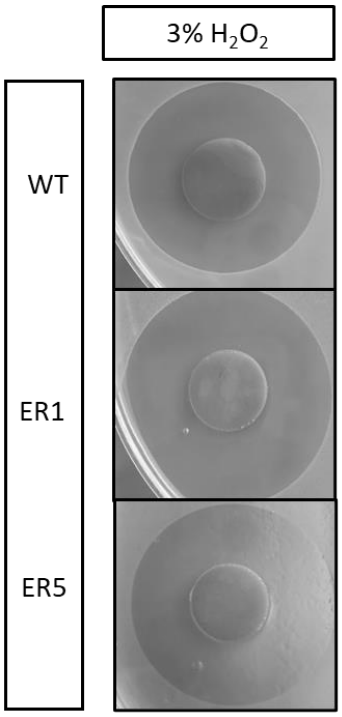


Figure S4. Colony PCR of emerging resistant colonies with *R. gnavus-specific* primers. Lanes 1-5 comprise resistant colonies; lane 6 is a negative control, and lane 7 is a positive control. DNA ladder: HyperLadder 100bp.



	Disc Size (cm)	Zone Size (cm)
WT	1.3	3
ER1	1.3	3.8
ER5	1.3	3.5

Figure S5. Treatment of *R. gnavus* with 3% hydrogen peroxide. WT and mutant strains ER1 and ER5 treated with a filter disc soaked in 3% H_2O_2. Measurements recorded in adjoining table.

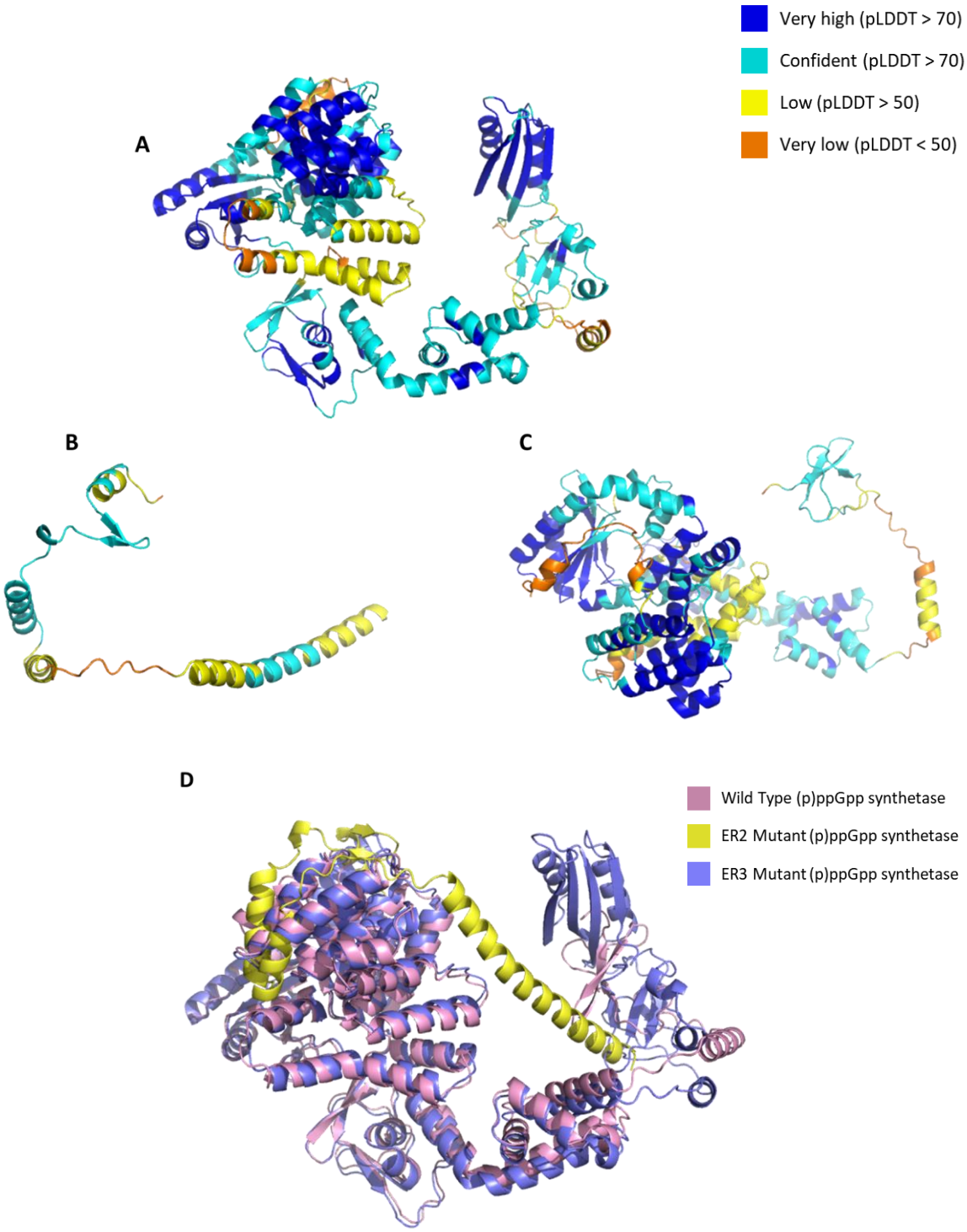


Figure S6. AlphaFold predictions of wild-type and mutant (p)ppGpp synthetase. Confidence scores of AlphaFold2.3 are indicated in the legend. **A)** (p)ppGpp synthetase of *R. gnavus* JCM6515^{T}. **B)** (p)ppGpp synthetase of *R. gnavus* JCM6515^{T} mutant ER2. **C)** (p)ppGpp synthetase of *R. gnavus* JCM6515^{T} mutant ER3. **D)** superimposed images of all structures to highlight conformational changes.

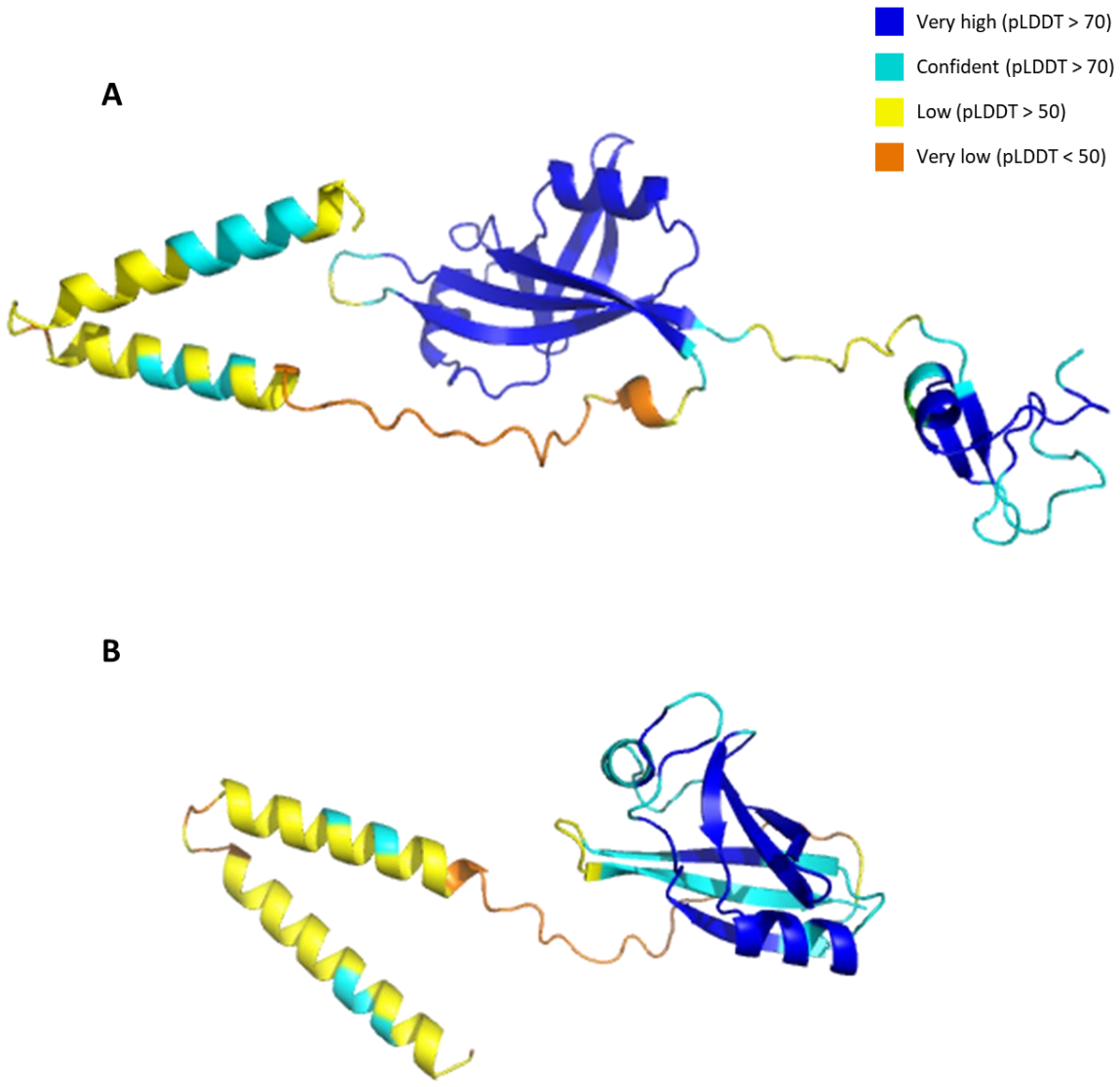


Figure S7. AlphaFold analysis of wild-type and mutant uncharacterised/HIRAN domain containing protein. Confidence scores of AlphaFold2.3 are indicated in the legend. **A)** Protein structure from the wild type strain **B)** Protein structure from ER6.

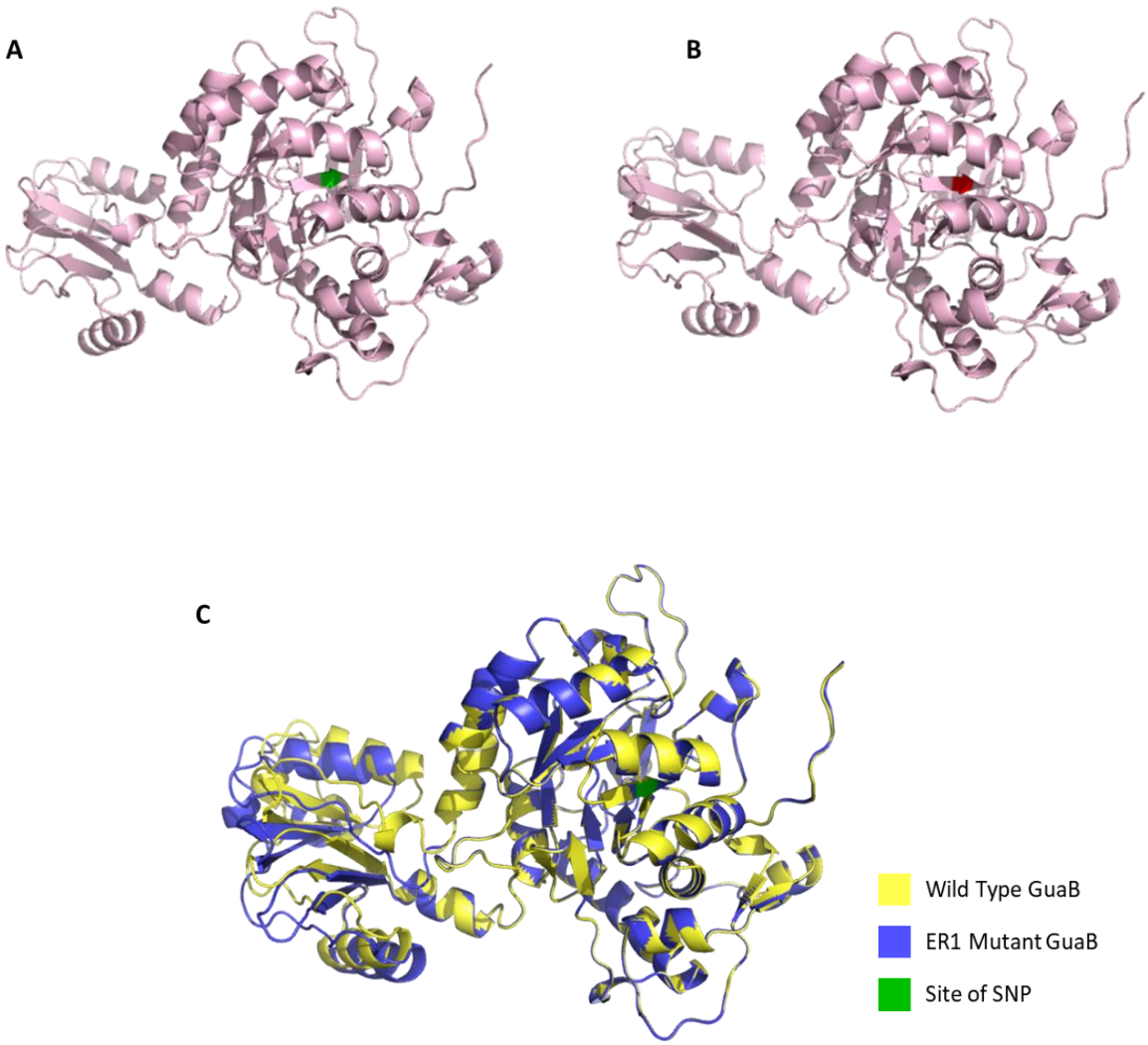


Figure S8. AlphaFold analysis of wild-type and mutant *GuaB* from ER1. A) Protein structure from the wild type strain with site of mutant SNP highlighted in green. **B)** Protein structure from the wild type strain with SNP highlighted in red. **C)** Protein structures of the WT and ER1 proteins superimposed and aligned. The site of the mutation is highlighted in green.

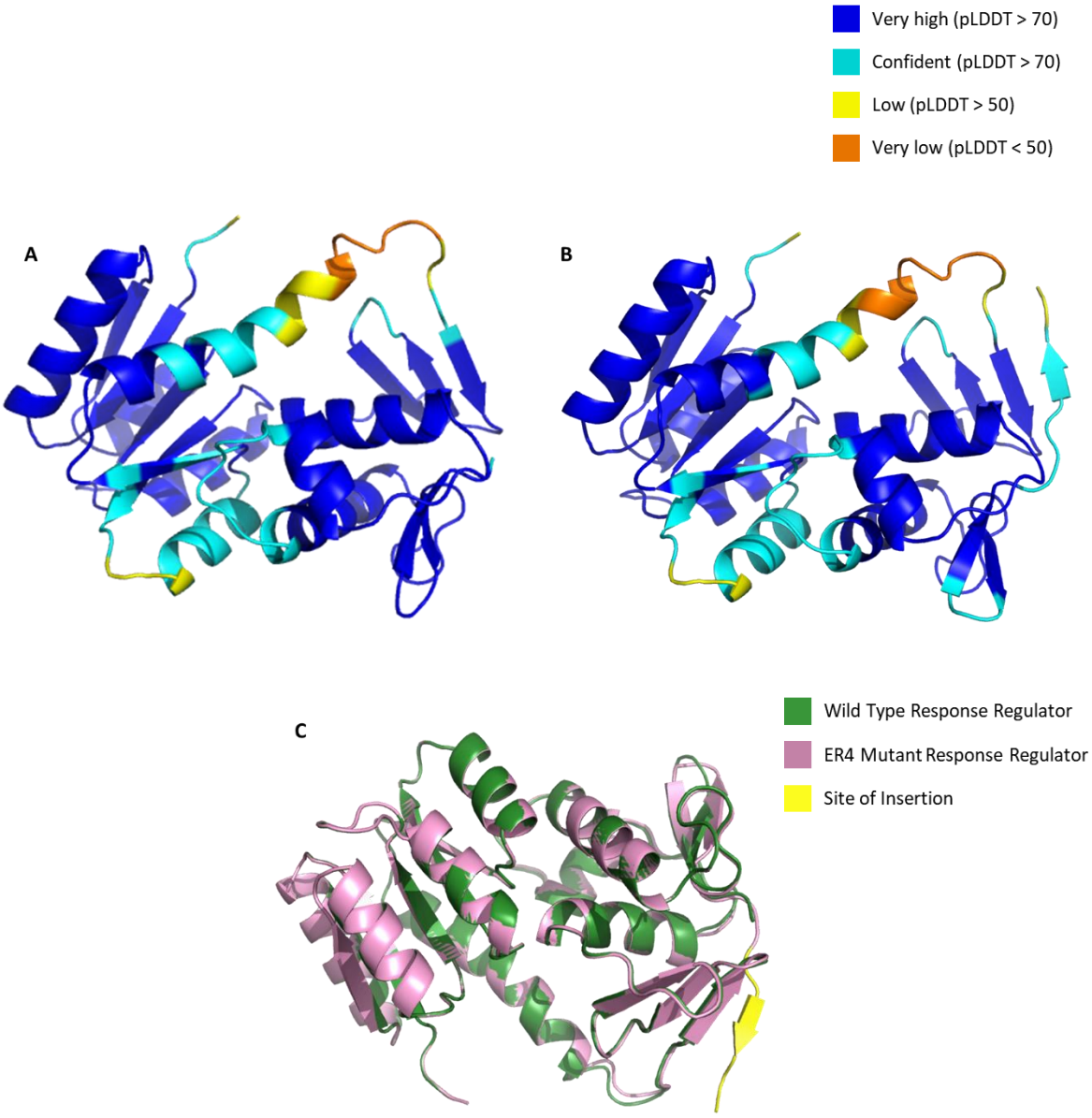


Figure S9. AlphaFold analysis of wild-type and mutant response regulator from ER4. A) Protein structure from the wild type strain. **B)** Protein structure from the mutant strain **C)** Protein structures of the WT and ER4 proteins superimposed and aligned. The site of the insertion is highlighted in yellow. Confidence scores of AlphaFold2.3 are indicated in the legend.

Figure S10. Positions of genes that have been subject to mutations in the genome of *R. gnavus* JCM6515^{T}, visualised with Geneious. **A)** Response regulator (ER4). **B)** TetR transcription al regulator (ER4). **C)** Extracellular solute binding protein (ER4). **D)** ABC transporter permease (ER4). **E)** GNAT family acetyltransferase (ER5). **F)** Trypsin-like serine protease (ER7). **G)** Pyruvate phosphate kinase.

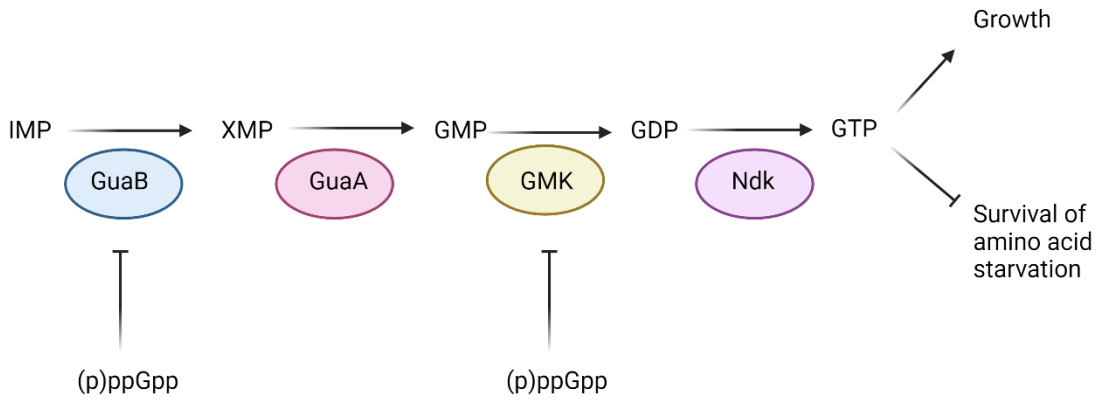


Figure S11. The relationship between *GuaB* and (p)ppGpp. Adapted from Hauryliuk et al., 2015. IMP is converted into XMP by *GuaB*. GuaA then converts XMP into GMP. GMP is converted to GTP, via phosphorylation. GMP kinase facilitates the conversion of GMP into GDP, which is then converted to GTP by nucleoside diphosphate kinase (Ndk). Created with Biorender.com.

Author Contributions

E.M., E.V.K., and A.S. conducted wet lab work and analysed the results; D.H., A.S., and E.M. performed bioinformatics. L.A.D. and N.N.A managed the project; R.P.R., A.N.S. and C.H. supervised the project. E.M., R.P.R., and C.H. wrote the manuscript.

References

[1] A. H. Togo et al., “Description of Mediterraneibacter massiliensis, gen. nov., sp. nov., a new genus isolated from the gut microbiota of an obese patient and reclassification of Ruminococcus faecis, Ruminococcus lactaris, Ruminococcus torques, Ruminococcus gnavus and Clostridium glycyrrhizinilyticum as Mediterraneibacter faecis comb. nov., Mediterraneibacter lactaris comb. nov., Mediterraneibacter torques comb. nov., Mediterraneibacter gnavus comb. nov. and Mediterraneibacter glycyrrhizinilyticus comb. nov,” Antonie Van Leeuwenhoek, vol. 111, no. 11, pp. 2107–2128, Nov. 2018, doi: 10.1007/S10482-018-1104-Y.

[2] E. H. Crost, E. Coletto, A. Bell, and N. Juge, “Ruminococcus gnavus: friend or foe for human health,” FEMS Microbiol Rev, vol. 47, no. 2, pp. 1–23, Mar. 2023, doi: 10.1093/FEMSRE/FUAD014.

[3] D. C. Baumgart and W. J. Sandborn, “Crohn’s disease,” The Lancet, vol. 380, no. 9853, pp. 1590–1605, 2012, doi: 10.1016/S0140-6736(12)60026-9.

[4] M. Zhao, L. Gönczi, P. L. Lakatos, and J. Burisch, “The Burden of Inflammatory Bowel Disease in Europe in 2020,” J Crohns Colitis, vol. 15, no. 9, pp. 1573–1587, Sep. 2021, doi: 10.1093/ECCO-JCC/JJAB029.

[5] S. S. Seyedian, F. Nokhostin, and M. D. Malamir, “A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease,” J Med Life, vol. 12, no. 2, p. 113, Apr. 2019, doi: 10.25122/JML-2018-0075.

[6] B. Gareb, A. T. Otten, H. W. Frijlink, G. Dijkstra, and J. G. W. Kosterink, “Review: Local Tumor Necrosis Factor-α Inhibition in Inflammatory Bowel Disease,” Pharmaceutics, vol. 12, no. 6, pp. 1–31, Jun. 2020, doi: 10.3390/PHARMACEUTICS12060539.

[7] “Inflammatory bowel disease - NHS.” Accessed: Aug. 08, 2022. [Online]. Available: https://www.nhs.uk/conditions/inflammatory-bowel-disease/

[8] A. N. Ananthakrishnan et al., “Environmental triggers in IBD: a review of progress and evidence,” Nature Reviews Gastroenterology & Hepatology 2017 15:1, vol. 15, no. 1, pp. 39–49, Oct. 2017, doi: 10.1038/nrgastro.2017.136.

[9] M. Joossens et al., “Dysbiosis of the faecal microbiota in patients with Crohn’s disease and their unaffected relatives”, doi: 10.1136/gut.2010.223263.

[10] A. B. Hall et al., “A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients”, doi: 10.1186/s13073-017-0490-5.

[11] E. H. Crost, L. E. Tailford, G. Le Gall, M. Fons, B. Henrissat, and N. Juge, “Utilisation of Mucin Glycans by the Human Gut Symbiont Ruminococcus gnavus Is Strain-Dependent,” PLoS One, vol. 8, no. 10, p. e76341, Oct. 2013, doi: 10.1371/JOURNAL.PONE.0076341.

[12] M. T. Henke, D. J. Kenny, C. D. Cassilly, H. Vlamakis, R. J. Xavier, and J. Clardy, “Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn’s disease, produces an inflammatory polysaccharide,” Proc Natl Acad Sci U S A, vol. 116, no. 26, pp. 12672–12677, 2019, doi: 10.1073/PNAS.1904099116/-/DCSUPPLEMENTAL.

[13] M. Talaat et al., “Increasing Antimicrobial Resistance in World Health Organization Eastern Mediterranean Region, 2017–2019 - Volume 28, Number 4—April 2022 - Emerging Infectious Diseases journal - CDC,” Emerg Infect Dis, vol. 28, no. 4, pp. 717–724, Apr. 2022, doi: 10.3201/EID2804.211975.

[14] S. Federici et al., “Targeted suppression of human IBD-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation,” Cell, vol. 185, no. 16, pp. 2879-2898.e24, Aug. 2022, doi: 10.1016/J.CELL.2022.07.003/ATTACHMENT/6D2D0D64-B056-46D4-83AA-34C5928ECB6A/MMC7.XLSX.

[15] E. Murray, L. A. Draper, R. P. Ross, and C. Hill, “The advantages and challenges of using endolysins in a clinical setting,” Viruses, vol. 13, no. 4, 2021, doi: 10.3390/v13040680.

[16] C. Buttimer et al., “Temperate bacteriophages infecting the mucin-degrading bacterium Ruminococcus gnavus from the human gut,” Gut Microbes, vol. 15, no. 1, 2023, doi: 10.1080/19490976.2023.2194794.

[17] L. Zimmermann et al., “A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core,” 2017, doi: 10.1016/j.jmb.2017.12.007.

[18] “AlphaFold Protein Structure Database.” Accessed: Jan. 08, 2024. [Online]. Available: https://AlphaFold.ebi.ac.uk/

[19] “PyMOL | pymol.org.” Accessed: Jan. 08, 2024. [Online]. Available: https://pymol.org/2/

[20] E. Guerin et al., “Isolation and characterisation of ΦcrAss002, a crAss-like phage from the human gut that infects Bacteroides xylanisolvens,” Microbiome, vol. 9, no. 1, pp. 1–21, Dec. 2021, doi: 10.1186/S40168-021-01036-7/FIGURES/7.

[21] D. E. Deatherage and J. E. Barrick, “Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq,” Methods Mol Biol, vol. 1151, p. 165, 2014, doi: 10.1007/978-1-4939-0554-6_12.

[22] “Amino Acid Physical Properties - IE”, Accessed: May 31, 2024. [Online]. Available: https://www.thermofisher.com/tr/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/amino-acid-physical-properties.html

[23] J. K. Hobbs and A. B. Boraston, “(p)ppGpp and the Stringent Response: An Emerging Threat to Antibiotic Therapy,” ACS Infect Dis, vol. 5, no. 9, pp. 1505–1517, Sep. 2019, doi: 10.1021/ACSINFECDIS.9B00204/ASSET/IMAGES/LARGE/ID-2019-00204J_0004.JPEG.

[24] M. A. Matilla, Á. Ortega, and T. Krell, “The role of solute binding proteins in signal transduction,” Comput Struct Biotechnol J, vol. 19, pp. 1786–1805, Jan. 2021, doi: 10.1016/J.CSBJ.2021.03.029.

[25] V. G. Lewis, M. P. Ween, and C. A. McDevitt, “The role of ATP-binding cassette transporters in bacterial pathogenicity,” Protoplasma 2012 249:4, vol. 249, no. 4, pp. 919–942, Jan. 2012, doi: 10.1007/S00709-011-0360-8.

[26] A. L. Davidson and J. Chen, “ATP-binding cassette transporters in bacteria,” Annu Rev Biochem, vol. 73, no. Volume 73, 2004, pp. 241–268, Jul. 2004, doi: 10.1146/ANNUREV.BIOCHEM.73.011303.073626/CITE/REFWORKS.

[27] I. C. Sutcliffe and R. R. B. Russell, “MINIREVIEW Lipoproteins of Gram-Positive Bacteria,” J Bacteriol, vol. 177, no. 5, pp. 1123–1128, 1995.

[28] M. van Kempen et al., “Fast and accurate protein structure search with Foldseek,” Nature Biotechnology 2023 42:2, vol. 42, no. 2, pp. 243–246, May 2023, doi: 10.1038/s41587-023-01773-0.

[29] C. F. Pope, D. M. O’sullivan, T. D. Mchugh, and S. H. Gillespie, “MINIREVIEW A Practical Guide to Measuring Mutation Rates in Antibiotic Resistance,” Antimicrob Agents Chemother, vol. 52, no. 4, pp. 1209–1214, 2008, doi: 10.1128/AAC.01152-07.

[30] M. U. Rahman et al., “Endolysin, a Promising Solution against Antimicrobial Resistance,” Antibiotics, vol. 10, no. 11, Nov. 2021, doi: 10.3390/ANTIBIOTICS10111277.

[31] V. A. Fischetti, “Bacteriophage lytic enzymes: Novel anti-infectives,” Trends Microbiol, vol. 13, no. 10, pp. 491–496, Oct. 2005, doi: 10.1016/j.tim.2005.08.007.

[32] J. M. Loeffler, D. Nelson, and V. A. Fischetti, “Rapid killing of Streptococcus pneumoniae with a bacteriophage cell wall hydrolase,” Science (1979), vol. 294, no. 5549, pp. 2170–2172, Dec. 2001, doi: 10.1126/SCIENCE.1066869/SUPPL_FILE/1066869S1_THUMB.GIF.

[33] L. Rodríguez-Rubio, B. Martínez, A. Rodríguez, D. M. Donovan, F. Götz, and P. García, “The Phage Lytic Proteins from the Staphylococcus aureus Bacteriophage vB_SauS-phiIPLA88 Display Multiple Active Catalytic Domains and Do Not Trigger Staphylococcal Resistance,” PLoS One, vol. 8, no. 5, p. e64671, May 2013, doi: 10.1371/JOURNAL.PONE.0064671.

[34] D. B. Gilmer, J. E. Schmitz, C. W. Euler, and V. A. Fischetti, “Novel bacteriophage lysin with broad lytic activity protects against mixed infection by Streptococcus pyogenes and methicillin-resistant Staphylococcus aureus,” Antimicrob Agents Chemother, vol. 57, no. 6, pp. 2743–2750, Jun. 2013, doi: 10.1128/AAC.02526-12.

[35] B. Wang, R. A. Grant, and M. T. Laub, “ppGpp coordinates nucleotide and amino-acid synthesis in E. coli during starvation,” Mol Cell, vol. 80, no. 1, p. 29, Oct. 2020, doi: 10.1016/J.MOLCEL.2020.08.005.

[36] A. Kriel et al., “GTP Dysregulation in Bacillus subtilis Cells Lacking (p)ppGpp Results in Phenotypic Amino Acid Auxotrophy and Failure To Adapt to Nutrient

Downshift and Regulate Biosynthesis Genes," J Bacteriol, vol. 196, no. 1, p. 189, Jan. 2014, doi: 10.1128/JB.00918-13.

[37] A. A. Pulschen, A. Z. N. Fernandes, A. F. Cunha, D. E. Sastre, B. E. Matsuguma, and F. J. Gueiros-Filho, "Many birds with one stone: targeting the (p)ppGpp signaling pathway of bacteria to improve antimicrobial therapy," Biophys Rev, vol. 13, no. 6, p. 1039, Dec. 2021, doi: 10.1007/S12551-021-00895-6.

[38] V. Hauryliuk, G. C. Atkinson, K. S. Murakami, T. Tenson, and K. Gerdes, "Recent functional insights into the role of (p)ppGpp in bacterial physiology," Nat Rev Microbiol, vol. 13, no. 5, p. 298, May 2015, doi: 10.1038/NRMICRO3448.

[39] A. N. Bittner, A. Kriel, and J. D. Wang, "Lowering GTP Level Increases Survival of Amino Acid Starvation but Slows Growth Rate for Bacillus subtilis Cells Lacking (p)ppGpp," J Bacteriol, vol. 196, no. 11, p. 2067, 2014, doi: 10.1128/JB.01471-14.

[40] M. M. Mwangi et al., "Whole-genome sequencing reveals a link between β-lactam resistance and synthetases of the alarmone (p)ppGpp in Staphylococcus aureus," Microbial Drug Resistance, vol. 19, no. 3, pp. 153–159, Jun. 2013, doi: 10.1089/MDR.2013.0053/ASSET/IMAGES/LARGE/FIGURE4.JPEG.

[41] R. M. Corrigan, L. E. Bellows, A. Wood, and A. Gründling, "PpGpp negatively impacts ribosome assembly affecting growth and antimicrobial tolerance in Grampositive bacteria," Proc Natl Acad Sci U S A, vol. 113, no. 12, pp. E1710–E1719, Mar. 2016, doi: 10.1073/PNAS.1522179113/SUPPL_FILE/PNAS.201522179SI.PDF.

[42] A. Salzer and C. Wolz, "Role of (p)ppGpp in antibiotic resistance, tolerance, persistence and survival in Firmicutes," microLife, vol. 4, pp. 1–12, Jan. 2023, doi: 10.1093/FEMSML/UQAD009.

[43] H. Takada et al., "Expression of Bacillus subtilis ABCF antibiotic resistance factor VmlR is regulated by RNA polymerase pausing, transcription attenuation, translation attenuation and," academic.oup.comH Takada, ZF Mandell, H Yakhnin, A Glazyrina, S Chiba, T Kurata, KJY Wu, BIC TrescoNucleic Acids Research, 2022•academic.oup.com, vol. 50, no. 11, pp. 6174–6189, 2022, doi: 10.1093/nar/gkac497.

[44] S. Aedo, A. T.-A. agents and chemotherapy, and undefined 2016, “Role of the stringent stress response in the antibiotic resistance phenotype of methicillin-resistant Staphylococcus aureus,” Am Soc MicrobiolS Aedo, A TomaszAntimicrobial agents and chemotherapy, 2016•Am Soc Microbiol, vol. 60, no. 4, pp. 2311–2317, Apr. 2016, doi: 10.1128/AAC.02697-15.

[45] J. Abranches, A. R. Martinez, J. K. Kajfasz, V. Chávez, D. A. Garsin, and J. A. Lemos, “The Molecular Alarmone (p)ppGpp Mediates Stress Responses, Vancomycin Tolerance, and Virulence in Enterococcus faecalis,” Am Soc MicrobiolJ Abranches, AR Martinez, JK Kajfasz, V Chávez, DA Garsin, JA LemosJournal of bacteriology, 2009•Am Soc Microbiol, vol. 191, no. 7, pp. 2248–2256, Apr. 2009, doi: 10.1128/JB.01726-08.

[46] E. S. Honsa et al., “Rela mutant enterococcus faecium with multiantibiotic tolerance arising in an immunocompromised host,” mBio, vol. 8, no. 1, Jan. 2017, doi: 10.1128/MBIO.02124-16/FORMAT/EPUB.

[47] S. Prem et al., “Towards an understanding of oleate hydratases and their application in industrial processes,” Microb Cell Fact, vol. 21, no. 1, pp. 1–15, Dec. 2022, doi: 10.1186/S12934-022-01777-6/FIGURES/5.

[48] L. Aravind and C. P. Ponting, “The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins.,” FEMS Microbiol Lett, vol. 176, no. 1, pp. 111–116, Jul. 1999, doi: 10.1111/J.1574-6968.1999.TB13650.X.

[49] S. Gebhard, “ABC transporters of antimicrobial peptides in Firmicutes bacteria – phylogeny, function and regulation,” Mol Microbiol, vol. 86, no. 6, pp. 1295–1317, Dec. 2012, doi: 10.1111/MMI.12078.

[50] A. Ahmad, S. Majaz, and F. Nouroz, “Two-component systems regulate ABC transporters in antimicrobial peptide production, immunity and resistance,” Microbiology (Reading), vol. 166, no. 1, pp. 4–20, 2020, doi: 10.1099/MIC.0.000823.

[51] R. J. P. Dawson and K. P. Locher, “Structure of a bacterial multidrug ABC transporter,” Nature 2006 443:7108, vol. 443, no. 7108, pp. 180–185, Aug. 2006, doi: 10.1038/nature05155.

[52] R. Tam and M. H. Saier, “Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria.,” Microbiol Rev, vol. 57, no. 2, p. 320, 1993, doi: 10.1128/MR.57.2.320-346.1993.

[53] M. A. Matilla, Á. Ortega, and T. Krell, “The role of solute binding proteins in signal transduction,” Comput Struct Biotechnol J, vol. 19, p. 1786, Jan. 2021, doi: 10.1016/J.CSBJ.2021.03.029.

[54] L. P. Tripathi and R. Sowdhamini, “Genome-wide survey of prokaryotic serine proteases: Analysis of distribution and domain architectures of five serine protease families in prokaryotes,” BMC Genomics, vol. 9, no. 1, pp. 1–28, Nov. 2008, doi: 10.1186/1471-2164-9-549/FIGURES/3.

[55] M. Wei, D. Ye, and D. Dunaway-Mariano, “Investigation of the role of the domain linkers in separate site catalysis by Clostridium symbiosum pyruvate phosphate dikinase.,” Biochemistry, vol. 40, no. 45, pp. 13466–13473, Nov. 2001, doi: 10.1021/BI0113061.

[56] P. García, B. Martínez, L. Rodríguez, and A. Rodríguez, “Synergy between the phage endolysin LysH5 and nisin to kill Staphylococcus aureus in pasteurized milk,” Int J Food Microbiol, vol. 141, no. 3, pp. 151–155, Jul. 2010, doi: 10.1016/J.IJFOODMICRO.2010.04.029.

[57] M. Thandar, R. Lood, B. Y. Winer, D. R. Deutsch, C. W. Euler, and V. A. Fischetti, “Novel engineered peptides of a phage lysin as effective antimicrobials against multidrug-resistant Acinetobacter baumannii,” Antimicrob Agents Chemother, vol. 60, no. 5, pp. 2971–2979, May 2016, doi: 10.1128/AAC.02972-15/SUPPL_FILE/ZAC004165080SO1.PDF.

[58] H. Yang, Y. Zhang, J. Yu, Y. Huang, X. E. Zhang, and H. Wei, “Novel chimeric lysin with high-Level antimicrobial activity against methicillin-Resistant staphylococcus aureus in vitro and in vivo,” Antimicrob Agents Chemother, vol. 58, no. 1, pp. 536–542, Jan. 2014, doi: 10.1128/AAC.01793-13/SUPPL_FILE/ZAC001142517SO1.PDF.

[59] “A Grammar of Graphics for Comparative Genomics • gggenomes.” Accessed: May 15, 2024. [Online]. Available: https://thackl.github.io/gggenomes/index.html

[60] J. Söding, A. Biegert, and A. N. Lupas, “The HHpred interactive server for protein homology detection and structure prediction,” Nucleic Acids Res, vol. 33, no. Web Server issue, p. W244, Jul. 2005, doi: 10.1093/NAR/GKI408.

[61] P. Jones et al., “InterProScan 5: genome-scale protein function classification,” vol. 30, no. 9, pp. 1236–1240, 2014, doi: 10.1093/bioinformatics/btu031.

[62] F. Sievers et al., “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega,” Mol Syst Biol, vol. 7, p. 539, 2011, doi: 10.1038/MSB.2011.75.

[63] M. Gouy, S. Guindon, and O. Gascuel, “SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building,” Mol Biol Evol, vol. 27, no. 2, pp. 221–224, Feb. 2010, doi: 10.1093/MOLBEV/MSP259.

[64] S. Chen, Y. Zhou, Y. Chen, and J. Gu, “fastp: an ultra-fast all-in-one FASTQ preprocessor,” Bioinformatics, vol. 34, no. 17, pp. i884–i890, Sep. 2018, doi: 10.1093/BIOINFORMATICS/BTY560.

[65] A. Bankevich et al., “SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing,” J Comput Biol, vol. 19, no. 5, pp. 455–477, May 2012, doi: 10.1089/CMB.2012.0021.

[66] O. Schwengers, L. Jelonek, M. A. Dieckmann, S. Beyvers, J. Blom, and A. Goesmann, “Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification,” Microb Genom, vol. 7, no. 11, p. 685, 2021, doi: 10.1099/MGEN.0.000685.

[67] D. E. Deatherage and J. E. Barrick, “Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq,” Methods Mol Biol, vol. 1151, pp. 165–188, 2014, doi: 10.1007/978-1-4939-0554-6_12.

Chapter 5. Functional Characterisation of a Novel Receptor Binding Domain from a Tail-Associated Lysin in a *Clostridioides difficile* Bacteriophage.

Ellen Murray, Shakhinur Islam Mondal, Christopher J. Turkington, Lorraine A. Draper, Neda Nezam Abadi, Paul R. Ross, Colin Hill

Abstract

Clostridioides difficile infection (CDI) is a significant healthcare problem. It is a nosocomial infection with high recurrence rates, and failure of treatment can have lethal consequences. To reduce the transmission of CDI in healthcare settings, it is important to be able to rapidly monitor *C. difficile* levels to inform treatment options and to control infection. In this study, we assess the potential of a receptor-binding domain, CdRBP, part of the multi-domain tail-associated lysin of *C. difficile* phage phiMMP01, to be used as a means of identifying *C. difficile*. Using a green fluorescent protein (GFP) linked to the binding domain, CdRBP was shown to bind *C. difficile* strains and is stable up to 50 ^{o}C and from pH 2-11. Given its demonstrated binding capacity against a panel of strains and its stability, CdRBP could prove to be a promising candidate as a diagnostic tool for *C. difficile*.

Introduction

Clostridioides difficile is a gram-positive, spore-forming bacterium that is commonly found in the intestinal tract of humans and animals. Although it is widely present in the hospital environment, in recent decades infections with this bacterium have become a significant healthcare problem. Spores of *C. difficile* in the environment may be spread via the faecal-oral route, which leads to a widespread source of infection *via* asymptomatic carriers, contaminated surroundings, and animal carriers including canine, feline, porcine, and avian species. It is estimated that up to 5% of adults and up to 70% of infants are carriers of *C. difficile* [1]. *C. difficile* infection (CDI), or *C. difficile*-associated disease (CDAD), is a severe illness that causes symptoms ranging from diarrhoea to life-threatening inflammation of the colon. CDI is a toxin-mediated disease as the bacterium is capable of producing three toxins that cause damage to the intestinal epithelium layer and incite an immune response that further exacerbates symptoms. These toxins are an enterotoxin (A), a cytotoxin (B) and *C. difficile* toxin (CDT). Risk factors for CDI include an age of 65 or older, prior hospitalisation, long-term hospital stays, and the administration of antibiotics [2]. The administration of antibiotics in a hospital setting, particularly in elderly patients, can cause a disruption to the patient's gut microbiome, consequentially clearing a niche for *C. difficile* to thrive [3]. Mortality as a result of CDI doubled from 1999 to 2004 and followed an upward trend until 2007 in the United Kingdom. This was followed by a decrease in disease incidence, but CDI is still a serious concern for healthcare providers [4].

Current medical guidelines recommend the administration of antibiotics such as vancomycin, metronidazole (first-line defence) and fidaxomicin for the eradication of the pathogen. Despite the sensitivity of *C. difficile* to these drugs, CDI has become increasingly associated with treatment failures and recurrence. The average treatment failure rate has been reported to be 22.3%, and the recurrence rate is 22.1%. It is also reported that treatment failures occur more commonly with the use of metronidazole [2]. Rates of recurrence can vary depending on where the disease was contracted, with a lower rate seen in community-associated diseases as compared to hospital-acquired infections [5]. To tackle CDI, there is an urgent need for alternative treatments that will act more specifically and minimise collateral damage to the microbiome. Recent developments in treatment options include the use of Faecal

Microbiota Transplant (FMT), bacteriophage (phage) therapy (including the use of endolysins), probiotics and bacteriocin treatments [6][7][8][9].

While there is a great need to develop strategies to specifically target and kill *C. difficile*, there is also a need for next-generation pathogen detection methods. Detection of *C. difficile* in a patient sample will help to identify those who may be at risk and allow for a specific treatment plan to be designed for these individuals, or for them to be placed in isolation wards with other carriers to limit transmission of disease to uninfected patients. Current methods for pathogen detection often include PCR, culturing methods and immunology-based methods. While these techniques are tried and trusted, they are also time-consuming and may take days to deliver a result. When dealing with a disease as severe as CDI, identification and quantifying the pathogen promptly is of utmost importance. Biosensors have the potential to make point-of-care testing available for CDI.

Phages are viruses that specifically infect and kill bacteria. Lysins are phage-encoded peptidoglycan hydrolases that play a role both at the start of a phage infection to allow the phage gain access to the bacterial host membrane, or towards the end of a phage replication cycle where they lyse the bacterial peptidoglycan to allow phage progeny to be released into the surrounding environment. Lysins involved at the beginning of the phage infection are generally termed tail-associated lysins, or TALs, while those at the end of the cycle are classed as endolysins (*endo* - acting from within) or cell wall hydrolases (CWHs). In recent decades, there has been a resurgence of interest in TALs and endolysins as novel therapeutics for bacterial infections. When applied exogenously, particularly in the case of gram-positive bacteria, the lytic activity observed during a phage infection cycle can be utilised as an antimicrobial agent. These phage-derived peptidoglycan hydrolases can efficiently lyse and kill bacterial cultures, making them promising candidates as antimicrobials [10]. Phage lysins generally have a multi-domain structure with each domain having designated functions. Enzymatically active domains (EAD) attack the peptidoglycan, and the cell wall binding domain (CBD) or Receptor Binding Protein (RBP) domains specifically attach the protein to the target cell wall. Both CBDs and RBPs have been harnessed for next-generation pathogen detection and diagnostic tools, with applications ranging from ELISA-based assays to fluorescence-based microscopy [3]. The binding

domains of these proteins are exciting tools for biotechnological applications due to their specificity, speed, stability, and cost-effectiveness.

CBDs of endolysins have already been utilised in the development of electrochemical impedance biosensors. In 2012, a biosensor for the detection of *Listeria monocytogenes* was described. This sensor consisted of a CBD immobilized on a gold screen-printed electrode (SPE), and the binding event was recorded by electrochemical impedance spectroscopy (EIS). This biosensor was successful in rapidly detecting *Listeria* in numbers as low as 1.1 $X10^4$ CFU/mL in both pure cultures and in 2% milk samples [11]. Both CBDs and RBPs of bacteriophages have been discussed as promising candidates in a range of biosensors, ranging from magnetic to electrochemical. RBPs and CBDs have the advantage of potentially lower detection limits, high specificity, and an ability to be engineered using molecular biology, which makes them viable candidates for the application of biosensors [12].

In a previous publication from our group, an enzyme targeting *C. difficile* was identified in phage phiMMP01 and its domains were characterised [13]. The 656-residue CWH was predicted to have a three-domain structure consisting of a C-terminal catalytic domain ($CWH_{351-656}$) and an N-terminal putative binding domain (CWH_{1-350}). The catalytic region contains two distinct peptidoglycan hydrolases - a glucosaminidase domain and an endopeptidase NIpC/P60 domain. The N-terminal domain had no significant homology to any known proteins, but some similarities were found between the query protein and bacteriophage tail proteins using HHpred, a tool that relies on structure prediction [14]. It was also noted that the CWH isolated did not resemble any other published characterized lysins at that time, but rather showed closer similarity to cell wall modifying enzymes of other bacterial pathogens. The CWH was truncated into its various domains, and these were recombinantly expressed in *Escherichia coli* to investigate their functions. While the EAD domain was able to lyse the target, the putative CWH binding domain had no lytic activity against any of the strains tested.

In this study, we further characterise this putative N-terminal binding domain of the phiMMP01 CWH. We compare this domain to other characterised lysins targeting *C. difficile* and expand upon its structural homology to their binding domains. To investigate the potential of the binding domain for diagnostic applications, we tested

its binding specificity against a panel of bacteria by creating a novel hybrid with green fluorescent protein. This allows for the biological identification of *Clostridioides* strains using fluorescent microscopy. The stability of the protein was also assessed in terms of heat and pH. We conclude that this novel protein domain could be a viable tool for the detection of *C. difficile*.

Results

Bioinformatic Analysis.

In the original publication describing the CWH of phiMMP01 (NCBI protein ID YP_009206142.1), no significant homology was found for the N-terminal domain of the endolysin, other than weak associations with phage tail proteins. Updated HHpred analysis on this domain revealed a number of hits with high probability. These include the *Shewanella oneidensis* prophage tail protein (Protein Data Bank (PDB) number 3CDD, 99.2% probability), a tail protein from *Neisseria meningitidis* MC58 (PBD number 3D37, 99.1% probability) and a tail protein from bacteriophage *Muvirus mu* (PDB number 1WRU, 99.9% probability). InterProScan search results indicated that the region 6-175 of the N-terminal domain was classified under “phage tail proteins”, Superfamily SSF6279.

To date, six additional lysins targeting *C. difficile* have been described in the literature. To further characterise the novel N-terminal domain of lysin from phiMMP01 we compared the binding domains of each of these proteins. To quantify the similarities between these proteins, a percentage identity matrix was generated **(Figure 1 A)**. Similarity was assessed at the amino acid level and revealed that the lysins are divided into two clusters. Lysins phiMMP01 and Ecd09610 have high sequence similarity (>90%) but are distinctly different to the remaining five endolysins. Multiple sequence alignments of the binding domains were performed using Clustal Omega **(Figure 1 B)** with the same conclusion. The PhiMMP01 and Ecd096010 putative binding domains are significantly longer than the other cell wall binding domains and have high sequence homology. Interestingly, there appears to be a region of the multiple sequence alignment (region 65-71) has conserved amino acid identity across all seven aligned proteins, including both clusters of binding domains, suggesting that this region may be important in binding to *C. difficile.*

The protein structure of the CWH from phiMMP01 was previously predicted for the active domain, but not for the N-terminal domain due to its lack of homology to known proteins at the time [13]. In this study, we aimed to expand upon the previous report and used AlphaFold2 to predict the structures of a panel of proteins **(Figure 2)**. The resulting predictions were of high confidence, with the predicted local distance difference test (pLDDT) on average greater than 70. To characterise the putative

binding domain of phiMMP01 further, we prepared structural predictions for the whole lysin **(Figure 2 A)**. As it was previously reported that the lysin had similarities to phage tail proteins, we also identified the endolysin (CWH2) (accession number: CEK40790.1) from the genome of phiMMP01 and used AlphaFold to predict its structure **(Figure 2 B).** AlphaFold predictions also were performed on the binding domains of a panel of the currently published *C. difficile* cell wall binding domains for comparison purposes, with the exception of the CD27L C-terminal binding domain which was imported from the X-ray crystallography structure found at PDB accession 4CU5 **(Figure 2 C)**.

Structural analysis also showed that the lysins fall into two distinct groups. The binding domains of PhiMMP01 and Ecd906010 both had a distinctive N-terminal domain made up of six anti-parallel β-sheets connected by a loop. This repeating structure is paired with sub-units consisting of three more anti-parallel β-sheets flanked by α-helices. The structure of these two binding domains shows greater structural similarity to Tail Associated Lysins (TALs) reported for other phages reported in the literature, rather than the clostridial endolysins cell wall binding domains [15] [16]. However, the PhiMMP01 genome also encodes a second endolysin (CWH2) that shows greater homology to the five endolysins reported in the literature (NCBI protein ID YP_009206154.1). CWH2 has a C-terminal putative cell wall binding domain, and an N-terminal active domain classified as N-acetylmuramoyl-L-alanine amidase by InterProScan. It is positioned next to the gene encoding the holin protein in the genome of phage phiMMP01 **(Figure 3)**. This indicates that the endolysin previously characterised as a CWH is actually the tail-associated lysin of PhiMMP01 with an N-terminal RBP, and that CWH2, with an amidase mode of action and a C-terminal CBD, is actually the endolysin.

The panel of endolysins taken from the literature follow a trend in structure. The C-terminal domain of CD27L has been previously described, and its crystal structure has been determined [17]. It possesses a novel fold that can be seen amongst other clostridial endolysins. As shown in our AlphaFold analysis **(Figure 2 C)**, the cell wall binding domains of *C. difficile* endolysins group together and share a structure that generally consists of four anti-parallel β-sheets, bordered by two α-helices at one terminal, and an additional α-helix at the opposite end of the protein.

Due to the bioinformatic analysis classifying the N-terminal domain of CWH as a tail-associated protein, the identification of the endolysin in the prophage genome, and the structural similarities between this N-terminal domain and other tail-associated lysins, we propose that the previously named CWH of phiMMP01 should be renamed CdTAL and that the N-terminal domain functions as a putative receptor binding protein (CdRPB). Hereafter, we refer to the CWH of phiMMP01 as CdTAL.

Cloning and Purification of CdRBP and a CdRBP-GFP Hybrid.

As outlined in the previous study [13], the recombinant N-terminal domain of CdTAL was overexpressed and purified. In this study, we also created a hybrid construct CdRBP-GFP in which the green fluorescent protein (GFP) is fused to the CdRBP binding domain. The gene fragment encoding CdRBP was initially cloned in the expression vector pET28b(+) [13], but to create CdRBP-GFP the same fragment was cloned in the vector pHTP9 which encodes the GFP protein **(Figure 4 A)**. Both proteins (CdRBP and CdRBP-GFP) were overexpressed in recombinant *E. coli* and purified using affinity chromatography. Elution fractions of the purification were analysed by 4-12% SDS-PAGE **(Figures 4 (B) and 3 (C))**. The molecular weight of CdRBP alone is predicted to be 42.3 kDa, and the fusion protein is 69.6 kDa.

Functional Analysis of CdRBP-GFP.

Binding assays were performed to functionally characterise the N-terminal domain of CdTAL. Previously, the domain was assigned as a putative CBD due to its lack of lytic activity [13].. The creation of a GFP-fusion construct allowed us to assess binding *via* fluorescent microscopy. The purified fusion protein (100 µg/mL) was incubated with different strains of *C. difficile*. These cells were then washed to release any unbound GFP, slides were prepared, and they were viewed under the microscope. These assays confirmed that the N-terminal domain of phiMMP01 CdTAL is involved in binding. Where binding occurs, GFP can be observed decorating the surface of *C. difficile* cells. Different strains of *C. difficile* displayed different efficiencies of binding **(Figure 5).**

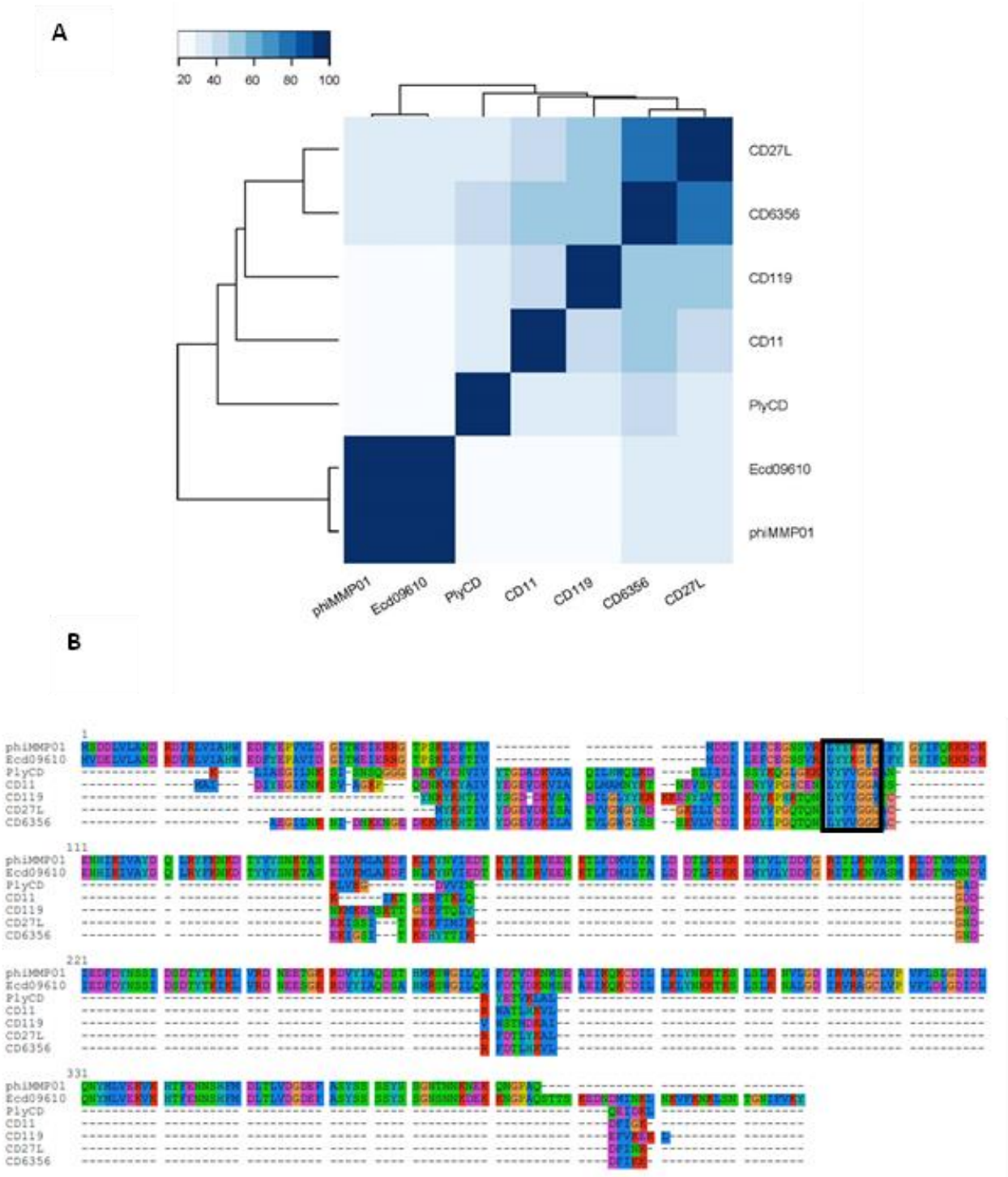


Figure 1. Bioinformatic Analysis of Cell Wall Binding Domains from *C. difficile* targeting endolysins. A) Percentage-identity analysis, percentage identity matrix calculated using Clustal Omega and visualised with Heatmap2. **B)** Clustal Omega alignment of PhiMMPO1 binding domains against other characterised *C. difficile* endolysins binding domains from the literature, visualised by SeaView4.

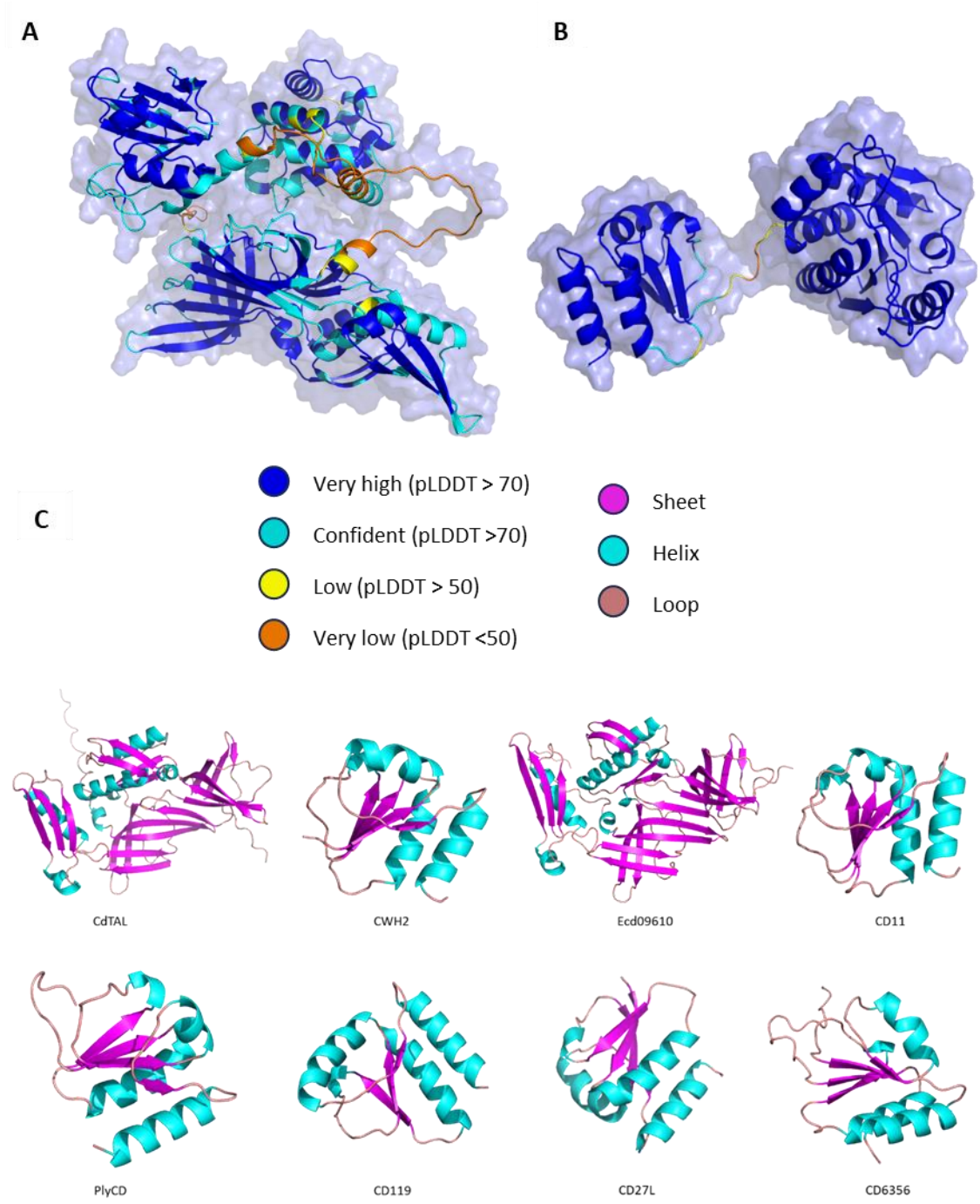


Figure 2. Structural predictions. A) Structural Analysis of phiMM01 whole tail-associated CWH **B)** Structural Analysis of endolysin from phiMM01. The ribbon structure is assigned colours based on pLDDT confidence scores as shown in the figure legend. **C)** Structural Analysis of Cell Wall Binding Domains and Receptor Binding Proteins (CdTAL, Ecd9610) of *C. difficile* lysins. Structures are assigned colours based on secondary structures as shown in the figure legend. CdTAL refers to the tail-associated lysin of phiMMP01, CWH2 refers to the endolysin of phiMMP01.

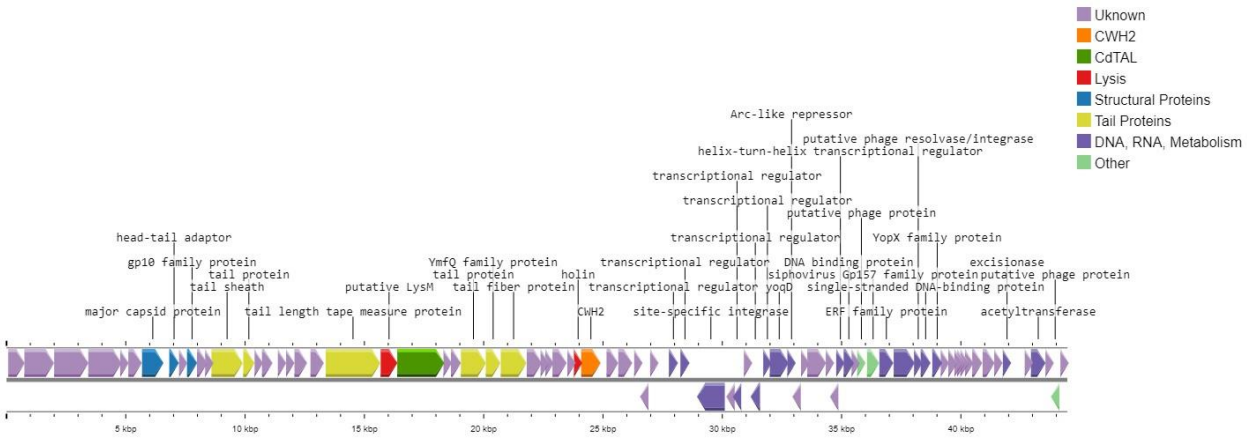


Figure 3. Genome map of phage phiMMP01. Complete phage genome obtained from NCBI, accession number NC_028883.1. Linear genome of 44,461 bp. Coding genes (CDS) are highlighted as per their NCBI annotation. The coloration of the CDS corresponds to their general functions as indicated in the legend. Genes whose function could not be determined are labelled “unknown”. Map created with Proksee [18].

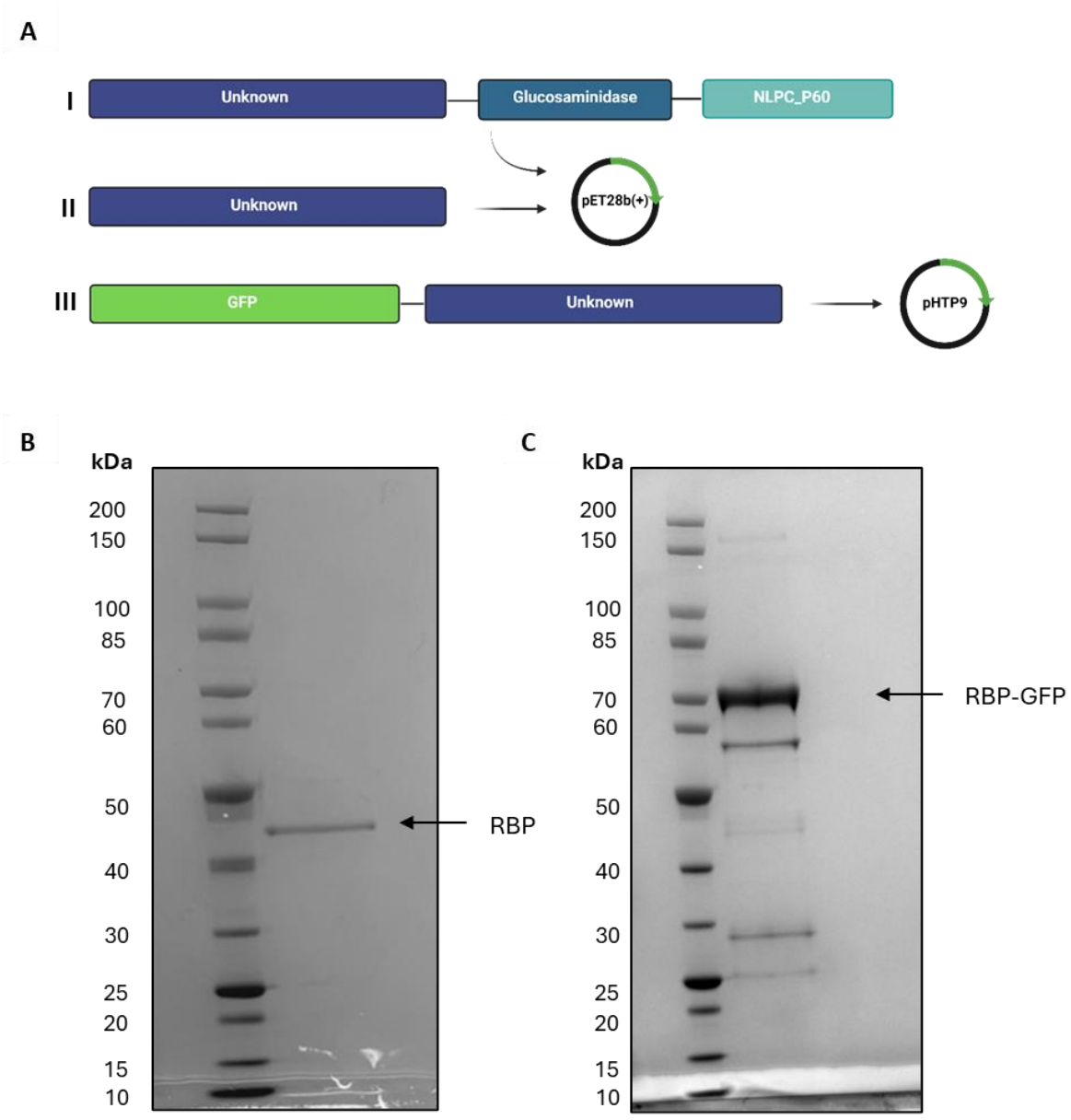


Figure 4. Expression and Purification of CBD constructs. **A)** Schematic of constructs designed from CdTAL of phiMMP01. Construct **(I)** was created in the previous study. Constructs **(II)** and **(III)** were created in this study. Created with BioRender.com **B)** Purification of phiMMP01 CdRBP, C) Purification of phiMMP01 CdRBP fused with GFP.

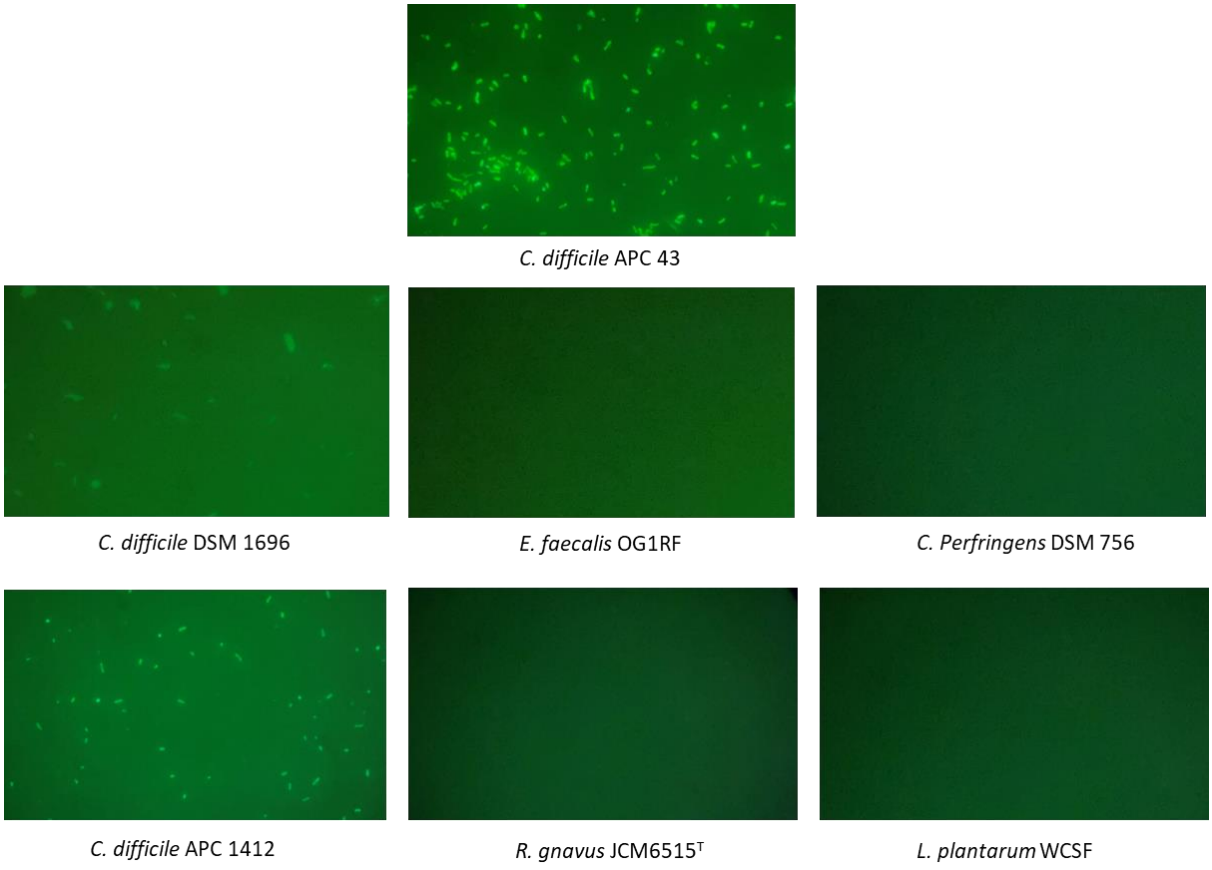


Figure 5. Binding Assays. Fluorescent microscopy was used to determine the binding specificity of CdRBP-GFP. A range of *Clostridioides* and non-*Clostridioides* strains were tested.

Potential for Diagnostics.

To assess the suitability of CdRBP-GFP as a candidate for biosensor technology, we tested its binding specificity and stability. For binding specificity, we initially tested a number of non-*Clostridioides* strains were tested. *Ruminococcus gnavus*, *Lactiplantibacillus plantarum*, *Clostridium perfringens* and *Enterococcus faecalis* are all gram-positive gut commensals that all gave a negative result which indicates that CdRBP-GFP does not bind non-specifically to this set of strains **(Figure 5)**. All samples were standardised to an OD_{600} of 0.8 for binding assays. Stability tests involved exposing CdRBP-GFP to a range of temperatures and pH. These tests were carried out against *C. difficile* APC43, as this was the most susceptible strain to CdTAL and showed the highest efficiency of binding of CdRBP-GFP. For temperature stability, the protein was pre-treated at various temperatures for 1 hour before the binding assay was performed. For the pH tests, *C. difficile* cells were pelleted and resuspended in 50 mM sodium phosphate buffer adjusted to a specific pH. Observations from binding assays showed that CdRBP-GFP was stable up to a temperature of 50 ^{o}C. At this temperature reduced binding was observed, and at 70 ^{o}C no binding was observed, and the protein appeared to be precipitated. pH tests showed that the function of the protein was retained across the pH range tested, indicating that CdRBP-GFP would be stable in a wide range of buffers **(Table 1).**

Analysis of Binding Efficiency in Complex Samples.

A desirable trait of biosensors is the ability to apply a complex sample to the sensor with minimal pretreatment. A biosensor that can give a reliable output with lower processing time is important for point-of-care diagnostics. As CdRBP-GFP showed promise in its initial assessment as a potential diagnostic tool, we wished to further test its suitability by carrying out binding assays in a complex sample. CdRBP-GFP was applied to a crude healthy human faecal sample. This sample consisted of 1g of faeces resuspended in PBS which was autoclaved for safety purposes. This faecal sample was treated with CdRBP-GFP at a concentration of 100 µg/mL in the same manner as the previous experiments. Once the incubation time had passed, the sample was washed three times in PBS to dislodge any unbound CdRBP-GFP. Fluorescent microscopy analysis revealed that CdRBP-GFP bound to a high number

of cells. It is possible that some of these cells were not *C. diffcile*. The protein also appeared to be significantly sequestered by debris in the faecal sample **(Figure 6(A))**.

To assess if the observations in **Figure 6 (A)** were due to the solid matter in the crude faecal sample, or if it was truly non-specific binding, we prepared a mixed bacterial community. This community was composed of a bank of 400 highly characterised gut commensal strains from our lab representing more than 100 species but that did not include any strains of *C. difficile*. The community was grown in liquid media, the cells were harvested, washed and resuspended to a final OD_{600} of 0.8. The bacterial community sample was also treated with CdRBP-GFP under the same conditions as the faecal sample. Analysis under the fluorescent microscope showed that in the absence of any solid debris or mucus that could be seen in the faecal sample, fluorescent cells could still be detected **(Figure 6B)**. A relatively high number of cells in the field of view appear to be decorated with CdRBP-GFP, showing that CdRBP has off-target binding. The bacterial community inoculum was prepared with an equal number of each strain, but as this initial inoculum was allowed to grow for 2 days it likely that the dominant members of the community have thrived and outcompeted other strains. As a result, the high number of cells that exhibit binding are likely due to an off-target effect being recorded on the dominant strain.

Table 1. Stability Tests. The binding abilities of CdRBP-GFP were tested at a range of temperatures and pHs. “+” represents no change in binding, “+/-“ for reduced binding and “-“ for no binding observed.

Temperature (°C)	Retained Function
4	+
25	+
37	+
50	+/-
70	-

pH	Retained Function
2	+
4	+
7	+
9	+
11	+

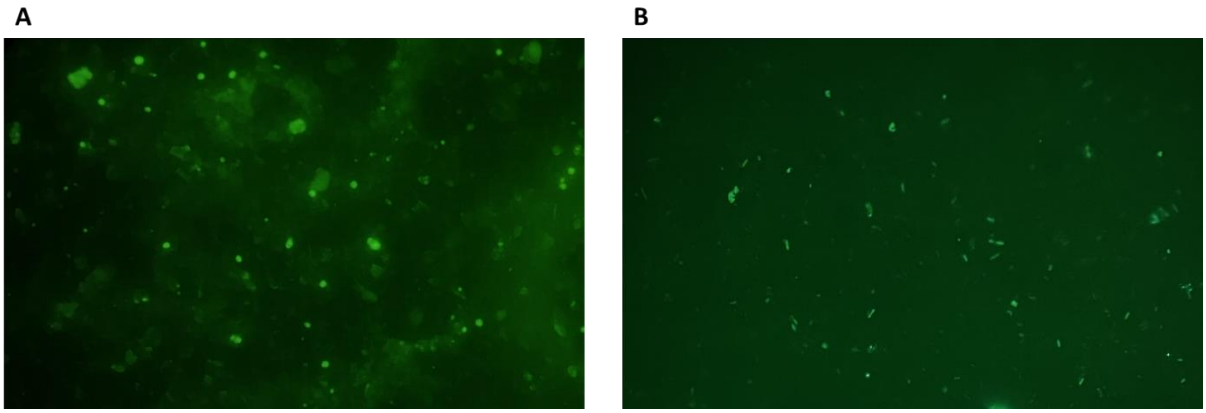


Figure 6. Binding assays in complex samples. Fluorescent microscopy was used to determine the binding specificity of CdRBP-GFP in a crude faecal sample and a mixed bacterial community. **A)** Crude faecal sample treated with 100 µg/mL CdRBP-GFP. **B)** Community of bacteria treated with 100 µg/mL CdRBP-GFP.

Discussion

In this study, we describe the characterisation of an RBP domain of a phage tail lysin (referred to as CdTAL) targeting *C. difficile*. The CdRBP domain was also fused to GFP, and this allowed for an expanded characterisation of the CdRBP. CdRBP is similar in predicted structure to the binding domain of another published tail-associated lysin (Ecd09610). In a small panel of gut-associated microbes the protein bound only to *C. difficile* strains. However, when the binding domain was assessed against a complex faecal sample, and a characterised bacterial community some off-target binding was observed. CdRBP has the stability and binding efficiency to indicate it could be a candidate for a diagnostic tool but there are challenges of non-specific binding that need to be overcome.

Comparative analysis and structural analysis of the N-terminal domain of CdTAL from phiMMP01 showed that rather than being grouped with endolysins targeting *C. difficile*, it clustered with lysin Ecd906010. These two binding domains have distinct sequence and structural differences from other reported lysins, which lead us to the reclassification of our endolysin as a tail-associated lysin and the CBD binding domain as an RBP. In 2022, Sekiya et al confirmed that Ecd09610 has lytic activity against *C. difficile* [19]. The gene encoding Ecd09610 was identified in the genome of *C. difficile* 630 (GenBank accession: AM180355.1). Ecd09610 also possesses an uncharacterised N-terminal domain and a catalytic C-terminal domain. It was reported that there was a 98% similarity between the N-terminal domains of the phiMMP01 CWH (now CdTAL) and Ecd09610. They also suggested that phiMMP01 is present as a lysogenic phage in *C. difficile* 630. In their study, truncation of the Ecd09610 protein was also carried out, and it was reported that the N-terminal domain bound ”slightly” to the target strain *C. difficile* 630, as assessed by SDS-page analysis. In our study, the use of a GFP-fusion protein has allowed us to confirm this binding, showing a clear distinction in the binding efficiency across a range of strains, and providing further evidence for the function of this domain.

A comprehensive study of the *C. difficile* endolysin cell wall binding domains, called CD16/50L, was published in 2022, however, the sequence of these lysins (CD16 and 50L) were not publicly available at the time of our analysis so they were not included in the bioinformatics workflow [20]. However, it was reported that the CBD

clustered with CD119 and CD11 which were included in our analysis. This study also investigated the interaction of the CBD CD16/50L with the *C. difficile* peptidoglycan-polysaccharides and confirmed its binding abilities [20]. As reported in the previous phiMMP01 study, the catalytic activity of CdTAL (previously CWH) does not depend on the specificity conferred by the RBP. This conclusion was also endorsed in the Ecd906010 study where it was suggested that the binding domain could even potentially hinder the catalytic activity. Truncated constructs of phiMMP01 CdTAL consisting of the enzymatic domains (either alone, or as a two-domain EAD) showed similar, and even increased lytic activity. This report is consistent with the literature. The study of endolysin CD16/50L also concluded that the lysin did not depend on its CBD for activity, and even reduced lytic activity [21]. Likewise, *C. difficile* phage endolysins CD27L, PlyCD, and CD11 do not require a CBD for their catalytic activity [22] [23] [24].

We successfully cloned and expressed the CdRBP domain and a CdRBP-GFP fusion protein **(Figure 3).** The purification of these proteins has allowed us to create a biological sensor that has demonstrated both specificity and stability. The highest binding was observed against *C. difficile* strain APC43 (alternatively named ATCC 43255 or ribotype 087, a high-level toxin-producing strain isolated from an abdominal wound), with varying efficiency against other strains, and a lack of non-specific binding observed from the non-clostridial strains tested in the initial host range analysis. We also demonstrated that CdRBP-GFP is stable up to 50 ^{o}C and has a pH range from 2-11.

To test the limitations of CdRBP-GFP in complex and microbe-rich samples, binding assays were carried out on a crude human faecal sample and a gut-derived bacterial community consisting of 400 strains. A significant amount of non-specific binding was observed **(Figure 6)**. Initially we hypothesised that the off-target binding in the faecal sample could be due to CdRBP-GFP being bound to mucus, solid matter and food debris. However, the preparation of a defined diverse bacterial community allowed us to test a sample that could be comparable to the bacterial diversity of a faecal sample but without the physical challenges. In this community, which specifically excluded *C. difficile*, CdRBP-GFP bound to a wide number of cells. These results highlight the need to assess non-specific binding of phage RBPs and CBDs when aiming to develop diagnostic tools. To avoid faecal debris blocking and

sequestering the binding domains the development of pre-treatments to crude samples (such as washing and centrifugation at varying speeds) could minimise this issue. However, the observation of non-specific binding is certainly a limitation of this specific CdRBP as a diagnostic tool. It may be a feature unique to this binding domain, or it may be that previous studies involving endolysin domains have not utilised sufficiently diverse bacterial target communities to allow the identification of non-specific binding.

In recent years, lysin-based detection methods have been gaining interest. Phage-derived binding proteins possess several traits that make them good candidates for these applications. They are much smaller than antibodies, are easy to produce and their recombinant production can be easily scaled up. They usually are reported to have high specificity, and they have low detection limits in comparison to antibodies [25]. Diagnostic methods that have involved the use of CBDs and RBPs include fluorescent detection, the use of magnetic beads and electrochemical biosensors. A 2018 study described the use of a tail fibre protein for the recognition of *Pseudomonas aeruginosa* [26]. The tail fibre protein P069 was utilised for two forms of detection – conjugation to magnetic beads for the separation of the target bacteria, and the use of fluorescent labels to capture and detect *P. aeruginosa*. The limit of the detection was recorded to be 10^2 CFU/mL for both the magnetic beads and bioluminescent methods. A multiplex system for the detection of *E. coli* and *P. aeruginosa* in blood samples was described in 2023 [27]. This study used recombinantly-produced RBPs in a microfluidic platform and it resulted in specific detection of the bacteria in whole blood within 70 minutes, with a detection limit of 10^3 CFU. It is clear from current studies encompassing both CBDs and RBPs of phages that these molecules hold promise as detection tools for a range of pathogens, but their ability to function specifically in microbe-rich environments must be carefully assessed. We propose that future work could investigate the viability of CdRBP to be adapted to a biological sensor platform for the detection of *C. difficile*. It is possible that with the creation of a multiplexed sensor, perhaps including a cocktail of the binding domains discussed in this study, non-specific binding could be minimised.

In conclusion, we have functionally characterised a novel RBP from the *C. difficile* phage PhiMMP01. The RBP alone and a GFP-fusion version of this domain were recombinantly expressed and characterised. The GFP-fusion protein allowed us

to investigate the function of this domain through binding assays and fluorescent confocal microscopy. CdRBP has high stability, and has been shown to function effectively. Given the traits we have discussed here, with optimisation to address issues of non-specific binding, CdRBP could be a part of a feasible diagnostic tool for *C. difficile* infections.

Materials and Methods

Bioinformatic Analysis.

Homology searches were performed for phiMMP01 BD in HHpred [14] and InterPro Scan [28]. Comparative analysis was performed on phiMMP01 CWH against several endolysins targeting *C. difficile* that have been characterized in the literature. These proteins can be found on GenBank at the following accession numbers: phiMMP01 CWH - YP_009206142.1, CD11 - WP_009895119.1, CD119 - YP_529586.1, CD6356 - YP_004306129.1, Ecd090610 - CAJ67813.1, PlyCD WP_003435466.1, CD27L - PBI50115.1. Multiple sequence alignments were generated using Clustal Omega and visualized using SeaView4 [29]. The percentage identity matrix was also generated using Clustal Omega and was visualised using Heatmap2. For structural analysis of endolysins and cell wall binding domains, AlphaFold v2.3.1 was used to predict the structures with the settings "--enable_gpu_relax=true --model_preset=monomer --use_gpu=true" [30].

Bacterial Strains and Growth Conditions.

All strains used in this study (including the community sample) were cultured in modified Brain Heart Infusion (Oxoid) medium with the addition of 0.5% yeast extract, 5 mg/L hemin (Sigma–Aldrich), 1 mg/ml cellobiose (Sigma–Aldrich), 1 mg/ml maltose (Sigma–Aldrich), and 0.5 mg/ml L-cysteine (Sigma–Aldrich) – LYHBHI as described by Soko et al, 2008 [31], except recombinant *E. coli* strains which were grown in Laura-Bertani (LB) broth (Fischer Scientific), supplemented with 50 µg/mL kanamycin (Sigma Aldrich) for maintenance of plasmids. Strains used for binding assays have been made available in the APC Culture Collection (APC Microbiome, Ireland). Aerobic strains were cultivated by incubating at 37 ^{o}C, and shaking at 150 rpm, while anaerobic strains were grown at 37 ^{o}C under strict anaerobic conditions (Type A vinyl anaerobic chamber, Coy Labs).

Cloning.

As outlined in the previous study by Mondal et al [13] the putative cell wall binding domain of endolysin from phage PhiMMP01 (NCBI Reference Sequence YP_009206142.1) was synthesised, with codon optimisation (GenScript), into the

vector pET28b(+) and introduced into *E. coli* BL21(DE3) for expression. For the purpose of this study, this previously created construct is named RBP.

The putative RBP with GFP fusion construct (RBP-GFP) was created using NZYEasy Cloning & Expression Kit IX (NZYtech) which created a construct in plasmid pHTP9 that contains GFP as a fusion tag. This construct was introduced into *E. coli* Trans-alpha competent cells. Transformants were screened for positive clones which were propagated for plasmid extractions. Extracted plasmids positive for the insert were introduced into One Shot BL21(DE3) chemically competent cells (Invitrogen) for protein expression. Sanger sequencing confirmed the integrity of the construct.

Recombinant Expression and Purification.

E. coli BL21(DE3) cells containing each of the constructs were overexpressed in the same manner. 3 L of cells were prepared in LB supplemented with 50 µg/mL kanamycin and incubated at 37 ^{o}C at 150 rpm shaking. Cells were grown to an OD 0.6 and expression was induced by the addition of 1 mM final concentration isopropyl-β-D-thiogalactopyranoside (IPTG) (Sigma Aldrich). Cultures were left to shake overnight at room temperature and were harvested the following day. Cells were pelleted at 4,500 rpm for 30 minutes. Cell pellets were lysed using sonication on ice, pelleted again by centrifugation and the supernatant was extracted for protein purification.

Purification of proteins was facilitated by the Akta Start Purification System (Cytiva), using the Talon Crude Purification columns (Cytiva). Purification fractions were assessed for quality and purity by SDS-PAGE (4-12% gradient gel) and imaged with iBright500. Fractions containing RBP were pooled, resuspended in a final concentration of 20% glycerol, and stored at -20 ^{o}C for short-term storage, and -80 ^{o}C for extended storage.

Binding Assays and Fluorescent Microscopy.

The binding of RBP to *C. difficile* strains was demonstrated through the following methods. Overnight cultures of bacterial candidates were prepared. Cells were harvested by centrifugation at 3,000 xg for 5 minutes. Cells were washed 3 times in PBS (phosphate-buffered saline) and resuspended to an OD_{600} of 0.8. 200 µL cells were combined with a final concentration of 100 µg/mL of purified protein, and shaken at 200 rpm at room temperature for 45 minutes. This cell suspension was also

harvested by centrifugation at 3,000 xg for 5 minutes and washed three times in PBS to remove unbound protein. Cells were resuspended to a volume of 200 uL, and 10 µL was added to a microscope slide, dried, and secured with a coverslip. For fluorescent microscopy, cells were viewed with a Zeiss Confocal Microscope at 63X magnification and excited with a green, fluorescent laser at 458 nm.

To prepare a complex sample for binding assays, 1 g of faecal sample was resuspended in 10 mL of phosphate buffered saline (PBS). The resuspension was thoroughly vortexed before autoclaving at cycle 121 ^{o}C for 15 minutes (VWR VAPOUR-Line Eco).

Temperature and pH Stability Tests.

Overnight cultures of bacterial candidates were prepared. Cells were harvested by centrifugation at 3,000 xg for 5 minutes. Cells were washed 3 times in PBS (phosphate-buffered saline) and resuspended to an OD_{600} of 0.8. For heat stability, the pure protein was treated at various temperatures (4 ^{o}C, 25 ^{o}C, 37 ^{o}C, 50 ^{o}C and 70 ^{o}C) for one hour. 200 µL cells were combined with a final concentration of 100 µg/mL of the temperature treated protein, and were shaken at 200 rpm at room temperature for 45 minutes.

To assess pH stability, Overnight cultures of bacterial candidates were prepared. Cells were harvested by centrifugation at 3,000 xg for 5 minutes. Cells were washed 3 times in PBS (phosphate-buffered saline) and resuspended to an OD_{600} of 0.8 in a buffer of set pH (2, 4, 7, 9 and 11). 200 µL cells in the pH adjusted buffer were combined with a final concentration of 100 µg/mL of the temperature treated protein, and were shaken at 200 rpm at room temperature for 45 minutes.

Efficiency of binding was assessed by fluorescent microscopy as previously described.

Author Contributions

EM and CH conceptualised the study. SIM identified and isolated the original protein. EM created the fusion protein, and carried out bioinformatic analysis, structural analysis, binding assays, and stability tests. CJT aided in the preparation of the bacterial community. LAD and NNA managed the project. EM, CH and PR wrote the manuscript.

References

[1] J. Czepiel et al., “Clostridium difficile infection: review,” European Journal of Clinical Microbiology and Infectious Diseases, vol. 38, no. 7, pp. 1211–1221, Apr. 2019, doi: 10.1007/S10096-019-03539-6/TABLES/1.

[2] K. Z. Vardakas, K. A. Polyzos, K. Patouni, P. I. Rafailidis, G. Samonis, and M. E. Falagas, “Treatment failure and recurrence of Clostridium difficile infection following treatment with vancomycin or metronidazole: a systematic review of the evidence,” Int J Antimicrob Agents, vol. 40, no. 1, pp. 1–8, Jul. 2012, doi: 10.1016/J.IJANTIMICAG.2012.01.004.

[3] J. Klumpp, M. Dunne, and M. J. Loessner, “A perfect fit: Bacteriophage receptor-binding proteins for diagnostic and therapeutic applications,” Curr Opin Microbiol, vol. 71, p. 102240, Feb. 2023, doi: 10.1016/J.MIB.2022.102240.

[4] P. N. Wiegand, D. Nathwani, M. H. Wilcox, J. Stephens, A. Shelbaya, and S. Haider, “Clinical and economic burden of Clostridium difficile infection in Europe: a systematic review of healthcare-facility-acquired infection,” Journal of Hospital Infection, vol. 81, no. 1, pp. 1–14, May 2012, doi: 10.1016/J.JHIN.2012.02.004.

[5] E. Finn, F. L. Andersson, and M. Madin-Warburton, “Burden of Clostridioides difficile infection (CDI) - a systematic review of the epidemiology of primary and recurrent CDI,” BMC Infect Dis, vol. 21, no. 1, pp. 1–11, Dec. 2021, doi: 10.1186/S12879-021-06147-Y/TABLES/3.

[6] M. C. Rea et al., “Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile,” Proc Natl Acad Sci U S A, vol. 107, no. 20, pp. 9352–9357, May 2010, doi: 10.1073/PNAS.0913554107.

[7] M. C. Rea, D. Alemayehu, R. Paul Ross, and C. Hill, “Gut solutions to a gut problem: Bacteriocins, probiotics and bacteriophage for control of Clostridium difficile infection,” J Med Microbiol, vol. 62, no. 9, pp. 1369–1378, Sep. 2013, doi: 10.1099/JMM.0.058933-0/CITE/REFWORKS.

[8] D. Drekonja et al., “Fecal Microbiota Transplantation for Clostridium difficile Infection,” https://doi.org/10.7326/M14-2693, vol. 162, no. 9, pp. 630–638, May 2015, doi: 10.7326/M14-2693.

[9] K. Selle et al., “In vivo targeting of clostridioides difficile using phage- delivered CRISPR-Cas3 antimicrobials,” mBio, vol. 11, no. 2, Mar. 2020, doi: 10.1128/MBIO.00019-20/SUPPL_FILE/MBIO.00019-20-ST002.DOCX.

[10] E. Murray, L. A. Draper, R. P. Ross, and C. Hill, “The advantages and challenges of using endolysins in a clinical setting,” Viruses, vol. 13, no. 4, 2021, doi: 10.3390/v13040680.

[11] M. Tolba, M. U. Ahmed, C. Tlili, F. Eichenseher, M. J. Loessner, and M. Zourob, “A bacteriophage endolysin-based electrochemical impedance biosensor for the rapid detection of Listeria cells,” Analyst, vol. 137, no. 24, pp. 5749–5756, Nov. 2012, doi: 10.1039/C2AN35988J.

[12] S. P. Costa, C. L. Nogueira, A. P. Cunha, A. Lisac, and C. M. Carvalho, “Potential of bacteriophage proteins as recognition molecules for pathogen detection,” Crit Rev Biotechnol, vol. 43, no. 5, pp. 787–804, 2023, doi: 10.1080/07388551.2022.2071671.

[13] S. I. Mondal, A. Akter, L. A. Draper, R. P. Ross, and C. Hill, “Characterization of an endolysin targeting clostridioides difficile that affects spore outgrowth,” Int J Mol Sci, vol. 22, no. 11, p. 5690, Jun. 2021, doi: 10.3390/IJMS22115690/S1.

[14] J. Söding, A. Biegert, and A. N. Lupas, “The HHpred interactive server for protein homology detection and structure prediction,” Nucleic Acids Res, vol. 33, no. Web Server issue, p. W244, Jul. 2005, doi: 10.1093/NAR/GKI408.

[15] S. Kanamaru et al., “Structure of the cell-puncturing device of bacteriophage T4,” Nature 2002 415:6871, vol. 415, no. 6871, pp. 553–557, Jan. 2002, doi: 10.1038/415553a.

[16] R. Linares, C. A. Arnaud, S. Degroux, G. Schoehn, and C. Breyton, “Structure, function and assembly of the long, flexible tail of siphophages,” Curr Opin Virol, vol. 45, pp. 34–42, Dec. 2020, doi: 10.1016/J.COVIRO.2020.06.010.

[17] M. Dunne et al., “The CD27L and CTP1L endolysins targeting Clostridia contain a built-in trigger and release factor,” PLoS Pathog, vol. 10, no. 7, 2014, doi: 10.1371/JOURNAL.PPAT.1004228.

[18] J. R. Grant et al., “Proksee: in-depth characterization and visualization of bacterial genomes,” Nucleic Acids Res, vol. 51, no. W1, pp. W484–W492, Jul. 2023, doi: 10.1093/NAR/GKAD326.

[19] H. Sekiya, H. Yamaji, A. Yoshida, R. Matsunami, S. Kamitori, and E. Tamai, “Biochemical Characterizations of the Putative Endolysin Ecd09610 Catalytic Domain from Clostridioides difficile,” Antibiotics, vol. 11, no. 8, p. 1131, Aug. 2022, doi: 10.3390/ANTIBIOTICS11081131/S1.

[20] W. Phothichaisri et al., “Potential Role of the Host-Derived Cell-Wall Binding Domain of Endolysin CD16/50L as a Molecular Anchor in Preservation of Uninfected Clostridioides difficile for New Rounds of Phage Infection,” Microbiol Spectr, vol. 10, no. 2, Apr. 2022, doi: 10.1128/SPECTRUM.02361-21/SUPPL_FILE/SPECTRUM02361-21_SUPP_1_SEQ8.PDF.

[21] W. Phothichaisri et al., “Potential Role of the Host-Derived Cell-Wall Binding Domain of Endolysin CD16/50L as a Molecular Anchor in Preservation of Uninfected Clostridioides difficile for New Rounds of Phage Infection,” Microbiol Spectr, vol. 10, no. 2, Apr. 2022, doi: 10.1128/SPECTRUM.02361-21.

[22] K. K. Mehta et al., “Newly identified bacteriolytic enzymes that target a wide range of clinical isolates of Clostridium difficile,” Biotechnol Bioeng, vol. 113, no. 12, pp. 2568–2576, Dec. 2016, doi: 10.1002/BIT.26029.

[23] Q. Wang, C. W. Euler, A. Delaune, and V. A. Fischetti, “Using a Novel Lysin To Help Control Clostridium difficile Infections,” Antimicrob Agents Chemother, vol. 59, no. 12, pp. 7447–7457, Dec. 2015, doi: 10.1128/AAC.01357-15.

[24] X. Wu, E. E. Paskaleva, K. K. Mehta, J. S. Dordick, and R. S. Kane, “Wall Teichoic Acids Are Involved in the Medium-Induced Loss of Function of the Autolysin CD11 against Clostridium difficile,” Scientific Reports 2016 6:1, vol. 6, no. 1, pp. 1–11, Oct. 2016, doi: 10.1038/srep35616.

[25] J. Bai, Y. T. Kim, S. Ryu, and J. H. Lee, “Biocontrol and rapid detection of food-borne pathogens using bacteriophages and endolysins,” Front Microbiol, vol. 7, no. APR, p. 189888, Apr. 2016, doi: 10.3389/FMICB.2016.00474/BIBTEX.

[26] Y. He et al., “Nonlytic Recombinant Phage Tail Fiber Protein for Specific Recognition of Pseudomonas aeruginosa,” Anal Chem, vol. 90, no. 24, pp. 14462–14468, Dec. 2018, doi: 10.1021/ACS.ANALCHEM.8B04160.

[27] S. P. Costa, C. R. F. Caneira, V. Chu, P. P. Freitas, J. P. Conde, and C. M. Carvalho, “A microfluidic platform combined with bacteriophage receptor binding proteins for multiplex detection of Escherichia coli and Pseudomonas aeruginosa in blood,” Sens Actuators B Chem, vol. 376, pp. 925–4005, 2023, doi: 10.1016/j.snb.2022.132917.

[28] P. Jones et al., “InterProScan 5: genome-scale protein function classification,” vol. 30, no. 9, pp. 1236–1240, 2014, doi: 10.1093/bioinformatics/btu031.

[29] M. Gouy, S. Guindon, and O. Gascuel, “SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building,” Mol Biol Evol, vol. 27, no. 2, pp. 221–224, Feb. 2010, doi: 10.1093/MOLBEV/MSP259.

[30] “AlphaFold Protein Structure Database.” Accessed: Jan. 08, 2024. [Online]. Available: https://AlphaFold.ebi.ac.uk/

[31] H. Sokol et al., “Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients,” Proc Natl Acad Sci U S A, vol. 105, no. 43, pp. 16731–16736, Oct. 2008, doi: 10.1073/PNAS.0804812105.

Chapter 6. Discussion

Concluding Remarks

The volume of endolysin-focused publications has expanded year on year over the last decade. According to the Web of Science database, 34 endolysin papers were published in 2013, compared to 119 papers in 2023. This research field is gaining increased interest largely due to the rise of antimicrobial resistance. Many of the reports in the literature highlight the positive aspects of endolysins as alternatives to traditional antibiotics. There is a particular focus on the lack of resistance development, ease of engineering, rapid lytic activity, and applications in biotechnology. This thesis set out to critically assess the therapeutic and diagnostic potential of endolysins, acknowledging the inherent advantages and challenges.

E. coli expression systems are a popular choice for the recombinant production of endolysins. For the vast majority of constructs created in this thesis, *E. coli* BL21(DE3) was more than suitable for protein production. This strain provided high yields of 6xHis-tagged proteins and is straightforward to genetically engineer. The exception to this general rule was the *C. acnes* endolysin, LysCa5A2, described in Chapter 2. The insolubility of this protein was a barrier in the assessment of the therapeutic potential of *C. acnes* endolysins. This observation is consistent with the current literature. Due to the lack of genetic diversity amongst *C. acnes* phages, it is likely that future isolates of *C. acnes* endolysins will encounter the same obstacles. This is an issue that could challenge the commercialisation of lysins for acne vulgaris, and it is possible that pharmaceutical companies may instead opt for alternative methods of treatment that require less fine-tuning, or that significant protein engineering will be required to overcome insolubility while retaining efficacy.

Isolation and characterisation studies of endolysins often follow a template that assesses the lytic activity of the endolysins against a target strain along with a host range performed against a small panel of other relevant strains. While these assays are valuable, and the results are important for future studies, we are limited by testing bacteria in a monoculture. As seen during the assessment of LysH1, the behaviour of bacteria in a monoculture can be substantially different to that in a community setting. The field of microbiome research is at an all-time high, and it is clear that when investigating a therapeutic candidate, the natural environment of the microbe needs to be considered. In a monoculture, the endolysin can act with high enzymatic

efficiency. The bacteria will likely grow to a high level in its optimal media, and the lytic effect will be significant. This result is not reflective of a microbiome. We fail to take into account the inhibitory effect of competitors, the possibility of the enzyme being physically sequestered by other inhabitants of the environment, and the risk of collateral damage to the surrounding population. It is important that future research gravitates towards more complex, microbiome-based research that can truly assess the impact of adding a lytic enzyme into a microbial ecosystem.

This same lack of complexity in many biological test systems is also evident in the use of endolysin binding domains for biosensor technology. The commercialisation of these phage-derived diagnostic tools is hindered by the lack of information regarding their effectiveness in microbe-rich environments. Chapter 5 of this thesis emphasizes the risk of non-specific binding and raises the question: are phage tail proteins a good fit for biosensor technology? The future of phage-derived diagnostics relies on optimization and should prioritise the use of multiplex systems to improve specificity. In the appendix of Chapter 5, we explore the potential of CdRBP as a receptor for an electrochemical biosensor. While a fluorescent construct of a single RBP may not be a suitable method for specific detection, electrochemical sensors provide an opportunity to develop multiplex systems. We propose that future electrochemical biosensors could be created with multiple *C. difficile* binding domains. The creation of a multi-domain sensor that has both RBP and CBD receptors could increase the specificity of these diagnostic tools. They can be designed in such a way that a positive result for *C. difficile* cells would only be received if multiple proteins report a change in charge. This research was conducted together with scientists from the Tyndall Institute and highlights the benefits of collaboration and emphasises how a multi-disciplinary approach to research can generate new ideas and drive problem-solving.

The cross-species activity of LysH1 was a particularly interesting finding. Chapter 3 was initially designed to investigate the therapeutic potential of endolysins to manage VRE. Testing revealed that enterococci were not particularly susceptible to the enzymatic activity of LysH1. However, the unexpected lytic activity against *R. gnavus* opened up a new avenue of research for LysH1. This was a fortuitous observation and led to the possibility of employing LysH1 as a bacterial editing tool in simplified community studies. While the testing of LysH1 against *R. gnavus* was

largely down to chance (as the strain was being worked on in parallel for Chapter 4), it does draw attention to the need for testing endolysins against a broad host range of bacterial genera. LysH1 has shown promise as a therapeutic candidate. Going forward, it would be interesting to assess LysH1 *ex vivo* in a faecal fermentation and then in pre-clinical models. The use of IBD and healthy faecal samples *ex vivo* would help us to assess the ability of LysH1 to act in a microbiome setting. Metagenomic analysis would provide evidence for (or against) the specificity of LysH1 and would allow us to investigate population dynamics on a large scale.

The observation of endolysin-resistant mutants in Chapter 4 of this thesis is a new phenomenon. The bacterial stringent response is obviously central to this event, but it would be interesting to determine how or if this pathway is compromised in the mutant strains that didn't show genomic changes that could be directly linked to (p)ppGpp, to further investigate the mechanisms at play. The emergence of resistance should be a worrying observation for the field of endolysin therapy. However, the fitness costs associated with resistance mechanisms encountered in this study indicate that even if resistance were to occur in a bacterium within a patient, it is highly unlikely that the mutant would be able to proliferate and survive in that environment. These mutants were difficult if not impossible to maintain in a laboratory setting, never mind in a competitive microbiome. There is no doubt that endolysin-resistant mutants will be reported upon in the future, and it will be critical to assess the consequence of these mutants on a case-by-case basis. Given the ability of endolysins to be genetically modulated, the engineering of lysins will be an important tool for tackling resistant strains.

Overall, I believe that endolysins do have therapeutic potential. However, to become a viable alternative or adjunct to antibiotics, they must be rigorously tested and their limitations should be clearly acknowledged. Lack of resistance mechanisms and high specificity are some of the key selling points for endolysin therapy, and yet, in this body of work, we can account for three instances where this wasn't the case (resistance to LysIBDN1, cross-species activity of LysH1, and non-specific binding of CdRBP). Nonetheless, with an emphasis on rigorous tests in community settings, detailed genomic and phenotypic analysis, and assessing efficacy in clinical samples, endolysin therapy could be a valuable tool for detecting and identifying pathogens, in helping to treat infections, and in managing antimicrobial resistance.

Acknowledgements

To my dad, Pat.

Though you are no longer here, your memory continues to inspire me.

In your too short time on this earth, you taught me curiosity, creativity, kindness, and patience. From looking at leaves under a microscope as a child to picking me up from my leaving cert study sessions, you always fostered my interest in science. I will be forever grateful for your positive influence and belief in me.

This thesis is dedicated to you.

My PhD journey has been amazing, but it wasn't without its ups and downs. I count myself so very lucky to have had an outstanding support network of family, friends, and colleagues to help me every step of the way.

Firstly, to my supervisors, Professor Colin Hill and Professor Paul Ross, I am deeply grateful for your unwavering advice, guidance, and support over the past four and a half years. I cannot express enough how much I appreciate every PhD meeting, lab talk discussion, and office visit where you both patiently and enthusiastically listened and helped me plan experiments. You have not only shaped me into the researcher I am today but gave me the confidence to begin a career in academia. I am indebted to you both for the opportunity to pursue my PhD under your supervision.

To everyone (past and present) in the Hill/Ross labs, thank you so much for everything you have done for me. I would never have made it without such a kind group of people around me. There was never a time when I felt alone or that I couldn't ask someone nearby for a helping hand. From life advice to science advice, laughter, quiz nights, conference trips, complaining about experiments together, and numerous dog visits, I am completely convinced we have the best lab groups going!

There are some people that I have to give special mentions to from the lab group. Andrey, thank you for your wisdom and guidance. Steve, thank you for Novogino. Chris, thank you for all the stories, tea, and brainstorming. Sam, thank you for all the lab chats, second opinions and weekend company (but I would like to distinctly not thank you for the smell of the fermenters). Fabian and Fergal, thank you for your off-the-wall conversations that kept me endlessly entertained. Remi, thank you for your bioinformatic help and for always making sure I know the cat quote of the day. Nati, you are the qPCR queen and are always so helpful. Julie, Miguel, Mariana, Des, Felipe, and Cara, thank you for being the most supportive postdocs and endless sources of knowledge.

Lorraine, thank you for being an amazing mentor when I was trying to find my science feet. You always listened to my problems and gave me great advice, picked me up when I was feeling lost and created such a great lab environment to work in. Neda, there was no one else I would rather take over as lab manager. You made the final stretch of my PhD possible with your support, kindness, Cheeto visits and encouragement. I am so grateful to have had such wonderful lab managers!

Ivan, thank you for your kindness and wisdom, and for being so much fun. Daragh, thank you for your pep talks, hilarious ranting, and great chats over a pint. You are both amazing, and I will always be grateful for all the offers of help you both gave me and for never making me feel silly for asking stupid questions.

To my dearest friends outside the lab. Lisa, our coffee dates have kept me sane, and you are a constant source of reassurance. You are always there to listen, laugh and keep me going. Your support never goes unnoticed. Sadhbh, I always count myself lucky to have randomly ended up living with you. You are the most hilarious person. Thank you for always being understanding and giving me a reality check when I need it. I can't wait to be annoying you more often now! Chloë, though you are a million miles away, I have never felt distanced from you. 10 years later, despite time zones and busy schedules, you still have my back, always make time for me, and I have always felt your support. You're one in a million, Champi forever.

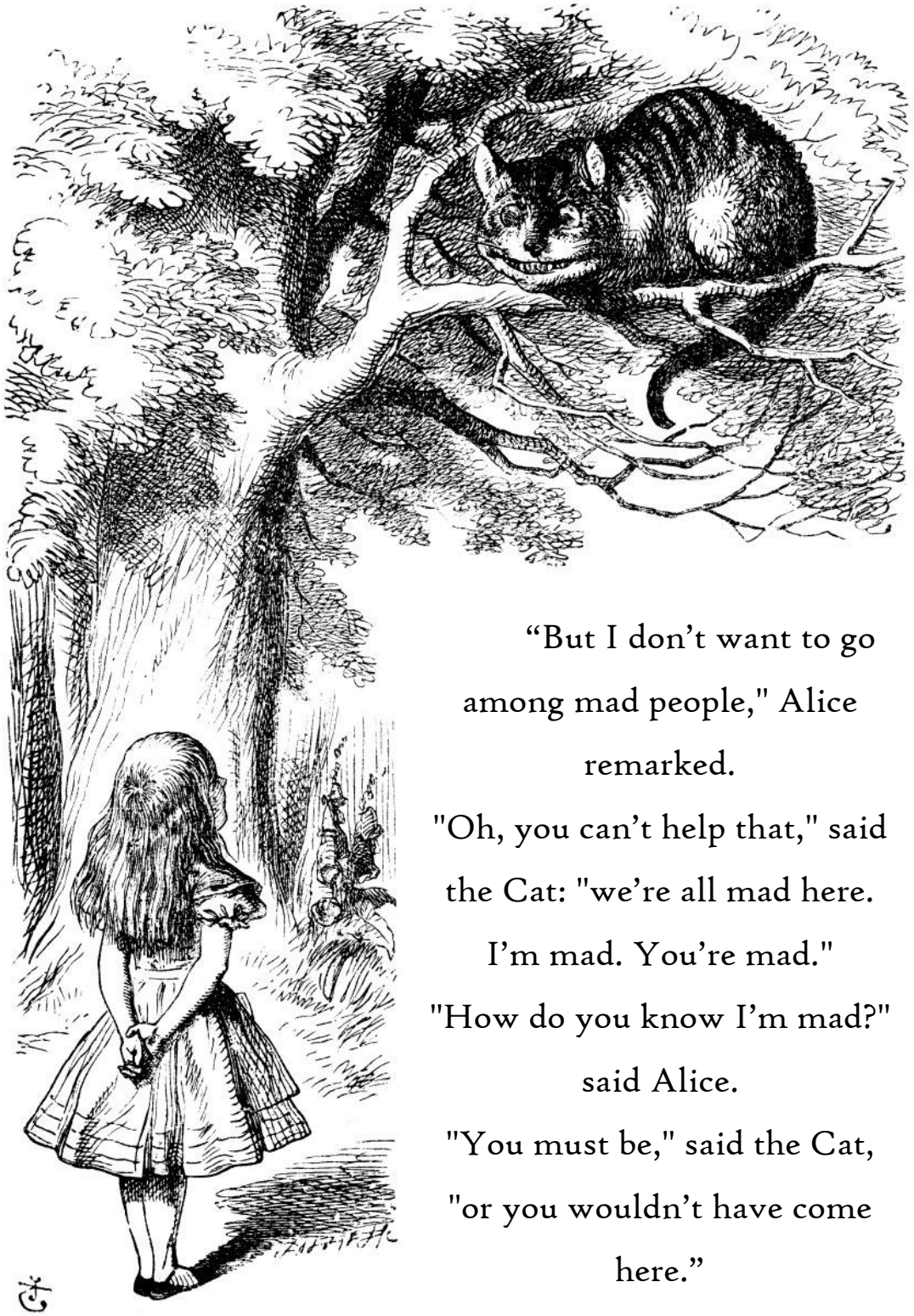
To my girlies that started as lab friends and are now forever friends. Katie, thank you for your positive energy, lunchtime chats and nights out. You have been a great addition to the 4th floor! Ciara, thank you for moving lab benches so I could torment you more easily and thank you for the laughter, shared panic (ak, ak, ak) and for listening to my rants. Hiba, thank you for your kindness, desk treats, and coffee making. Your ability to make people feel included is unmatched. You have both been the most amazing friends to me, and I can truly say I have made connections for life with you both. Julie, thank you for befriending me when I started in the phage lab. You are such a source of positivity and warmth. The most wholesome girlie, and life is better with you in it. Michelle, I can't even begin to put words on how strong you are. You have shown incredible resilience over the past few years and have had to deal with more than we could ever imagine, yet you kept laughing and entertaining us throughout! Thanks for always being a source of laughter and cheering me up when I needed it, and for the oat-milk flat whites. Lauren, thank you for being the most thoughtful human being. I couldn't have done this without you. You are hilarious, so much fun to be around and I count myself very lucky to have your friendship.

Sal, *felis catus*, you can't read this, but thanks for being the best furry friend. No matter how busy the lab was, coming home to a cosy kitten always brightened my day. I am so grateful to you (and to Dave for surprising me with you).

Dave, I can’t even begin to describe how much you mean to me. I am so glad we got to experience this journey together. Thank you for all your brainstorming, for making me hundreds of coffees, for listening to me giving out about bacteria, for patiently teaching me computery stuff and everything in between. Most importantly, thank you for believing in me when I didn’t believe in myself. You are an amazing source of optimism, silliness, ideas, and support. I am so lucky to have you in my life, and I look forward to our next adventures together.

Mam, without you none of this would be possible. You have always encouraged me to follow my dreams and have provided me with the support and belief to do so. I am so very grateful to have such a kind and thoughtful role model. Thank you for listening to me, for all your advice and motivation and for all the post you sent me over the years! You’re the best.

None of this work would have been possible without such a wonderful group of people.



“But I don’t want to go among mad people," Alice remarked.

"Oh, you can’t help that," said the Cat: "we’re all mad here. I’m mad. You’re mad."

"How do you know I’m mad?" said Alice.

"You must be," said the Cat, "or you wouldn’t have come here.”