A Treatment for *Clostridioides difficile* Based on Cell Wall Lysins Isolated from *C. difficile* Specific Bacteriophages

Khalid Abdullah Alyahya

School of Pharmacy and Pharmaceutical Sciences, Cardiff University

Thesis submitted to Cardiff University as a requirement for the degree of Doctor of Philosophy (PhD)

April 2024

Plagiarism statement

This project is conducted in the School of Pharmacy and Pharmaceutical Sciences at Cardiff University under the supervision of Prof. Les Baillie. I certify that the research described in this work is original and that any parts of the work that I have been conducted in collaboration are clearly indicated. I also certify that I have written all the text herein and have clearly indicated by suitable citation any part of this project that has already appeared in publication.

Khalid Abdullah Alyahya

Acknowledgements

I am deeply grateful to have had the guidance and support of numerous individuals throughout my doctoral journey. First and foremost, I would like to express my sincerest appreciation to my supervisor, Prof. Les Baillie, whose support, mentorship, and expertise have been invaluable in shaping this research. I am also grateful to my second and third supervisors, Dr. Charles Heard and Dr. Youcef Mehellou, for their feedback and guidance in my journey.

Special thanks go to Prof. Andrea Brancale, Dr. Salvatore Ferla, and Dr. Marcella Bassetto for their assistance and guidance in aligning protein structures. I am grateful for their expertise and collaborative spirit. I am grateful to Dr. Michael Pascoe for his guidance and expertise in understanding electron microscopy, which was helpful in clarifying and advancing my research.

To my friends and labmates, thank you for your support and countless stimulating discussions. Your contributions have enriched my academic experience.

I am deeply appreciative of the generous support provided by my sponsor, King Saud University, without which this research would not have been possible.

Last but certainly not least, I extend my profound gratitude to my parents, my wife, and my children. Your love, encouragement, and sacrifices have been the cornerstone of my academic pursuits. I am grateful for your support and understanding throughout this journey.

To every individual who has played a positive part, no matter how small, in my journey, I offer my heartfelt thanks. Your contributions have made a positive impact on my project and professional development.

Publications and presentations related to the study

Publication

Alyahya, K.; Baillie, L. Assessing the Feasibility of Employing a Combination of a Bacteriophage-Derived Endolysin and Spore Germinants to Treat Relapsing *Clostridioides difficile* Infection. *Microorganisms* 2023, *11*, 1651. https://doi.org/10.3390/microorganisms11071651

Poster Presentations

A Treatment for *Clostridioides difficile* Based on Cell Wall Lysins Isolated from *C. difficile* Specific Bacteriophages. **Khalid Alyahya**, Les Baillie and Charles Heard. Pharmacy school postgraduate research day, School of Pharmacy and Pharmaceutical Sciences, Redwood Building, Cardiff University, Cardiff, UK (April 2021).

A Treatment for *Clostridioides difficile* Based on Cell Wall Lysins Isolated from *C. difficile* Specific Bacteriophages. **Khalid Alyahya**, Les Baillie and Charles Heard. Speaking of Science Conference, Doctoral Academy, Cardiff (May 2022).

Oral presentation

A Treatment for *Clostridioides difficile* Based on Cell Wall Lysins Isolated from *C. difficile* Specific Bacteriophages (extended). **Khalid Alyahya**, Les Baillie and Charles Heard. Pharmacy school postgraduate research day, School of Pharmacy and Pharmaceutical Sciences, Redwood Building, Cardiff University, Cardiff, UK (April 2022).

Abstract

Clostridioides difficile is a Gram-positive, anaerobic, spore-forming bacillus and is a major cause of healthcare-associated infections. While the vegetative form of the pathogen is susceptible to antibiotic treatment, its ability to persist in the gut as antibiotic-resistant spores leads to reinfection in cases where protective microflora is not reestablished. Utilizing recombinant endolysins as treatment of CDI is promising because of their activity against antibiotic and bacteriophage resistant strains. In this study, the recombinantly expressed endolysin LysCD6356 and its enzymically active domain (EAD) were shown to lyse the vegetative forms of a panel of clinical isolates of *C. difficile* from the UK, including hypervirulent 027 ribotype strains. While these results were promising endolysins, like antibiotics and bacteriophages, endolysins have no effect against the spore form of the pathogen, which is responsible for recurrent infections following successful treatment.

To address this issue, a combination of germinants and a bactericidal agent were employed to target the more sensitive form of the pathogen. In this study, exposing of *C. difficile* spores to germinants followed by endolysin several hours later proved to be the most effective approach suggesting a promising approach to treat relapsing CDI. Additionally, the binding of LysCD6356 and its EAD to the spore and vegetative forms of *C. difficile* would leave the endolysin ideally placed to target the emerging vegetative cells. This also raise the possibility of developing a diagnostic tool that could be used to detect the pathogen.

Bioinformatic studies revealed a significant similarity between LysCD6356 and its EAD to CD27L and $CD27_{1\text{-}179}$, respectively, suggesting a possibility of sharing common motifs. However, there was no similarity of LysCD6356 and its EAD with previously published spore binding domains, suggesting the presence of a new spore binding region. An attempt to investigate the binding of endolysins to spore surface proteins, CdeC and CdeM, was unsuccessful possibly due to deficiencies in the experimental approach or the fact that the endolysins recognize other spore surface targets. The combination of our strategy of targeting spores with more traditional approach, such as fecal microbiota transplant, could provide more efficient treatment of relapsing CDI.

Table of Contents

List of tables:

List of Figures:

List of abbreviations

ATCC ---------- American Type Culture Collection
BCA ---------- Bicinchoninic acid
BHI ---------- Brain-Heart Infusion medium
BLAST ---------- Basic Local Alignment Search Tool
Ca-DPA ---------- Calcium dipicolinic acid
CBD ---------- Cell wall binding domain
CCFA ---------- Cycloserine-Cefoxitin-Fructose Agar
CDI ---------- *Clostridium difficile* infection
CDMN---------- *Clostridium difficile* Moxalactam Norfloxacin
CFU ---------- Colony forming unit
Clustal omega ---------- Multiple sequence alignment program
DNA ---------- Deoxyribonucleic acid
E-value ---------- Expected value
E. coli ---------- *Escherichia coli*
EAD ---------- Enzymatic Active Domain
EDTA ---------- Ethylenediamine-tetra acetic acid
GI tract---------- Gastrointestinal tract
I-TASSER ---------- Iterative Threading ASSEmbly Refinement
IPTG ---------- Isopropyl β-D-1-thiogalactopyranoside
kDa ---------- Kilo Dalton
LB ---------- Luria-Bertani medium
MOE---------- Molecular Operating Environment program
MWCO ---------- Molecular wight cut-off
NAG ---------- *N*-acetylglucosamine
NAM ---------- *N*-acetylmuramic acid
NCBI ---------- National Centre for Biotechnology Information
NCTC ---------- National Collection of Type Culture
Ni-NTA ---------- Nickel-nitrilotriacetic acid
OD ---------- Optical density
PBS ---------- Phosphate-Buffered Saline
PBST ---------- PBS-tween-20
PCR ---------- Polymerase chain reaction
PDB ---------- Protein Data Bank
PG ---------- Peptidoglycan
PVDF ---------- Polyvinylidene fluoride
SDS ---------- Sodium Dodecyl Sulphate
SDW ---------- Sterile Distilled Water
SEM ---------- Scanning electron microscope
SOC medium ---------- Super Optimal broth with Catabolite repression
Tc ---------- Taurocholate
UV ---------- Ultraviolet

Chapter 1 General introduction

1.1 Introduction

1.1.1 Historical review

In 1935, Hall and O'Toole were the first to isolate *Clostridioides difficile* (Hall and O'Toole 1935). Due to the shape and the difficulties in isolating the organism, it was originally called *Bacillus difficile*. However, due to the microorganism being a strictly anaerobic Gram-positive bacilli with large and sub-terminal spores, it was later reclassified as a member of Clostridium genus, and renamed as *Clostridium difficile* (Brazier and Borriello 2000). Recently, Lawson and colleagues proposed to reclassify *Clostridium difficile* as *Clostridioides difficile*, to highlight the differences between *C. difficile* and other members of Clostridium genus (Lawson et al., 2016). The organism is commonly known as *C. diff*. or *C. difficile* (Lawson et al., 2016).

C. difficile is ubiquitous in nature and can be found in soil, medical facilities, water, human and animal gastrointestinal tract (GIT) (Janezic et al., 2016; Dharmasena and Jiang 2018). The replication and growth of the vegetative form of *C. difficile* has been reported in the animal and human GIT, while dormant spores have been recovered from environments, such as soil and puddle water (Janezic et al., 2016).

1.1.2 Characteristics of *Clostridioides difficile*

C. difficile exists in one of two forms, depending on the environment. When conditions are conducive to growth, it forms vegetative bacteria, but when the conditions become hostile to survival, it can form resistant, biologically inert spores (Figure 1.1).

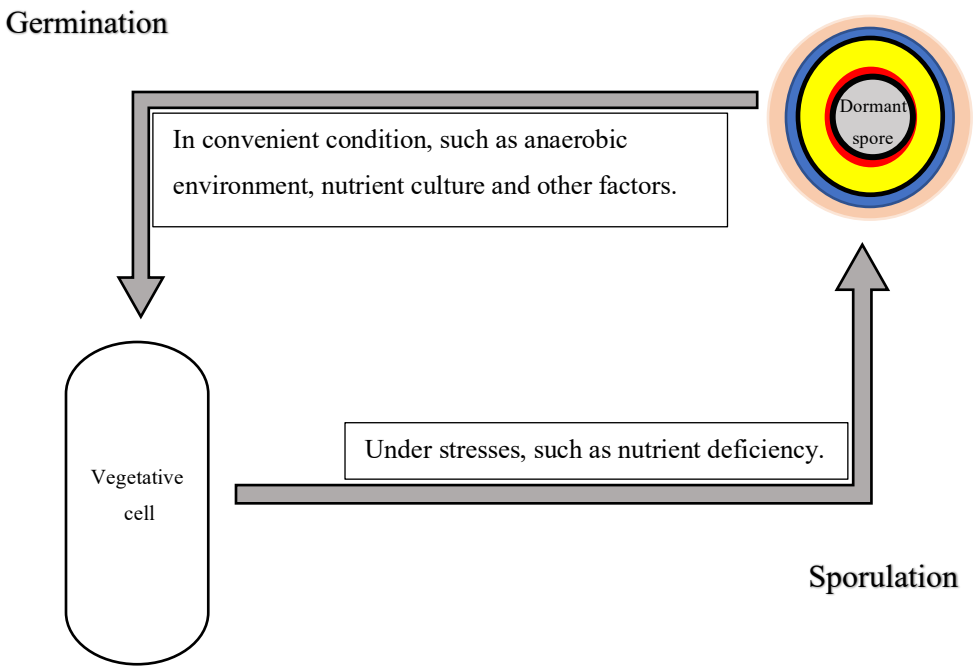


Figure 1.1 The different physical forms of *C. difficile*.

C. difficile has two forms, the vegetative and spore forms. The vegetative form initiates the infection and produce spores to be able to protect itself from antibiotics and other treatments and to be able to transmit to other individuals. The transition between the two forms occurs through processes called sporulation and germination.

1.1.3 The genome of *C. difficile*

In 2006, Sebaihia and colleagues determined the *C. difficile* genome sequence for the first time. The isolate which was sequenced was a virulent, multidrug resistant strain called *C. difficile* 630 which was isolated from a patient suffering from severe pseudomembranous colitis (PMC). The sequencing revealed the genome to be composed of a circular chromosome of 4,290,252 bp with a GC content of 29.06% and a 7,881 bp plasmid with a GC content of 27.9% (Sebaihia et al., 2006). This contrasts to the average GC content of *E. coli* which is 50.6% (Mann and Chen 2010).

This is an important point to be considered when attempting to express *C. difficile* genes from an *E. coli* host as the efficiency of gene expression is affected by the codon preference of the host (Gustafsson et al., 2004). Strategies to mitigate the impact of gene codon differences are described in detail below (section 1.1.5).

Among all bacterial species, *C. difficile* has the highest level of genomic diversity (Knight et al., 2015). Stabler and colleagues screened 75 *C. difficile* strains and found that only 734 genes (19.7%) out of 3723 genes were conserved (Stabler et al., 2006). Additionally, another study screened 74 strains and showed only 153 genes were conserved across tested strains. In other words, approximately 16% of whole *C. difficile* genome is highly conserved (core genome), while the rest (approximately 84%) represents accessory genes (Janvilisri et al., 2009). Indeed, most of the conserved genes are responsible for essential cellular activities, such as metabolism, biosynthesis, and cellular and regulatory processes (Stabler et al., 2006). Additionally, large percentage of the bacterium genome (11%) represents mobile genetic elements (Sebaihia et al., 2006). The relatively low level of core gene conservation coupled with the presence of mobile genetic elements go some way to explaining the marked genetic diversity of *C. difficile* as a species.

Bacterial typing methods are an essential tool for studying epidemiology. Through the last two decades, several typing techniques have been developed for *C. difficile*. Three typing methods, pulsed field gel electrophoresis (PFGE), restriction enzyme analysis (REA) and polymerase chain reaction ribotyping (PCR ribotyping), have commonly been used. PFGE and REA were used in North America, while PCR-ribotyping predominates in Europe.

While PFGE technique utilizes a rare restriction enzyme, such as *SmaI*, REA employs a common restriction enzyme, usually *HindIII* (Clabots et al., 1993; Northey et al., 2005). Both techniques utilize restriction enzymes to digest genomic DNA and the resulting DNA fragmentation patterns are compared in order to differentiate between strains. In the PCR ribotyping method, primers for conserved regions within the 16S and 23S ribosomal genes are amplified. Subsequently, variations in intergenic spacers are revealed by running the PCR products on an agarose gel (Bidet et al., 1999). An example of the results obtained for the same isolate using different typing methods resulted in the classification of a clinical isolate BAA-1803 as PCR ribotyping 027 and REA type B1 (027 / B1) (Figure 1.2).

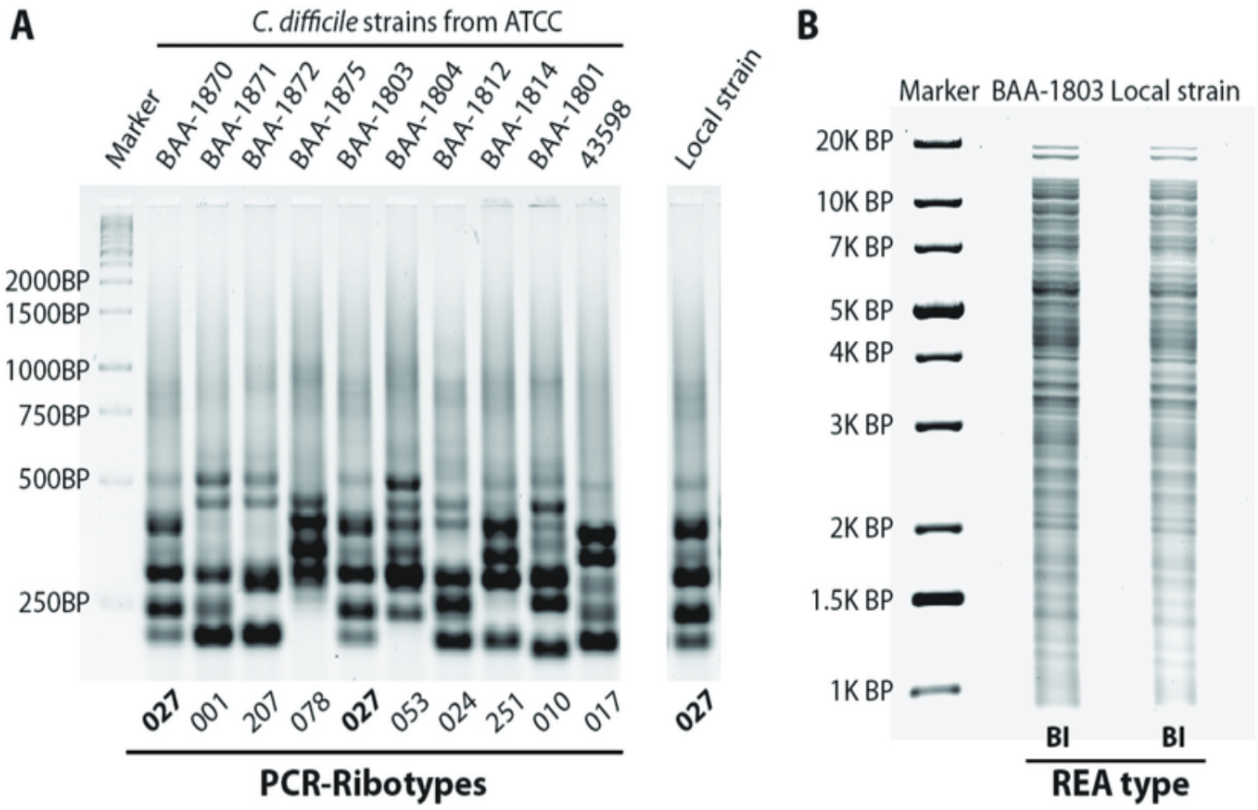


Figure 1.2 Genetic typing of *C. difficile.*

Representation of different genetic typing of *C. difficile* including PCR-ribotyping (A), REA (B) (Shen et al., 2016).

Newer genomic techniques, such as whole genome sequencing (WGS) and multilocus variable number repeat analysis (MLVA), provide a higher level of differentiation between *C. difficile* strains. MLVA targets tandem repeats, which are short repeated sequences within the DNA. Each bacterium has a characteristic of these repeats of a set length, which facilitates discrimination between strains (Van Belkum 2007).

WGS facilitates the typing of several bacterial strains, especially closely related strains, comparing whole bacterial genome instead of small genome fragments and provides the opportunity to study epidemiology at the molecular level (Eyre et al., 2019; Janezic and Rupnik 2019).

1.1.4 Vegetative *C. difficile*

C. difficile is an obligate anaerobic bacterium, indicating that the bacterium ferments carbohydrates and amino acids in order to make energy (ATP). In energy metabolism, five amino acids are essential: (leucine, isoleucine, tryptophan, proline, and valine). Additionally, one amino acid (glycine) enhances bacterial growth significantly, especially at 37 ºC, the optimal temperature for growth (Gibbs 2009). The bacterial shape is rod and usually arranged in pairs. In terms of size, the bacterium's length ranges from 3 to 4 µm with curved edges (Figure 1.3). Older colonies of *C. difficile* may exhibit Gram stain variability even though the bacterium is predominantly Gram positive (Brazier and Borriello 2000). While *C. difficile* strains utilize peritrichous flagella for motility, strains vary in the motility from motile strains (*C. difficile* 630) to non-motile strains (CM-26) due to the absence of the genes encoding the flagella (Stabler et al., 2006; Twine et al., 2009; Purcell et al., 2012).

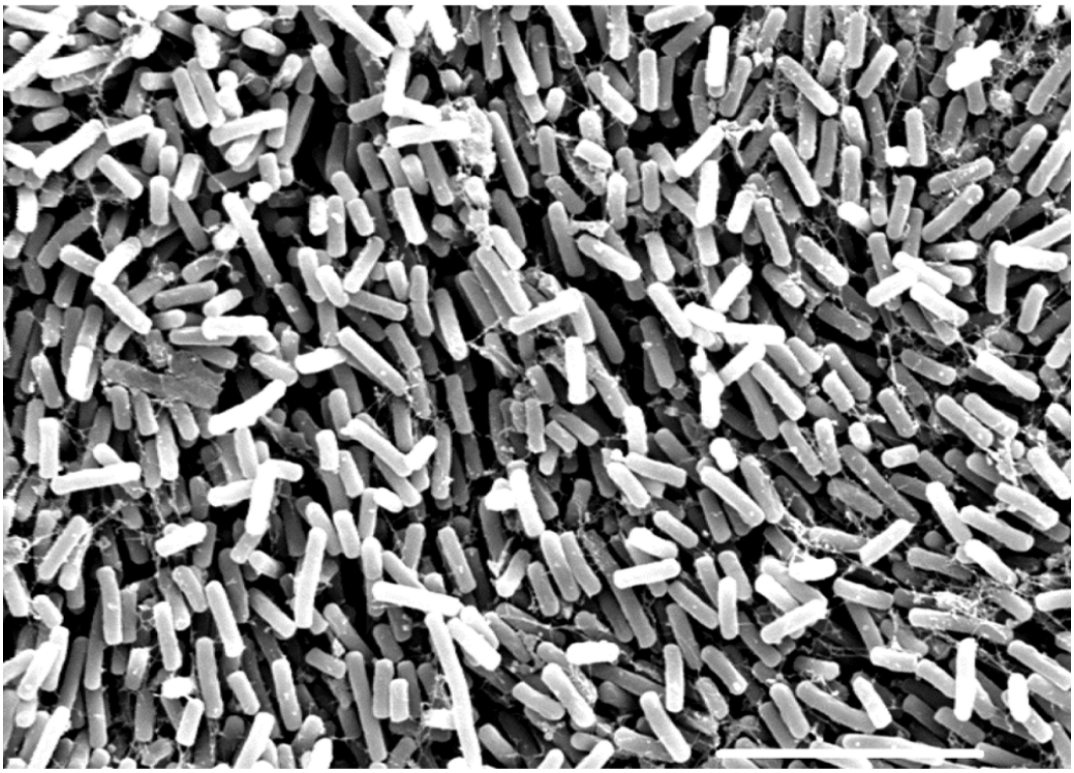


Figure 1.3 Scanning electron microscopic of *C. difficile* vegetative cells.
C. difficile BI17 following three days of cultivation in BHI. Scale bar is 10 µm (Semenyuk et al., 2014).

As a Gram positive bacterium, the cell wall is composed of teichoic acid and a thick peptidoglycan layer (up to 80 nm) compared to Gram negative bacteria (10 nm) (Figure 1.4). Peptidoglycan has an alternative repetition of two amino sugars forming linear chains. These amino sugars, N-acetylglucosamine (GlcNAc or NAG) and N-acetylmuramic acid (MurNAc or NAM), are connected via a β-(1,4) glycosidic bond. Each MurNAc, located between two GlcNAc, is connected to a small peptide chain (approximately 5 amino acids). Each chain of amino sugar connects to the next amino sugar chain through peptide chains. All of the chains are connected to each other, forming a mesh structure. Peptidoglycan, sometimes called murein, plays an essential role in protecting the bacterium from osmotic lysis.

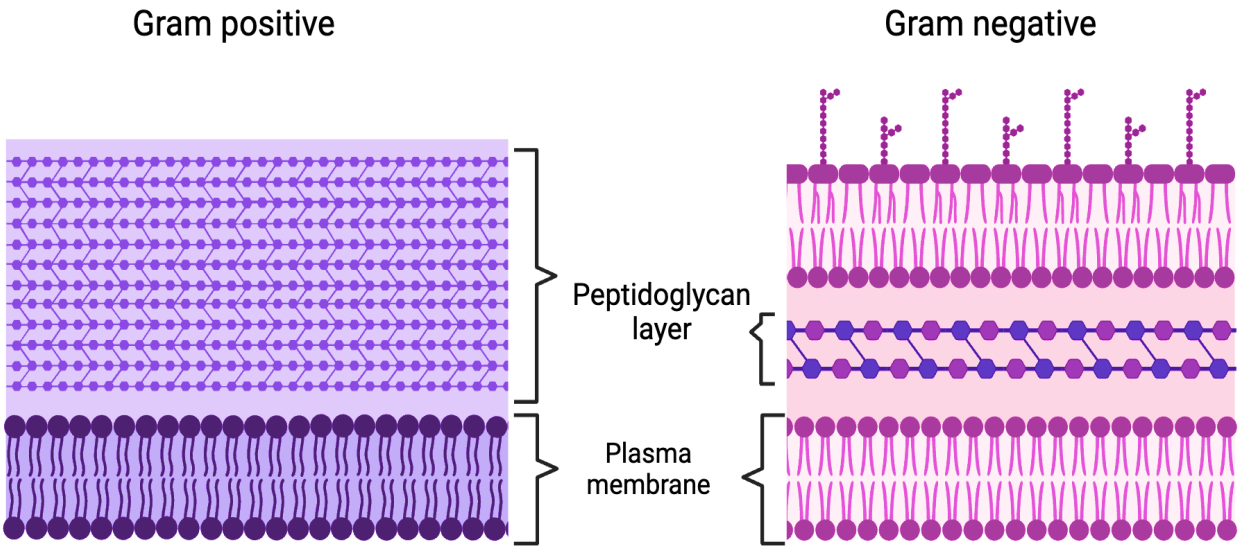


Figure 1.4 Gram positive and Gram negative cell wall.
The outer membrane is missing in Gram positive bacteria, but it has a big layer of peptidoglycan compared to Gram negative bacterial cell wall (Nazzaro et al., 2013).

The morphology of *C. difficile* colonies varies and depends on the strain, incubation period and culture medium. Generally, after 24 to 48 hours of anaerobic incubation on blood-based culture media, colonies have irregular edges, are grey in colour, are opaque, and show no haemolysis on blood agar. The appearance of the colonies is similar to that of a fried egg as a result of the formation of spores in the centre of the colonies due to a deficiency of nutrients (Figure 1.5)

(Brazier and Borriello 2000). The diameter of the colonies varies based on the media. For example, they range from around 4.5 mm on blood agar to approximately 7.5 mm on cycloserine cefoxitin fructose agar (CCFA) (George et al., 1979).

Figure 1.5 Colony morphology of *C. difficile.*
C. difficile NCTC 12727 on Blood agar under anaerobic condition at 37º C for 48 hours.

Under aerobic conditions, the vegetative form of *C. difficile* is able to survive for a limited period of time and must form spores if it intends to survive in the presence of oxygen. *C. difficile* is strictly anaerobic, which indicates a lack of detoxification pathways such as peroxidase, catalase, and superoxide dismutase. As a consequence, exposure to oxygen results in the accumulation of reactive oxygen species, such as H_2O_2, leading to cell death (Oliver Riebe 2009). The survival time of the vegetative form of the bacterium in the presence of oxygen varies from minutes to hours, depending on factors such as humidity. For example, in the presence of oxygen, *C. difficile* is able to survive for up to six hours in a humid environment (Jump et al., 2007), while it can only survive for 15 minutes in a dry environment (Buggy et al., 1983).

1.1.5 *C. difficile* spores

Since *C. difficile* is a strictly anaerobic Gram-positive bacilli, oxygen exposure, nutrient deficiency, stomach acidity, and other stressors threaten the life of *C. difficile* vegetative cells. To overcome these issues, *C. difficile* forms a structure known as a spore, a structure that was described for the first time in *Bacillus subtilis* by Robert Koch and Ferdinand Cohn (Henkin 2016). The dimensions of *C. difficile* spores vary depending on the isolate with an average length of 1.0 – 1.5 µm and an average diameter of 0.5 – 0.7 µm (Snelling et al., 2010). Spores of *C. difficile* can survive under aerobic conditions, and in this dormant and non-reproductive state, it can persist from months to years (Rupnik et al., 2009).

In general, spores are composed of a spore core, inner spore membrane, germ cell wall, cortex, outer spore membrane, and spore coat, as represented in Figure 1.6 (Zhu et al., 2018). A structure known as an exosporium, which was first named by Flugge in 1886, surrounds some but not all bacterial spores (Gerhardt and Ribi 1964).

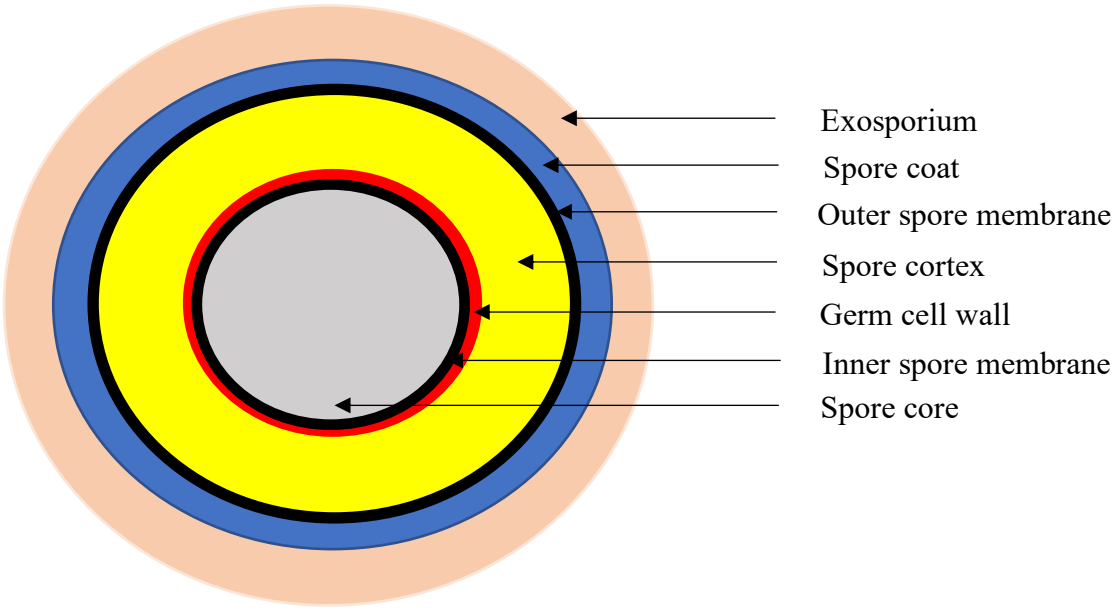


Figure 1.6 Bacterial spore structure showing structural layers.
The diagram is not to scale. Modified from Zhu, Sorg and Sun (2018).

In terms of content and function, the outer exosporial layer is mainly composed of proteins, polysaccharides, and lipids (Matz et al., 1970), in contrast to the inner and outer spore membranes, which are primarily composed of phospholipid. The exosporial layer is believed to participate in epithelial cell attachment in addition to protecting the spore (Joshi *et al.,*, 2012). The spore coat, which has a small role in regulating spore germination, increases the strength of the spore and protects it from degradation in the GI tract. The spore's cortex is composed mainly of peptidoglycan, which is believed to increase the stability of the spore and has a rapid degradation rate during germination. The bacterial genome, ribosomes, mRNA, and proteins are located within the spore's core (Zhu et al., 2018).

In terms of spore structure, differences have been observed between *C. difficile* strains. While some strains possess an exosporial layer, such as *C. difficile* DS1813, others lack this layer, such as DS1771 and DS1748 (Figure 1.7) (Joshi *et al.,*, 2012). Moreover, the exosporium layer in the *C. difficile* 630 strain appears to be a smooth and electron dense layer, while for strain R20291, it exhibits the presence of hair like structures (Paredes-Sabja et al., 2014). The exosporial layer of *C. difficile,* when present, is directly attached to the spore coat, unlike that of members of the *B. cereus* group, in which there is a distinct space between the exosporial layer and the spore coat (Calderón-Romero et al., 2018).

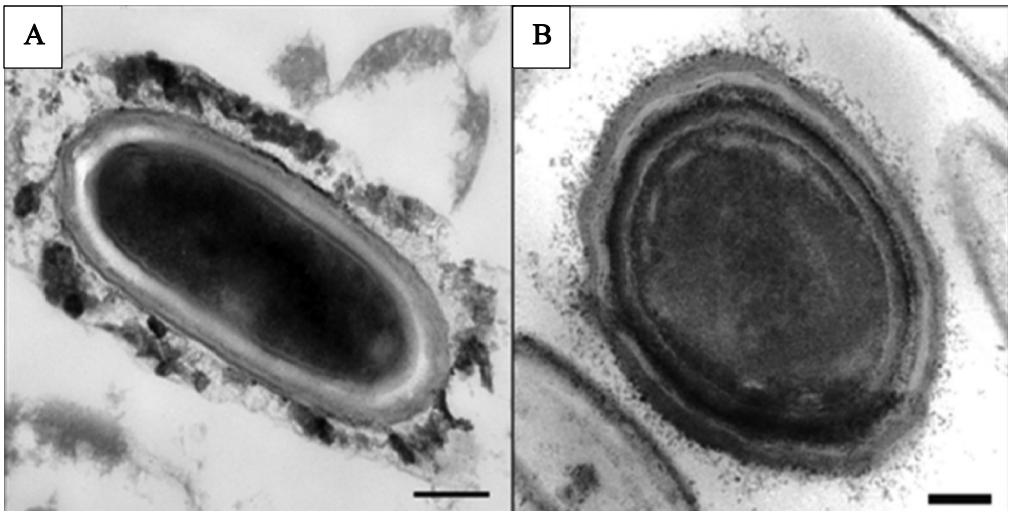


Figure 1.7 Transmission electron microscopic images of *C. difficile* spore.
TEM images of varies *C. difficile* spores produced by DS1813 (A) with visible exosporial layer and DS1771 (B) with an invisible exosporial layer. Scale bars length is 100 nm. (Joshi et al.,, 2012).

The exosporial layer is composed of several proteins, including collagen like glycoproteins such as BclA1, BclA2, and BclA3, in addition to cysteine rich proteins such as CdeA, CdeB, CdeC, and CdeM (Paredes-Sabja et al., 2022). Researchers focused on some exosporial proteins and found they participate in important structural and functional features in the spore, such as CdeC and CdeM, while other exosporial proteins are not fully investigated (Barra-Carrasco et al., 2013; Antunes et al., 2018; Calderón-Romero et al., 2018).

CdeC and CdeM are highly conserved proteins in *C. difficile* strains that have been identified to date. They are described as *Clostridioides difficile* exosporium cysteine rich proteins due to the presence of relatively large numbers of cysteine residues. CdeC is comprised of 405 amino acids, including 37 cysteine residues, which represent 9% of the whole protein (Calderón-Romero et al., 2018). CdeC is believed to be involved in connecting the exosporial layer to the spore coat (Stewart 2015). It plays an important role in enhancing spore resistance to external stresses since its removal increases the permeability of spores to ethanol and makes them more vulnerable to heat (Barra-Carrasco et al., 2013).

CdeC is conserved in *C. difficile* and in all *Peptostreptococcaeace* family members and has been shown to be immunogenic and capable of protecting rodents from a lethal *C. difficile* spore

challenge (Ghose et al., 2016; Pizarro-Guajardo et al., 2018). Moreover, Díaz-González and colleagues revealed that spore associated CdeC is accessible to antibodies, suggesting that it is surface exposed, making it a potential target for therapeutic agents (Díaz-González et al., 2015).

A second exosporial protein, CdeM, plays an important role in delaying the germination process. It is comprised of 161 amino acids, including 14 cysteine residues, which represent 8.7% of the whole protein (Calderón-Romero et al., 2018). The protein is unique and conserved in *C. difficile* (Pizarro-Guajardo et al., 2018). As was the case for CdeC, the protein, when used as a vaccine, was able to protect rodents from a lethal *C. difficile* spore challenge (Ghose et al., 2016). Its surface location suggests that, like CdeC, it could provide a spore specific target for the rational design of therapeutic agents (Barra-Carrasco et al., 2013; Díaz-González et al., 2015).

The characterization of these proteins has been made possible by the development of *E. coli* based recombinant protein expression systems. To date, both proteins have been expressed in different *E. coli* strains and in different vectors in order to enhance protein expression and solubility. For example, *E. coli* BL21(DE3)-Ril has been employed to express CdeC due to the fact that it has transfer RNA molecules which recognize codons which are rarely seen in *E. coli* but are relatively common in Gram positive bacteria such as *C. difficile* (Brito-Silva et al., 2018). Additionally, the SHuffle T7 *E. coli* host has been employed to express CdeC, as it supports the formation of disulfide bonds in cysteine rich proteins, which enhances protein folding (Lobstein et al., 2012). BL21(DE3) has also been utilized to express codon optimized versions of CdeC and CdeM in an effort to overcome codon differences (Ghose et al., 2016).

1.1.6 Sporulation

Sporulation is a process by which the bacterium converts from its vegetative form to a resistant spore. This process is stimulated by several environmental stresses, such as low nutrients availability and other unidentified stresses. Deakin and colleagues revealed that Spo0A is a major transcriptional regulator which plays an essential role in the sporulation of *C. difficile* (Deakin et al., 2012). Under stress, cognate membrane-associated sensor histidine kinase (Kin) senses and

phosphorylates Spo0A to form Spo0A-P, which triggers the sporulation process (Deakin et al., 2012; Pettit et al., 2014).

Sporulation specific sigma factors, especially σ^E, σ^F, σ^G, and σ^K play an important role in the sporulation process (Fimlaid and Shen 2015) and are under the control of Spo0A regulation. Pairs of sigma factors are specific for one of the two compartments of sporulation process (the mother cell and the forespore). So, σ^E and σ^K are specific for the mother cell while σ^F, σ^G are specific for the forespore and play an essential role in the development of the forespore (Zhu et al., 2018).

The sporulation process involves five main stages (Figure 1.8). Stage zero represents the vegetative mother cell. Under stress, the mother cell decides to form a spore and starts genomic replication (Stage 1). The vegetative cell initiates asymmetric separation to form two different size compartments, the larger mother cell retains the original genome while the smaller compartment has the replicated genome (stage 2). Similar to phagocytosis, the mother cell engulfs the small compartment in order to form a second layer around the small compartment (stage 3). The small compartment is called the forespore and it is located in the cytoplasm of the mother cell. Both the mother cell and the forespore participate in the development and maturation of the spore (Shen et al., 2019). Dipicolinic acid (DPA) is formed and chelate Ca^{2+} ions to form Ca-DPA in the mother cell and is incorporated into the forespore. The incorporation of Ca-DPA into the spore core displaces the water content, creating an environment that protects the bacterial genome from damage during the sporulation process (Setlow et al., 2006), enhances spore resistance to heat (Francis and Sorg 2016) and inhibits metabolism (Setlow 2014).

During stage four, a cortex layer, which is composed of peptidoglycan, forms between the two forespore layers, and a spore coat and exosporium layer are developed. In the final stage, the mother cell lyses to release the fully matured dormant spore to the environment.

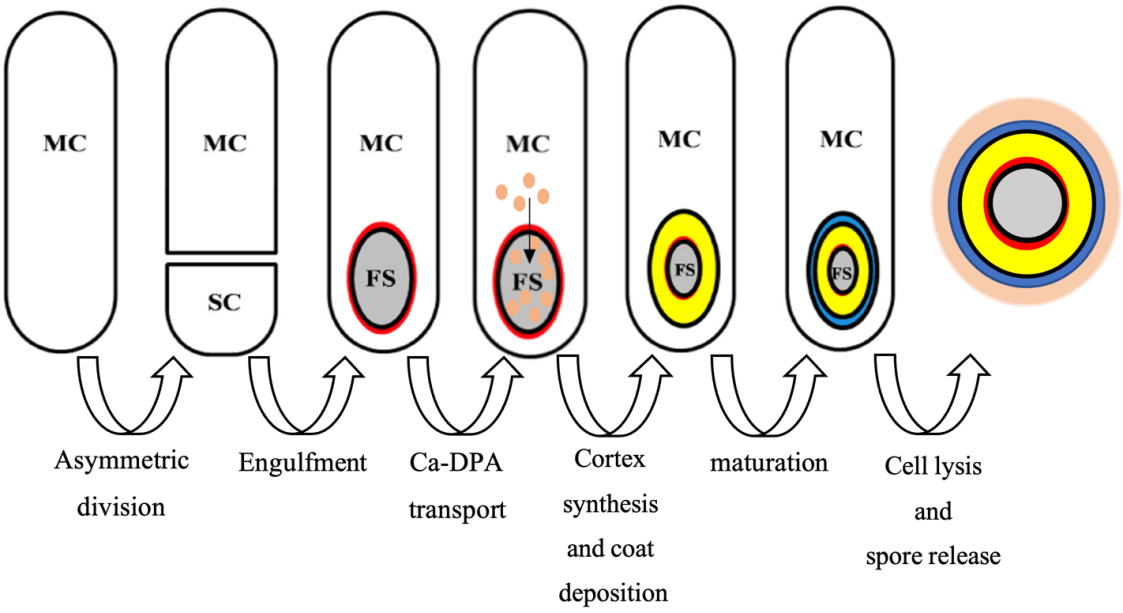


Figure 1.8 Sporulation process of *C. difficile.*
MC refers to mother cell; SC refers to small compartment; FS refers to forespore compartment. Figure modified from (Zhu et al., 2018; Shen et al., 2019).

1.1.7 Germination

Germination is the process by which a dormant spore forms a metabolically active vegetative cell. In order to start the germination process, the spore must first be triggered by small molecules, which indicates that the local environmental conditions are conducive to growth.

These small molecules are called germinants and are usually nutrients with low molecular weight, such as purine nucleosides, sugars, amino acids, and inorganic salts (Driks and Eichenberger, 2015; Paredes-Sabja *et al.,*, 2008; Foerster and Foster, 1966; Setlow, 2013). In 2008, Sorg and Sonenshein found that the bile salt derivatives, taurocholate and glycocholate, stimulated the germination of *C. difficile* spores. The intestinal microbiota metabolizes cholate and produces deoxycholate, a cholate metabolite. Interestingly, deoxycholate was reported to stop the growth of the vegetative form of *C. difficile* and act as a spore germinant (Sorg and Sonenshein 2008). Therefore, the disruption of the microbiota by external factors, such as the administration of broad-spectrum antibiotics, minimizes the role of the microbiota in defeating pathogens and facilitates the initiation of CDI.

The reactivation of biological activities within the spore requires water. However, the high levels of Ca-DPA maintained within the spore cortex prevent the uptake of water. Thus, the interaction of germinants with the germinant receptor, a subtilisin-like *C. difficile pseudoprotease* (CspC), located on the surface of the outer membrane, activates a subtilisin-like *C. difficile* protease (CspB), which activates a spore peptidoglycan hydrolase (SleC). SleC hydrolyzes the cortex layer of the spore, leading to the opening of a Ca-DPA channel (SpoVA), resulting in the release of Ca-DPA. As a consequence, the spore takes up water, which leads to an increase in the volume of the core. Finally, the spore loses its specific features and converts to a metabolically active vegetative cell (Figure 1.9) (Setlow et al., 2017).

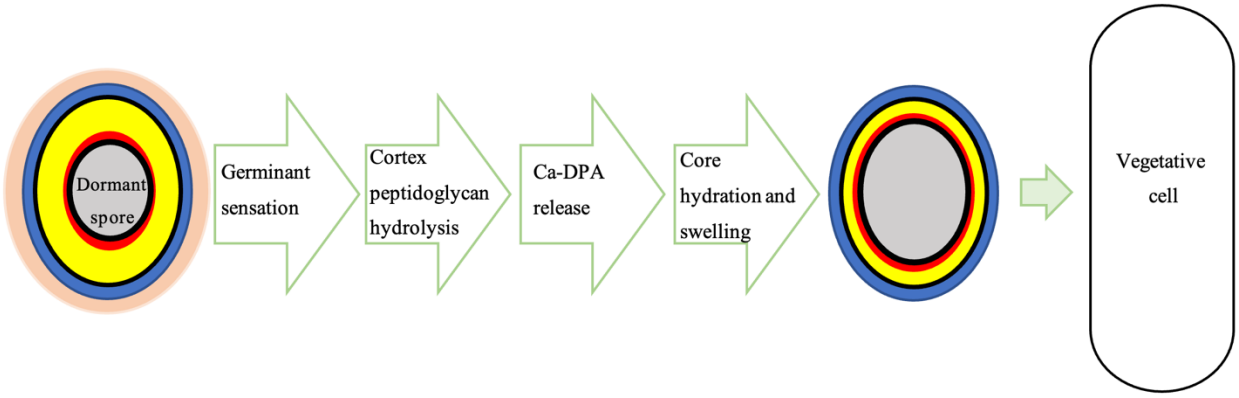


Figure 1.9 *C. difficile* spore germination.
In a permissive environment, spore sense germinants and follow a sequence of actions resulting in formation of a biologically active *C. difficile* cell. Modified from (Setlow et al., 2017; Shen et al., 2019).

1.1.8 Pathogenicity

The bacterium is a major cause of antibiotic associated diarrhea, which occurs following the use of certain classes of antibiotics, such as clindamycin. In 1974, a study which included 200 patients reported that approximately 10% of patients treated with clindamycin developed a condition known as pseudomembranous colitis (PMC) (Tedesco et al., 1974). Subsequently, Larson and colleagues identified the presence of bacterial toxins in the feces of patients with PMC (Larson et al., 1977). Consequently, Bartlett et al., recovered a toxin producing organism from patients with PMC following antibiotic treatment and suggested that a *Clostridia* species was the causative agent of PMC, which was later identified as *C. difficile* (Bartlett et al., 1978).

Through the last two decades, *C. difficile* has become a major worldwide nosocomial pathogen (Paredes-Sabja et al., 2014; Lessa et al., 2015; Feuerstadt et al., 2023). In the USA, not only is CDI the leading cause of healthcare associated infections by being responsible for 12.1% of all hospital acquired infections (Regunath. 2019), but it is also responsible for approximately 500,000 cases annually, of which almost 30,000 are fatal (Lessa et al., 2015; Feuerstadt et al., 2023). Based on the latest update of CDI cases in Wales, there were 298 deaths in 2016 (Office for National Statistics 2017). The pathogen is also responsible for approximately 25% of all cases of antibiotic associated diarrhea (Paredes-Sabja et al., 2014). Moreover, in some medical facilities, it was reported that the number of cases of CDI surpassed the number of infections caused by methicillin resistant *Staphylococcus aureus* (MRSA) (Miller et al., 2011).

C. difficile is transmitted horizontally from patient to patient via the fecal-oral route (Figure 1.10) (Paredes-Sabja et al., 2014). Symptoms range from mild to severe diarrhea with fever. In some cases, severe conditions, such as toxic megacolon and PMC, can be developed (Paredes-Sabja et al., 2014). The cycle of the disease starts following the ingestion of *C. difficile* spores, which protect the bacteria from stomach acidity and antibiotics. Upon reaching the gut, they lay dormant until conditions are conducive for germination.

Normally, the gut microbiota protects the host from infection using a mechanism called the colonization resistant or barrier effect, which prevents exogenous microorganisms from colonizing the gut (Pérez-Cobas et al., 2015). Additionally, members of the intestinal microbiota degrade *C. difficile* germinants, thus inhibiting germination (Sorg and Sonenshein 2008). Antibiotic administration removes the members of the resident microflora responsible for the degradation of germinants, resulting in an increase in the availability of the bile salts, which trigger germination via the CspC spore receptor (Smits et al., 2016).

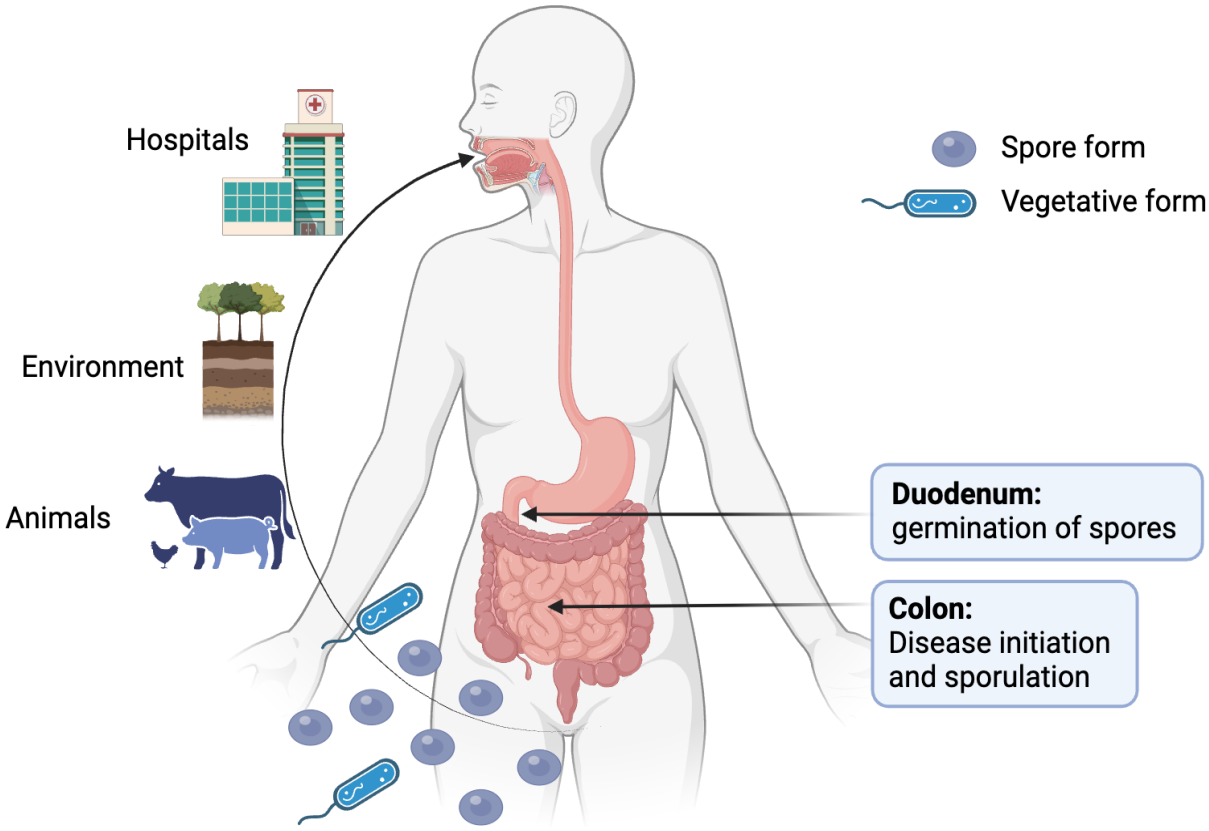


Figure 1.10 *C. difficile* infection cycle in human body.
Ingestion of spores that passes the stomach acidity and germinate in the duodenum is the start point of CDI. Once the vegetative *C. difficile* is attached to the colonic epithelial cells, CDI symptoms appears on the patient. Then, spores leave the human body and become able to infect others. The Figure was reproduced from (Smits et al., 2016) using BioRender software.

Binding of the vegetative form of the pathogen to the host intestinal colon initiates infection. In 2013, Merrigan and colleagues found that surface layer protein A (SlpA) is one of the major contributors to *C. difficile* attachment to the intestinal epithelial cells in addition to flagellin (FliC), flagellar cap protein (FliD), and fibronectin binding proteins mediate attachment to the intestinal epithelial cells (Tasteyre et al., 2001; Hennequin et al., 2003; Barketi-Klai et al., 2011; Merrigan et al., 2013). Additionally, since the exosporium layer is the outermost layer of the spore, it is likely to contribute to the attachment of this structure (Janganan et al., 2016). It has been reported

that two exosporial proteins, CdeC and CdeM, contribute to the spore attachment in the host intestine (Mora-Uribe et al., 2016; Calderón-Romero et al., 2018).

The probability of becoming infected with *C. difficile* is dependent on several risk factors. The highest risk group comprises individuals over 65 years of age who are more than 10 times more likely to become infected (Bignardi 1998). Once infected, the mortality rate can be as high as 25% (NICE 2017). Similarly, long term antibiotic administration increases the risk of CDI by seven to ten-fold. Some antibiotics, such as clindamycin, cephalosporin, penicillin, and fluroquinolone, present a higher risk of CDI than sulphonamides, tetracyclines, macrolides, and aminoglycosides (Marra and Ng 2015). Other high risk groups include individuals with weakened immune systems as a result of disease, such as HIV/ AIDS, or medication, such as immunosuppressants (Collini et al., 2013). Moreover, the use of proton pump inhibitors (PPIs) has been described as a risk factor (Dial et al., 2004). Cancer patients, with or without antibiotic treatment, also have a high risk of developing CDI (Raza et al., 2009). Other risk factors, such as spending time on an intensive care unit (ICU), long term hospitalization, and organ transplantation, have also been documented (Barbut and Petit 2001; Marra and Ng 2015).

1.1.9 Virulence factors

While there are strains of *C. difficile* that never cause CDI, the majority of disease causing isolates owe their pathogenicity to the production of potent toxins located on a mobile genetic element known as the pathogenicity locus (PaLoc) (Tonna and Welsby 2005). The PaLoc is a DNA sequence (19.6 kb) that encodes for two major secreted toxins (Figure 1.11).

Figure 1.11 PaLoc gene of toxigenic *C. difficile* strain CD630.
Figure was reproduced from (Carter et al., 2014) using BioRender software.

The first toxin is called cytotoxin TcdB, which plays an important role in inducing cell apoptosis, while the second toxin is an enterotoxin called TcdA, which damages the actin of the intoxicated cell. Actin damage leads to increased secretion of IL-8 as a result of macrophages and monocyte stimulation, leading to neutrophil infiltration and damage to intestinal epithelial cells (Tonna and Welsby 2005; Fatima and Aziz 2019).

Even though TcdB targets three receptors on the host cell: frizzled receptor (FZD), chondroitin sulfate proteoglycan 4 (CSPG4), and poliovirus receptor-like 3 (PVRL3), the most important host cell receptor in terms of cytotoxicity is CSPG4. Bezlotoxumab, a TcdB specific antibody, interrupts the binding of TcdB to CSPG4 (Gupta et al., 2017). The TcdA host cell receptor binding site is still unknown (Orrell et al., 2017).

The attachment of toxin A and toxin B to their receptors on the host cell triggers the engulfment of toxins through the endocytic process. Exposure of the toxin to low pH in early endosomes results in unfolding of hydrophobic regions, which inserts the protein into the endosomal membrane, forming a hole. The interaction of inositol hexakisphosphate (IP_6) with the toxins facilitates the release of the glycosyltransferase domain (GTD), the enzymically active region into the cytoplasm, inducing glycosylation of Rho-GTPase, which promotes cellular changes and finally leads to cell apoptosis (Figure 1.12) (Orrell et al., 2017; Icho and Melnyk 2018).

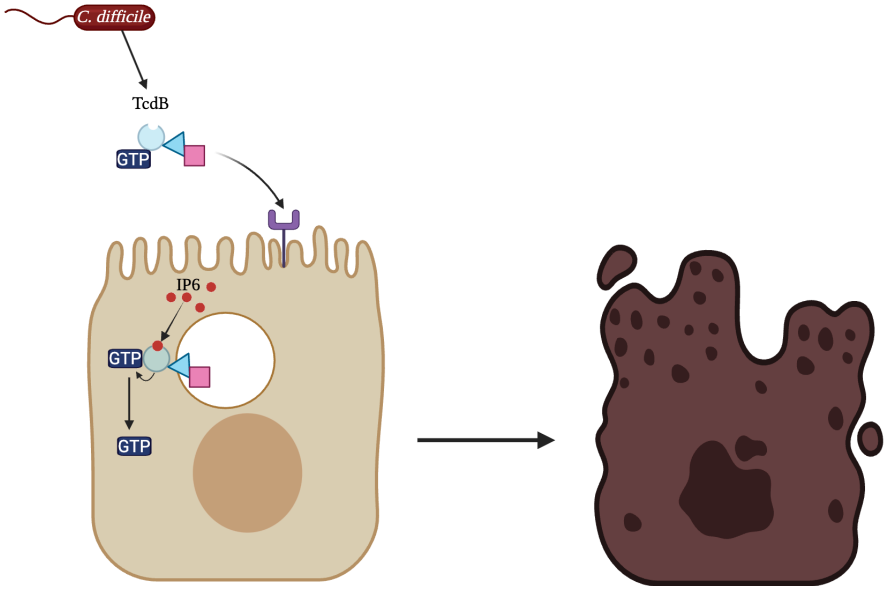


Figure 1.12 Toxication mechanism of TcdB

Following the secretion of *C. difficile* toxins, TcdB binds to the CSPG4 receptor on intestinal epithelial cells which mediate internalization of TcdB through endocytosis. Exposing TcdB to low pH unfold the hydrophobic region which leads to insertion of TcdB into endosomal membrane. The attachment of IP6 to the exposed area of TcdB leads to the release of GTD. Glycosylation of Rho-GTPase by GTD results in cellular changes and cell death. Figure was reproduced from (Icho and Melnyk 2018) using BioRender software.

A third gene called TcdC, which is also present in the PaLoc, is thought to regulate toxin gene expression. A deletion in the TcdC (from 232 amino acids to 61 amino acids) gene has been reported to increase toxin production by 20 fold compared to a control strain (Spigaglia and Mastrantonio 2002; Fatima and Aziz 2019). A deletion in this gene has been observed in one particular strain known as NAP1/B1/027. This strain is hypervirulent and has been associated with several outbreaks and increased mortality and morbidity compared to other *C. difficile* strains.

In contrast to TcdA and TcdB, a third toxin not located within the PaLoc is called binary toxin (CDT) and is composed of CDTa and CDTb. CDTb is responsible for the attachment of CDT to the lipolysis stimulated lipoprotein receptor (LSR) on the host epithelial cell and the subsequent delivery of CDTa to the cytoplasm of the epithelial cell. Then, CDTa, which is an ADP-ribosyltransferase, targets the cellular monomer actin, preventing the formation of the polymer

actin (F-actin). As a result, epithelial cells form protrusions, which enhance the bacterial cell attachment (Gerding et al., 2013).

C. difficile toxins vary in their potency. All *C. difficile* toxins (TcdA, TcdB, and CDT) belong to one family, large clostridial glucosylated toxins. The major toxins, TcdA and TcdB, are secreted by the majority of *C. difficile* strains, while CDT is secreted by a relatively few strains. TcdB is the most potent toxin and is at least 100 fold more active than toxin A (Di Bella et al., 2016). Even though CDT has not been studied extensively, it can inhibit intestinal immunity (eosinophils) by interacting with the toll like receptor 2 (TLR2) (Cowardin et al., 2016). Its overall contribution to toxicity is unclear, as strains lacking CDT are still capable of causing infections.

1.1.10 Current treatment strategies

The antibiotics metronidazole, vancomycin, fidaxomicin, and tigecycline are the only recommended medications for the treatment of *C. difficile* (Johnson et al., 2021; NICE 2021; van Prehn et al., 2021). Metronidazole eradicates the pathogen through the degradation of bacterial DNA, while vancomycin attaches to D-Ala-D-Ala subunits in the peptidoglycan of the vegetative cell wall, arresting development and subsequently leading to cell death.

Tigecycline is a broad-spectrum bacteriostatic antibiotic that inhibits protein translation by binding reversibly to the 30S ribosomal subunit, preventing the attachment of aminoacyl-tRNA. Fidaxomicin is a bactericidal antibiotic that inhibits the RNA polymerase of *C. difficile.* Due to its price, fidaxomicin is recommended as a second line treatment for CDI patients (Marković et al., 2014; NICE 2021). However, regardless of cost, the last update of Infectious Disease Society of America/Society for Healthcare epidemiology of America (IDSA/SHEA) and the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) guidelines recommended fidaxomicin as first line treatment for the first episodes of the infection, prioritizing the safety and efficacy (Johnson et al., 2021; van Prehn et al., 2021).

The recurrence of infection following successful antibiotic therapy increases morbidity and treatment costs. Up to 40% of CDI patients suffer from a recurrence of the disease within two to

eight weeks of successful treatment, and up to 65% of individuals develop more than two repeat infections (Barbut et al., 2000; Marsh et al., 2012). The reinfection may result from the same strain, germination of surviving spores which were not affected by antibiotics, or from a different strain (Chilton et al., 2018). All guidelines recommended fidaxomicin as the drug of choice for the treatment of recurrent infections, and the IDSA/SHEA and ESCMID guidelines recommend the inclusion of bezlotoxumab as an adjunctive treatment with fidaxomicin (Johnson et al., 2021; NICE 2021; van Prehn et al., 2021).

ESCMID recommends tigecycline in the treatment of severe CDI in combination with standard antibiotic therapy (van Prehn et al., 2021). In 2016, a study demonstrated that tigecycline gave significantly better results for the treatment of severe CDI compared to the standard treatment (a combination of vancomycin and metronidazole) in terms of cure rate and disease symptoms (Gergely Szabo et al., 2016).

Regarding metronidazole, it has been recommended as a third line treatment for non-severe CDI in cases where vancomycin and fidaxomicin have proved ineffective (Johnson et al., 2021; NICE 2021; van Prehn et al., 2021).

A non-antibiotic approach which is currently employed to treat CDI patients is fecal microbiota transplantation. This involves the transfer of stool (as a solution) from a healthy adult donor to the intestine of an infected individual. The goal of this technique is to restore a protective microbiota in the gut of the infected individual (Gupta et al., 2016). This technique provides a high success rate, and it is recommended by IDSA/SHEA and ESCMID for the treatment of patients with recurrent infections (Thabit et al., 2019; Johnson et al., 2021; van Prehn et al., 2021).

Within the last decade, monoclonal antibodies have demonstrated their value as a means of preventing recurrent CDI by targeting *C. difficile* toxins. Actoxumab and bezlotoxumab are human monoclonal antibodies designed to target toxin A and toxin B, respectively. The combination of actoxumab and bezlotoxumab or bezlotoxumab alone has been shown to decrease recurrent CDI compared to the placebo group (Wilcox et al., 2017). In 2017, bezlotoxumab was approved by the

FDA and is recommended as an adjunctive treatment by ESCMID and IDSA/SHEA (Johnson et al., 2021; van Prehn et al., 2021).

1.1.11 Experimental treatment strategies

C. difficile is known to be resistant to several antibiotics. Recent reports demonstrated the isolation of strains of *C. difficile* resistant to metronidazole and vancomycin (Barbut et al., 2000; Gonzales-Luna et al., 2021; Darkoh et al., 2022). The mechanism of resistance is unclear, but it is thought to be related to alteration of vancomycin and metronidazole targets in the amino acids of the cell wall or the metabolic pathway, respectively. An additional resistant mechanism could be biofilm formation, which works as a barrier enabling the bacteria to convert to their spore form, which is not affected by antibiotics. Thus, the development of new therapeutic options is recommended (Peng et al., 2017).

Bacteriophages are currently being explored as a treatment for *C. difficile*. These viruses, which target bacteria, were used as antibacterial agents prior to the discovery of antibiotics and are still used in Eastern Europe and the former Soviet Union (Sulakvelidze et al., 2001). One of the biggest advantages of utilizing bacteriophages as a treatment strategy is their specificity, which reduces the damage caused to bystander bacteria (Nale et al., 2018). Thus, targeting *C. difficile*, in the GIT without affecting other protective members of the gut microbiota would overcome the limitation of broad-spectrum antibiotics. To date, several bacteriophages have been utilized to target and kill *C. difficile* (Meader et al., 2010; Nale et al., 2018). Govind and colleagues examined the activity of the bacteriophage CD119 in Syrian golden hamsters, which were chosen as an experimental model due to the similarity of CDI clinical symptoms to those seen in humans. The study revealed that treatment with this phage extended the time to death by three days (Govind et al., 2011).

To date, the *C. difficile* phages which have been identified belong to one of two double stranded DNA (ds-DNA) bacteriophage families. The first family is called *Myoviridae* which have a long contracted tail connected to an icosahedral head that protects the viral genome (Phothichaisri et al., 2018). Example of this family include φC2, φCD27, φMMP04, φMMP02, φMMP01, and φCDKM9. The second family is called *Siphoviridae* and is characterized by a long noncontractile

tail connected to an isometric head (Cai et al., 2019). Example of members of the *Siphoviridae* family include φCD5763, φCD2955, φCD146, φCD111, and φCD6356 (Figure1.13).

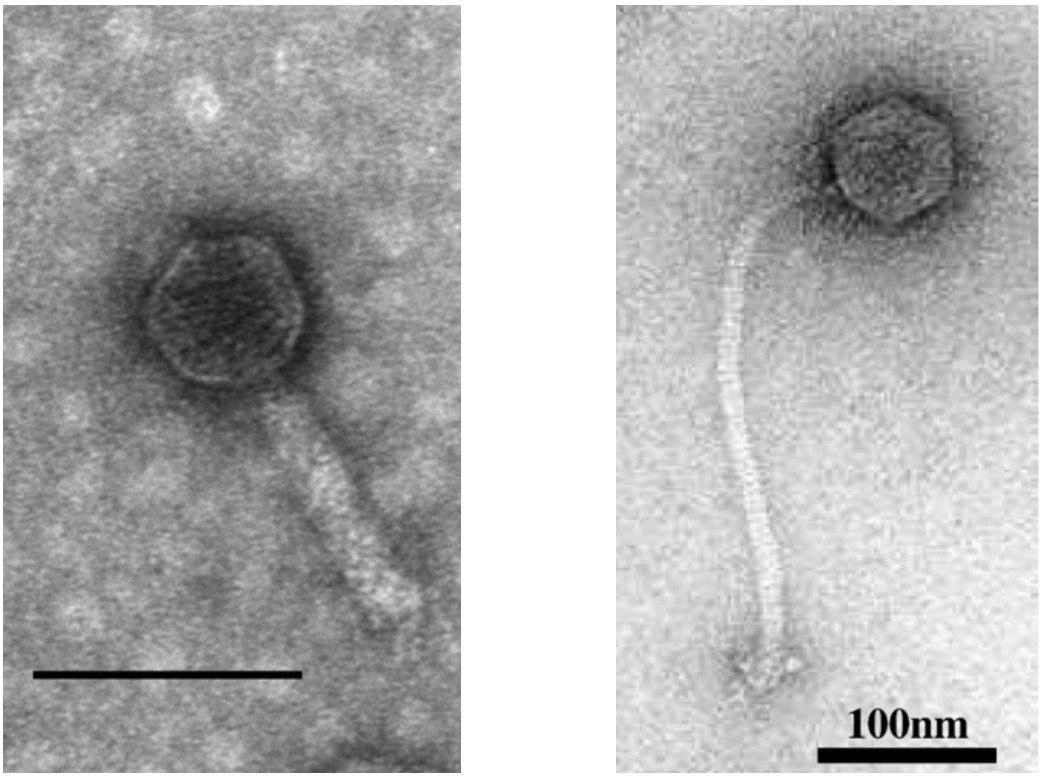


Figure 1.13 Transmission electron microscope images of *C. difficile* bacteriophages.
TEM images are showing an example of two families of bacteriophages including (a) MMP04 which belong to *Myoviridae* family and (b) CD6356 which belong to *Siphoviridae* family. The black bar represents a length of 100 nm. Images were obtained from (Horgan et al., 2010; Meessen-Pinard et al., 2012).

The replication of bacteriophages follows one of two pathways. The lytic cycle hijacks the bacterial cell to reproduce viral particles, leading to cell death and the release of bacteriophages due to the actions of phage encoded lysin, which degrades the bacterial cell wall. Bacteriophages that utilize this pathway for reproduction are called virulent phages. The lysogenic cycle refers to the integration of viral DNA into bacterial genome to form a prophage, which replicates at the same time as the rest of the bacteria. These silent copies of the phage can be reactivated by the

action of radiation, antibiotics, or DNA damage to form active phages, which go on to lyse the bacterial host. Bacteriophages that utilize this mechanism are known as temperate phages (Figure 1.14) (Fortier 2018).

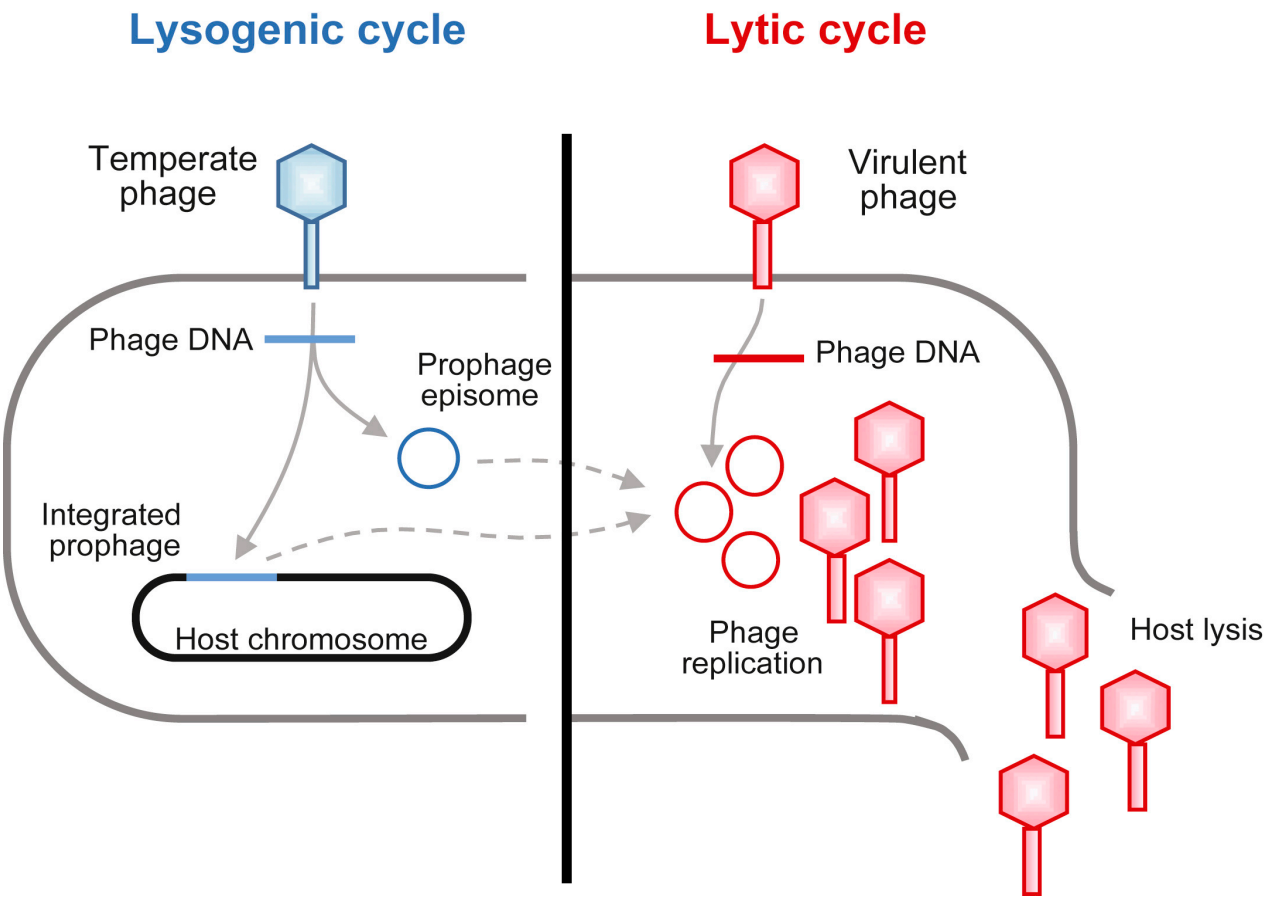


Figure 1.14 Bacteriophage replication pathways.
Figure was modified from (Fortier 2017).

The problems with developing a phage-based therapy for the treatment of *C. difficile* include:

1. *C. difficile* phages are difficult to propagate, so we would need to develop a bespoke production and purification system.
2. There is no single phage that infects all isolates of *C. difficile*.
3. Bacteria can develop resistance to individual phages.
4. To overcome the issue of resistance, we would need to use a cocktail of phages, which target different receptor sites.

1.1.12 Phage lysins

Endolysins are bacteriophage produced enzymes that specifically degrade the bacterial peptidoglycan cell wall, resulting in the release of bacteriophage particles from the infected cell, resulting in the death of the bacteria. Endolysins are divided into at least four main classes according to their target and include muramidase, glucosaminidase, amidase, and endopeptidase, as presented in Figure 1.15 (Ho et al., 2022; Bae et al., 2023; Liu et al., 2023).

Endolysins are often composed of two domains: an N- terminal enzymatically active domain (EAD) and a C- terminal cell wall binding domain (CBD). CBD is present in some but not all lysins, and its role is to target and bind specifically to the bacterial cell wall. The EAD, on the other hand, is present in all lysins and is responsible for cell wall specific peptidoglycan bond degradation and subsequent bacterial cell death.

Their ability to lyse bacteria when employed as an externally applied recombinant protein speaks to their potential use as antibacterial agents. Indeed, the feasibility of this approach has been demonstrated through the use of endolysins to remove *A. streptococci* and *S. agalactiae* from membrane surfaces in animals (Nelson et al., 2001; Cheng et al., 2005).

In order for an endolysin to act, the peptidoglycan layer has to be exposed to the environment. Because *C. difficile*, a Gram positive bacteria, does not have an outer membrane, endolysins are considered a suitable technique to kill the vegetative form of the bacterium. In previous work, the endolysin LysCD6356 (807 bp, 268 amino acid length) demonstrated the highest lytic activity against the following *C. difficile* ribotypes, 001, 027, 106, 014, 002, 005, 078, 020, 010 and 023 (Alyousef 2013).

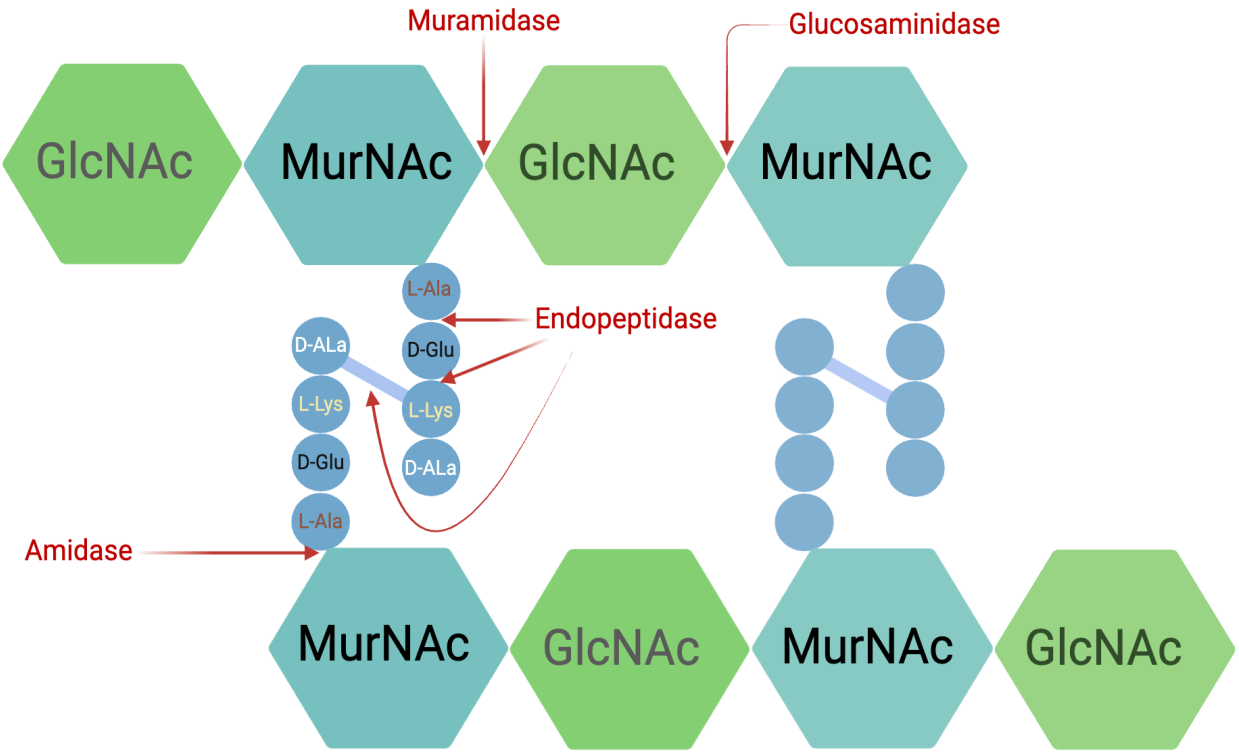


Figure 1.15 The Cell wall peptidoglycan structure of bacteria showing different endolysins cleavage sites.

Four main categories of endolysins (red) and their cleavage sites on the peptidoglycan layer of bacterial cell wall. N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc), L-alanine (L-Ala), D-glutamic acid (D-Glu), L-lysine (L-Lys) and D-alanine (D-Ala). Figure was reproduced from (Ho et al., 2022; Bae et al., 2023; Liu et al., 2023) using BioRender software.

Endolysins have been successfully employed to treat animals infected with *C. difficile.* In 2019, Peng and colleagues determined the feasibility of employing a fusion between LCD, a *C. difficile* phage (phiC2) endolysin, and the human defensin protein HD_5, which has broad spectrum antimicrobial activity, called LDH as a treatment for mice with CDI. The study found that administration of 400 µg of the fusion twice a day for seven days stopped water loss on day 4 of treatment, and the mice survived compared to the control group (Peng et al., 2019).

Another recombinantly expressed endolysin, PlyCD, derived from *C. difficile* CD630, has also been shown to be capable of treating *C. difficile* infected animals. Similar to LHD, PlyCD exhibited lytic activity against specific strains of *C. difficile*. The EAD of PlyCD ($PlyCD_{1-174}$) demonstrated

significantly higher activity against a wider range of *C. difficile* clinical isolates compared to full length PlyCD.

Combining $PlyCD_{1-174}$ with a sub-inhibitory dose of vancomycin (10µg/mL) had a synergistic effect, with treatment efficacy increasing with increasing endolysin dose. *In-vivo* analysis using infected mice showed that $PlyCD_{1-174}$ decreased morbidity and mortality compared to the control. An *ex-vivo* study using an infected mouse colon model found that $PlyCD_{1-174}$ significantly lowered *C. difficile* colonization compared to the control (Wang et al., 2015).

When compared to a phage based approach the use of lysins offers the following advantages:

1) They target multiple isolates of *C. difficile* including phage resistant mutants.
2) Relatively simple to produce as made using recombinant protein technology.

Disadvantages:

1) Susceptible to proteolytic degradation in the GIT
2) Unable to replicate, as is the case for phages.

While lysins are active against the vegetative form of *C. difficile,* like phages, they have no effect against the spore form of the pathogen, which is carried in the gut of healthy individuals. Indeed, it is the ability of this spore form to resist the action of antimicrobial agents and is central to its ability to infect individuals following treatment with broad-spectrum antibiotics.

1.1.13 Protein based therapeutics

Therapeutic proteins are an important field of pharmaceutical biotechnology, considering their therapeutic potential against fatal diseases, such as diabetes, heart disease and cancer (Akash et al., 2015). Additionally, recombinant proteins as therapeutics have increased their importance due to their effective use as vaccines (Dimitrov 2012; Akash et al., 2015). Therefore, the approval of therapeutical proteins and peptides by the FDA increased and exceeded 240 products, and the sales of recombinant protein products in the USA only were $108 billion in 2010 (Dimitrov 2012; Craik et al., 2013; Pham 2018; Puetz and Wurm 2019).

Before a protein can be employed as a therapeutic agent, a number of factors need to be addressed, which include safety, immunogenicity, stability, degradation, interaction between protein and excipient, metabolism, and elimination of the product from the body (Akash et al., 2015). Most of the factors are related to the systemic usage of therapeutic proteins. Moreover, therapeutic proteins are usually administered via parenteral routes (intravenous, intramuscular, or subcutaneous) to avoid degradation in the GIT (Akash et al., 2015).

In terms of oral administration of therapeutic proteins, the stability of the product is the biggest concern. Generally, therapeutic proteins have low bioavailability through GIT as a result of the inherent properties of proteins (Philippart et al., 2015). The acidic environment of the stomach tends to degrade, denature, and precipitate therapeutic proteins. The second major factor that affects protein stability is enzymatic degradation. The presence of proteases and peptidases, which help in the digestion of dietary proteins in the GIT, has the potential to degrade therapeutic proteins (Akash et al., 2015).

Oral administration of therapeutic proteins has been an attractive field for researchers and pharmaceutical companies. A number of strategies have been adopted to address this issue, such as enteric coating of therapeutic proteins to overcome the stomach acidity and the use of proteolytic enzyme inhibitors to overcome enzymic degradation (Chen et al., 2022).

1.1.14 Strategies to target the spore form of the pathogen.

Several spore binding domains have been identified in endolysins. Yang and colleagues constructed several truncated versions of the endolysin PlyG to investigate its ability to bind to *B.anthracis* spores and identified a 60 amino acid sequence from residues 106 to 165 (Yang et al., 2012). In 2019, Kong and colleagues used an *in-silico* analysis to identify the spore binding domain within the endolysin of PBC2 (LysPBC2). They found high similarity between the two sequences (57% percentage of identity) and identified a 62 amino acid region (residues 108 to 170) which mediated binding to *B. cereus* spores (Kong et al., 2019). Interestingly, the spore binding domain of PlyG and LysPBC2 were mainly located within the catalytic domain and only mediated binding

to the spore form of the bacterium (Yang et al., 2012; Kong et al., 2019). In contrast, Gomez-Torres and colleagues observed that the catalytic domain of CPT1L, an endolysin of the *Clostridium tyrobutyricum* phage CTP1, failed to attach to the vegetative form of some Clostridium species, such as *C. beijerinckii, C. tyrobutyricum, C. butyricum,* and *C. sporogenes*; however, its CBD was able to attach to the vegetative and spore forms of clostridial species (Gómez-Torres et al., 2018).

Since the spore form is responsible for infection and the recurrence of the disease, the ability to identify a spore binding tool would facilitate the detection of the pathogen. The recurrence of the infection is directly correlated with increasing morbidity and the cost of health care systems (Barbut et al., 2000; Marsh et al., 2012).

An approach which could be used to combat latent spores in the gut is to stimulate germination so they convert to their more susceptible vegetative form, which can be killed by the action of antibiotics, phages, or lysins. Germinants have been identified for *C. difficile*, which could potentially be combined with lysin to treat CDI.

The feasibility of using bactericidal agents following induction of spore germination in the environment has been demonstrated for spore forming bacteria such as *B. anthracis* and *C. difficile* (Celebi et al., 2016a; Nerandzic and Donskey 2016; Nerandzic and Donskey 2017).

The practicality of using recombinant endolysins to target newly germinated spores, following exposure to a germinant, was recently investigated by Furlon and colleagues who exposed *C. difficile* spores to 1% taurocholate and 12 µg/mL of the catalytic domains of the following endolysins: CD27L, CDBJ082, CDHM11, PlyCD, CDS9P1, CD6356, CDC192 and CDW0003.

They observed a reduction in bacterial viability for all lysins with the exception of CDC192 (Furlon et al., 2021).

1.2 Project aim and research objectives:

1.2.1 Project aim

The aim is to advance the development of a new therapeutic strategy that protects vulnerable individuals infected with *C. difficile* (CDI) and prevents them from developing active infections. To achieve this aim, a strategy was developed in which the enzymatic region of a *C. difficile* specific endolysin can be used to target vegetative *C. difficile* in combination with germinants, which triggers the conversion of resistant spores into the susceptible vegetative form of the pathogen.

1.2.2 Research objectives

1. Using an active endolysin previously identified in our lab, a truncated version of the lysin was design, comprising the EAD.
2. LysCD6356 and its EAD were expressed as recombinant proteins using an *E. coli* expression system.
3. The activity of expressed endolysins will be assessed against a collection of UK clinical isolates of *C. difficile.*
4. The ability of the EAD region of LysCD6356 to target *C. difficile* spores in the presence of germinants will be assessed.
5. The ability of LysCD6356 and its EAD to bind to the surface of the vegetative and spore forms of *C. difficile* will be determined.
6. Potential surface binding regions in LysCD6356 and its EAD will be identified using a bioinformatic based approach.
7. We will investigate the ability of LysCD6356 and its EAD to bind to previously identified surface exposed *C. difficile* spore targets.

Chapter 2 Materials and methods

2.1 Microbiological materials and methods

2.1.1 Growth and sporulation media

Brain heart infusion (BHI) broth and agar were purchased from Merck Life Science Limited (Gillingham, UK). *Clostridioides difficile* Moxalactam Norfloxacin (CDMN) agar, Luria Bertani (LB) broth, and LB agar were purchased from Fisher Scientific Limited (Leicestershire, UK). Tryptone soya broth (TSB) and tryptone soya agar (TSA) were purchased from Oxoid Ltd. (Basingstoke, UK).

Each medium wasprepared according to the manufacturer's instructions and sterilized using aCertoClav EL autoclave (CertoClav sterilizer GmbH, Austria), which was set at 121 ºC for 15 minutes. Following autoclaving, CDMN agar plates were prepared by allowing the media to cool to approximately 50 ºC after which 7% (v/v) defibrinated horse blood and *Clostridioides difficile* supplement antibiotics were added (Fisher Scientific Ltd., UK).

Prior to culturing anaerobic bacteria, media were incubated under anaerobic conditions at 37 ºC for 24 hours in an anaerobic workstation (Ruskinn Technology Ltd., UK) supplied with an anaerobic gas mixture comprising 5% H_2, 10% CO_2, and 85% N_2. Incubating under anaerobic environment helps remove any residual dissolved oxygen. Cultures were then incubated at 37 ºC for 48 hours under anaerobic conditions.

2.1.2 Bacterial strains

A collection of 16 clinical isolates of *C. difficile* representing a range of Ribotypes, including the principal types encountered in the UK, was kindly donated by the Anaerobe Reference Laboratory, University Hospital of Wales, Cardiff (Table 2.1). The bacillus strains used in the study were *B. subtilis*ATCC 27370 and *B. cereus* ATCC 6464, which were obtained from the American Type Culture Collection (ATCC) (Table 2.2).

Strains were stored in protect beads in Mirobank™ vials (Scientific Laboratory Supplies Ltd., UK) at -80 ºC as described below (section 2.1.5).

Table 2.1 *C. difficile* isolates and their clinical relevance.

Strains	Ribotypes	Clinical relevance	Reference
NCTC 12727	001	- Increased resistance to vancomycin. - 6th most common ribotype in Wales, UK.	(Public Health Wales 2013; Public Health England 2016a)
R31760	002	- The most common ribotype in England and the 7th in Wales, UK. - 3rd most common ribotype in Colombia.	(Public Health Wales 2013; Salazar et al., 2017; Herbert et al., 2019)
R31762	005	- 3rd most common ribotype in England and the 4th in Wales, UK.	(Public Health Wales 2013; Herbert et al., 2019)
DS1684	010	- 5th most prevalent ribotype in Colombia.	(Salazar et al., 2017)
CD630	012	- The first strain, that came from an outbreak in Zurich, Switzerland, to be genome sequenced.	(Sebaihia et al., 2006)
R31755	014	- 9th most common ribotype in Colombia. - 3rd most common ribotype in Wales and 4th in England, UK.	(Public Health Wales 2013; Salazar et al., 2017; Herbert et al., 2019)
R31774	020	- 5th most prevalent ribotype in England, UK	(Herbert et al., 2019)
DS1665	023	- 5th most common ribotype in Wales and 10th in England, UK.	(Public Health Wales 2013; Herbert et al., 2019)
DS1813 R20291	027	- Hypervirulent. - R20291 isolated from an outbreak in Stoke Mandeville Hospital, UK in 2006. - The most common ribotype in Texas and responsible for several outbreaks across the USA. - The most common ribotype in Wales, UK.	(Clifford McDonald et al., 2005; Public Health Wales 2013; Gonzales-Luna et al., 2020)
R30776	045	- Binary toxin positive (CDT^{+}), TcdA^{+} and TcdB^{+}.	(Gerding et al., 2013)
R31263	046	- Ribotype reported to be responsible for the Sweden outbreak in 2011.	(Magnusson et al., 2022)
R25961	047	- Binary toxin negative (CDT^{-}), TcdA^{-} and TcdB^{+}	(Stare et al., 2007; Crivaro et al., 2023)
R31312	056	- Binary toxin negative.	(Bauer et al., 2011)

		- Reported to contribute in complicated cases and may lead to death of old patients (75 years old and more).	
R31777	078	- 6th most common ribotype in England and 8th in Wales, UK.	(Public Health Wales 2013; Herbert et al., 2019)
DS1787	106	- The most common ribotype in USA. - 2nd most common ribotype in Colombia. - 9th most common ribotype in Wales, UK.	(Public Health Wales 2013; Salazar et al., 2017; Kociolek et al., 2018)
R30967	110	- Binary toxin negative (CDT^-), $TcdA^-$ and $TcdB^+$	(Crivaro et al., 2023)

Table 2.2 Bacillus isolates used and their importance in this study.

Species	Strain	Importance	Reference
B. cereus	ATCC 6464	- Gram positive, form spores surrounded by an exosporium layer.	(Kailas et al., 2011)
B. subtilis	ATCC 27370	- Gram positive, form spores which are **not** surrounded by an exosporium layer.	(Mckenney et al., 2013)

2.1.3 Culture and growth of Anaerobic bacteria

C. difficile isolates were recovered from -80 ºC freezer stocks and added to pre-reduced BHI broth. Cultures were incubated in an anaerobic atmosphere comprised of 5% H2, 10% CO_2, and 85% N_2 at 37 ºC for 48 hours in a Bugbox Plus anaerobic workstation (Ruskinn Technology Ltd., UK). The integrity of the anaerobic atmosphere was monitored daily using an anaerobic indicator (Oxoid, UK).

2.1.3.1 Culture in broth

Freshly prepared and autoclaved BHI broth was used to culture *C. difficile* in a broth. Broth was incubated for at least 24 hours at 37 °C under anaerobic conditions using an anaerobic workstation (Ruskinn Technology Ltd., UK). Freshly purified *C. difficile* cells (-80 °C) were then cultured in reduced BHI broth under anaerobic conditions at 37 °C for 48 hours.

2.1.3.2 Culture on selective media (CDMN agar)

Clostridium difficile Moxalactam Norfloxacin (CDMN) agar (Fisher Scientific, Leicestershire, UK), selective media, supplemented with 7 % defibrinated horse blood (Fisher Scientific, UK), as instructed by the manufacturer, were reduced prior to use by incubation of agar plates for 24 hours at 37 ºC under anaerobic conditions. Following inoculation with the test culture, plates were incubated for 48 hours at 37 ºC under anaerobic conditions, in an anaerobic workstation (Ruskinn Technology Ltd., UK). Following incubation, plates were examined for the presence of colonies with characteristic morphology. A reference strain, *C. difficile* NCTC 12727 was included to confirm that the culture medium and incubation environment supported the growth of *C. difficile*. An uncultured plate was incubated for 7 days as a negative control.

2.1.4 Culture and growth of aerobic bacteria

E. coli strains with plasmid transformants, including BL21(DE3), BL21(DE3) pLysS, and SHuffle T7, were recovered from -80 ºC freezer stocks and cultured in LB broth or agar plates, as appropriate, supplemented with ampicillin (100 µg/mL). Broth cultures were incubated overnight in a shaker incubator (300 rpm) under aerobic environment at 37 ºC. Agar plates were incubated in an aerobic incubator at 37 ºC.

Bacillus species, *B. cereus* and *B. subtilis,* were recovered from -80 ºC freezer stocks, added to 40 mL of TSB, and incubated aerobically in a shaker incubator (120 rpm) at 37 ºC for 24 hours.

2.1.5 Production of vegetative cultures for endolysin binding studies

Vegetative *C. difficile* R20291 and DS1787 were grown anaerobically in BHI at 37 ºC overnight. On the other hand, *B. cereus* and *B. subtilis* cells were grown aerobically in tryptic soy broth and incubated overnight at 37 ºC. 10 µl of each culture was re-cultured into 45 mL broth media (pre-reduced BHI broth or tryptic soy broth, respectively) at 37 ºC to mid-logarithmic phase (OD_{600} 0.4 - 0.6). Isolates were then centrifuged (5000 X g for 5 minutes) and the pellets were resuspended in 15 mL PBS (pH 7.4). Cells then become ready for binding analysis.

2.1.6 Bacterial storage

For the long term storage of bacterial isolates, two approaches were employed.

Cryoprotect beads: A disposable sterile plastic inoculation loop (10 µl) (Fisher Scientific Ltd, UK) was used to transfer a single colony of pure culture into 10 mL of pre-reduced BHIS (brain heart infusion medium with yeast extract) broth containing 0.03 % L-cysteine and incubated overnight at 37 ºC under anaerobic conditions. The next day, a sterile swab (Thermo Scientific, UK) was used to transfer and mix the bacterial culture with protect beads in a Mirobank™ vial (Scientific Laboratory Supplies Ltd., UK) for 2 minutes under sterile conditions, after which excess solution was discarded and the vial was stored at -80 ºC.

Glycerol storage: An 850 µl sample of a 48-hour bacterial broth culture was mixed with 150 µl glycerol and then transferred to 1.5 mL cryogenic storage vial (Fisher Scientific Limited, UK). The mixture was vortexed and stored at -80 ºC.

2.1.7 *C. difficile* Spore Preparation

C. difficile spores were prepared as previously described (Sorg and Sonenshein 2008a; Sorg and Sonenshein 2010a). Briefly, the bacterium was cultured on pre-reduced BHI agar and incubated under anaerobic conditions for four days. Using a disposable sterile plastic inoculation loop (5 µl), cells were collected and suspended in 1 mL sterile cold water after which suspensions were incubated overnight at 4 ºC. The next day, the suspension was centrifuged at 5000 g for 5 minutes, and the pellet was resuspended in 1 mL of ice-cold sterile distilled water (SDW). This was repeated five times. The suspension was then overlaid on top of 10 mL of 50% sterile sucrose solution and centrifuged at 3,200g for 20 minutes using a swinging rotator (Heraeus Megafuge 8R centrifuge, Thermo Scientific Ltd., UK). The sucrose solution was gently discarded, and the pellet was washed five times with ice cold SDW. The final pellet was resuspended in SDW and stored at 4 ºC until needed.

2.1.8 Production of spores for endolysin binding studies

Spores of the following bacteria were produced and purified as described in Section 2.1.7). *C. difficile* spores of R20291 and DS1787, *B. subtilis* ATCC 27370 and *B. cereus* ATCC 6464 were assessed. The *Bacillus* spores were cultured on tryptic soy agar and incubated aerobically at 37° C for at least four days.

2.1.9 Methods for identifying the purity of *C. difficile* samples.

2.1.9.1 Colony Characteristics

To confirm the purity of individual *C. difficile* cultures, samples were cultured on a sheep blood agar plate containing 5% defibrinated sheep blood (Fisher Scientific Ltd., UK) and incubated aerobically and anaerobically at 37 ºC for 48 hours. There should be no growth following aerobic culture because *C. difficile* is an obligate anaerobic bacterium. Following anaerobic culture, *C. difficile* forms distinctive colonies that are grayish white in colour, irregular in shape, and range from 4 to 6 mm in diameter. Additionally, colonies produce a characteristic horse barn odor, and emit a yellowish green fluorescence under long wave UV light (Curry 2010; Public Health England 2016b).

2.1.9.2 Gram stain

C. difficile is a Gram positive rod shaped bacteria. To confirm the expected morphology and staining characteristics, the follow staining method was employed.

Using a disposable sterile 10µl plastic inoculation loop (Fisher Scientific Ltd., UK), a 10 µl aliquot of culture was spread across the surface of a microscopic slide (Rapid Electronics Ltd., UK) and left to air dry. The bacterial smear was heat fixed by passing the slide through the flame several times. Crystal violet staining reagent was applied to the smear for 60 seconds, after which it was gently removed using running water. The smear was then flooded with Gram's iodine, the mordant, and left for 60 seconds before being removed by running water. The slide was slanted, and a few

drops of 95% ethanol were run over the smear to remove any unfixed stain, after which it was washed as previously described. Next, the smear was flooded with safranin (the counterstain) and left for 60 seconds before a final wash with water. The slide was left to dry before observation under oil using the 100X oil immersion lens of a brightfield microscope, Leica DM750 (Leica Microsystems, UK).

2.1.9.3 Spore strain

Using a disposable sterile plastic inoculation loop (10 µl) (Fisher Scientific Ltd., UK), an aliquot of culture was spread across the surface of a microscopic slide (Rapid Electronics Ltd., UK) and left to air dry. The bacterial smear was heat fixed by repeated passage through a Bunsen flame. The slide was flooded with malachite green (Sigma-Aldrich, UK) solution and heated using a Bunsen burner for 5 minutes. Additional malachite green was added to prevent the slide from drying out. Following this staining step, the slide was gently washed using running water. The slide was then flooded with safranin and left for 60 seconds after which it was washed as described earlier. The slide was left to dry before observation of the staining results under a 100x oil immersion objective lens of a brightfield microscope, Leica DM750 (Leica Microsystems, UK).

2.2 Genetics and protein analysis

2.2.1 Codon optimization.

The 20 amino acids that are used to make proteins are encoded by 61 codons (three adjacent nucleotides), meaning that an amino acid can be represented by more than one codon sequence. Since bacterial species differ in the codons they use for each amino acid, this can cause problems when attempting to express a particular gene sequence in a new host. To optimize protein expression from *E. coli*, gene sequences derived from *C. difficile* were modified to reflect the preferences of the host strain without changing the amino acid sequence (Fu et al., 2020).

Molecular biological companies use proprietary algorithms to codon optimize gene sequences. The recombinant protein expressed in this study, the endolysins LysCD6356 and its EAD and *C.*

difficile spore surface proteins CdeC and CdeM, were codon optimized using GenSmart™ codon optimization service (GenScript, UK). The codon optimized sequences utilized in this study were included in appendix 2. Codon optimized gene sequences were then inserted into a pET-19b expression vector as described below (section 2.2.2).

2.2.2 Insertion of the codon optimized gene sequences into the expression vector.

Restriction sites NdeI (CA/TATG) and BamHI (G/GATCC) that were identified in the multiple cloning site of the pET-19b expression vector were incorporated into the 5' and 3' regions of the codon optimized gene sequences during synthesis. To confirm that incorporation of these sites did not adversely affect the reading frame of the gene, the modified sequences were assessed using the bioinformatic tool ExPASy (https://web.expasy.org/translate/).

2.2.3 PCR primer design

PCR primers were designed as described in Appendix 3.

2.2.4 Polymerase Chain Reaction.

PCR was performed using the HotStarTaq DNA polymerase kit (Qiagen, UK). Standardized concentrations of DNA probe to achieve a final concentration of 0.2 µM in the PCR reaction were prepared. The reaction mix was prepared based on the manufacturer's instructions. Template DNA (5 ng/µl) and reaction mix (table 2.4) were added to BRAND® PCR tubes (Merck, UK) and completed to a final volume of 25 µl with molecular biology grade water (nuclease-free) (Fisher Scientific, UK). Reagents and DNA samples were mixed via aspiration of the content up and down using a micropipette and kept on ice throughout the preparations.

Table 2.3 PCR reagents and DNA concentrations used in the present project.

	Concentrations
PCR Buffer (including $MgCl_2$)	1 X
Q-solution	1 X
dNTP mix	200 µM of each
FWD primer	0.2 µM
RVS primer	0.2 µM
HotStar Taq DNA polymerase	2.5 unit/ reaction
Neuclease free water	Variable
DNA samples	5 ng /25µl

The reaction tubes were then transferred into a Bio-Rad T100^{TM} Thermal Cycle (Bio-Rad, UK), and the PCR reaction was amplified using the following program (ECDC, 2018):

a) Activation of Taq polymerase at 95 ºC for 15 minutes
b) 30 cycles, comprising 94 ºC for 1 minute, annealing at 60 ºC for 1 minute, and elongation at 72 ºC for 1 minute.
c) The final extension is set at 72 ºC for 30 minutes.

2.2.5 DNA extraction.

To confirm successful cloning and transformation, DNA was extracted from the transformed cells using the following methods:

2.2.5.1 Boiling method

This method of DNA extraction was described by Ahmed and Dablool in 2017. Transformed *E. coli* cells were subcultured in 10 mL of LB broth and incubated overnight at 37 ºC using a shaking incubator. The next day, the culture was centrifuged at 4,500 rpm for 5 min at 4 °C. The supernatant was discarded, and the pellet was resuspended in 50 µl of nuclease free water. Suspended cells

were exposed to 100 ºC (boiling water) for five minutes. Following heating, samples were centrifuged at 3000 g for 10 minutes, and the supernatant was transferred to a sterile new 2 mL Eppendorf tube. 35 mL of cold absolute ethanol (96-100%) (Fisher Scientific, UK) was added to the tube, which was centrifuged at 3000 g for 20 minutes using Heraeus biofuge primo R centrifuge (Thermo Scientific, UK). The DNA precipitate was retained, and the aqueous solution was removed. The washing process was repeated with a lower ethanol concentration (70%), and centrifugation of the suspension at 3000 g for 20 minutes was performed to remove most of the ethanol. DNA pellet was left to dry before being resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). DNA samples were stored at -20 °C.

2.2.5.2 Qiagen kit for DNA extraction

DNA extraction of BL21(DE3)pLysS, BL21(DE3), and SHuffle T7 was performed to confirm successful cloning and introduction of the expression plasmid into the host bacteria using a Qiagen kit (DNeasy Blood and tissue Kit, Germany).

Based on the manufacturer's instructions, bacterial cells were sub-cultured in Luria-Bertani (LB) broth supplemented with 100 µg/mL ampicillin and incubated at 37 ºC for 16 to 18 hours using a shaking incubator. The next day, cells were centrifuged using a Heraeus biofuge primo R centrifuge (Thermo Scientific, UK) for 10 minutes at 5000 x g (7500 rpm), and the supernatant was discarded. Pellets were resuspended in 180 µl of enzymatic lysis buffer and transferred into a 2 mL microcentrifuge tube (Fisher Scientific Ltd., UK). These tubes were incubated for a minimum of 30 minutes at 37 ºC using a TECHNE Dri-Block heaters DB-3D (Keison, UK). Following incubation, 200 µl of buffer AL, a chaotropic agent, and 25 µl of proteinase K were added to each tube and vortexed mixed. The tube was incubated at 56 ºC for a further 30 minutes. Following incubation, 200 µl of absolute ethanol (96- 100%) (Fisher Scientific, UK) was added to each tube and vortexed to achieve a semi-homogenous solution. All components of the tube were transferred onto a DNease mini spin column placed in a 2 mL collection tube. These tubes were centrifuged at 6000 x g (8000 rpm) for one minute, and the filtrate was discarded. 500 µl of washing buffer (AW1), were added to column and centrifuged at 6000 x g (8000 rpm) for 1 minutes with the filtrate being discarded. 500 µl of washing buffer (AW2) were added to the column and centrifuged at 20,000 x g (14000 rpm) for three minutes. Again, the filtrate was discarded, and the

centrifuge step was repeated for a further one minute in order to remove any traces of ethanol that may interfere with subsequent steps.

Collection tube was replaced by a new 2 mL microcentrifuge tube (Fisher Scientific Ltd., UK). 200 µl of elution buffer (AE) was added to the DNease membrane and incubated for one minute at room temperature. The tube was then centrifuged at 6000 x g (8000 rpm) for one minute in order to elute DNA. The elution step was repeated to maximize DNA recovery. DNA samples were stored at -20 ºC.

2.2.6 Determining of DNA concentration using a NanoDrop One Spectrophotometer

Following DNA extraction, a NanoDrop One spectrophotometer (ThermoScientific, UK) was utilized to determine DNA concentration and sample purity. After selecting double stranded DNA (dsDNA) on the system, 2 µl of blank (AE buffer) was loaded, after which 2 µl of the test sample was loaded. Each sample was vortex mixed for 5 seconds prior to loading. The device was blanked every three readings of each test sample. DNA concentrations at different wavelength ratios 260/230 nm and 260/280 nm were recorded.

2.2.7 DNA analysis: Agarose gel electrophoresis.

To determine the size of the PCR amplicons, samples were subjected to electrophoresis. A 10 µl sample of PCR products was mixed with 1 µl of DNA loading buffer (Meridian Biosciences, UK). 10 µl of this mixture was then loaded onto a 1% agarose gel containing agarose powder (TopVision Agarose, Thermo Scientific, UK) dissolved in 1X TBE electrophoresis buffer and followed by the addition of 3.5 µl of Safeview to visualize DNA amplicons. 5 µl of DirectLoad™ wide range DNA marker (Merck, UK), which covers sizes from 50 to 10,000 bp, was loaded into a separate well to provide a means to determine the size of the PCR products.

Following loading, the gel was run at 80V, 95mA, and 30W for 1 hour (Amersham Biosciences, UK). Individual DNA bands were visualized using a Bio-Rad Universal Hood II Gel Doc System (Bio-Rad, UK).

2.2.8 Host strains and plasmids.

Escherichia coli BL21 (DE3), *E. coli* BL21(DE3) pLysS (GenScript, UK) and SHuffle T7 strains (New England Biolabs Ltd., UK) with and without pET-19b were cultured in LB media with or without a selective agent (ampicillin 100 µg/mL). Cultures were incubated using a shaker incubator (Thermo-scientific MAX^{QTM} Mini4450) where the rotator was set at 200 rpm and the temperature was set at 37 ºC. Details regarding the characteristics of each strain and plasmid are listed in Table 2.4.

Table 2.4 Bacterial strains and plasmids used to express recombinant proteins.

E. coli strain and plasmid	Relevant characteristics	Source
BL21(DE3)	- T7 expression *E. coli* strain. - Utilized in CdeC and CdeM expression.	GenScript, Netherlands
SHuffle T7	- T7 expression *E. coli* strain that enhances disulfide bond formation in cytoplasm. - Utilized in CdeC and CdeM expression.	NEB Ltd., UK
BL21(DE3) pLysS	- T7 expression *E. coli* strain. - Suitable for expression of toxic proteins. - Utilized in endolysins, LysCD6356 and its EAD, expression.	GenScript, Netherlands
pET-19b	- *E. coli* expression vector.	Merck Life Science, UK
pET-19b CdeC	- pET-19b containing CdeC labeled with 10 Histidine tag at the N terminal between NdeI/BamHI restriction sites.	GenScript, Netherlands
pET-19b CdeM	- pET-19b containing CdeM labeled with 10 Histidine tag at the N terminal between NdeI/BamHI restriction sites.	GenScript, Netherlands
pET-19b LysCD6356	- pET-19b containing LysCD6356 labeled with 10 Histidine tag at the N terminal between NdeI/BamHI restriction sites.	GenScript, Netherlands
pET-19b EAD of LysCD6356	- pET-19b containing EAD of LysCD6356 labeled with 10 Histidine tag at the N terminal between NdeI/BamHI restriction sites.	GenScript, Netherlands

2.2.9 Heat shock Transformation

Plasmids were introduced into the *E. coli* expression host, competent SHuffle T7, using the heat shock method of transformation. Prior to the start of this experiment, the following steps were completed: SOC outgrowth media (New England Biolabs Ltd, UK) (comprising 2% vegetable peptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) was warmed to room temperature. A water bath (Fisher Scientific, Leicestershire, UK) was adjusted to 42 ºC. LB agar plates (with and without ampicillin 100 µg/mL) were warmed at 30 ºC. A shaking incubator (Thermo-scientific MAX^{QTM} Mini4450) with a rotator set to 250 rpm, ice, and an incubation temperature at 30 ºC was prepared.

An aliquot of SHuffle T7 competent cells (New England Biolabs Ltd., UK) previously stored at -80 ºC and an aliquot of plasmids stored at -20 ºC were thawed on crushed ice for 10 minutes. Under sterile conditions, 10 ng of plasmid was introduced to the bottom of the SHuffle T7 tube containing competent cells and mixed using the micropipette's tip. Following the addition of the plasmid, the mixture was incubated on ice for 30 minutes. Next, the mixture was heat shocked at 42 ºC in the water bath for exactly 30 seconds. The tube was moved gently to the ice for a further five minutes of incubation. Then, 950 µl of SOC media was added, the suspension was mixed, and then incubated for 60 minutes at 250 rpm using a shaking incubator. A 200 µl aliquot of transformed cells was then cultured on selective and non-selective agar plates and incubated overnight at 30 ºC. The next day, colonies were counted, and the transformation efficiency (TE) per µg of plasmid was determined using the following formula:

$$\boldsymbol{Ttransformation\ efficiency} = \frac{Number\ of\ colonies\ (1\ ml)}{Total\ plasmid\ DNA\ (\mu g)}$$

The transformation process for the second *E. coli* host, BL21 (DE3), was performed by GenScript for both the CdeC and CdeM pET-19b constructs.

2.2.10 Culturing and storing transformed cells

BL21(DE3)pLysS cells carrying either full length LysCD6356 or the EAD of LysCD6356 were delivered as glycerol stock, which was stored at -80º C. Glycerol stocks were thawed on ice for 15 minutes upon culturing. Using a disposable sterile plastic inoculation loop (Fisher Scientific Ltd, UK), 10 µl of glycerol stock was sub-cultured in 40 mL LB broth supplemented with ampicillin 100 µg/mL and incubated overnight in an orbital shaker incubator at 37 ºC and 200 rpm speed.

The next day, a sterile swab (Thermo Scientific, UK) was used to transfer and mix bacterial culture with protect beads in Mirobank™ vial (Scientific Laboratory Supplies Ltd., UK) for 2 minutes under sterile conditions after which the solution was discarded from each storage bead vial and stored at -80 ºC for long term storage.

2.2.11 Cloning strategy.

The sequences of the full length LysCD6356 endolysin, the Enzymatically Active Domain (EAD) of LysCD6356 and spore proteins (CdeC and CdeM) were codon optimized for expression in *E. coli* (GenScript, Netherlands). Optimized gene sequences were inserted into the pET-19b expression vector by incorporating NdeI and BamHI restriction sites at the 5' and 3' regions of the gene sequence. A bioinformatic resource portal, Expasy (https://web.expasy.org/translate/), was utilized to ensure that the introduced restriction sites kept the nucleotide sequence in frame with the start codon.

Following digestion with these enzymes, the sequences were inserted into the multiple cloning site of the plasmid so that the sequence was downstream of a 10 x Histidine affinity tag and enterokinase cleavage site and was under the control of an inducible T7 promoter (Figure 2.1).

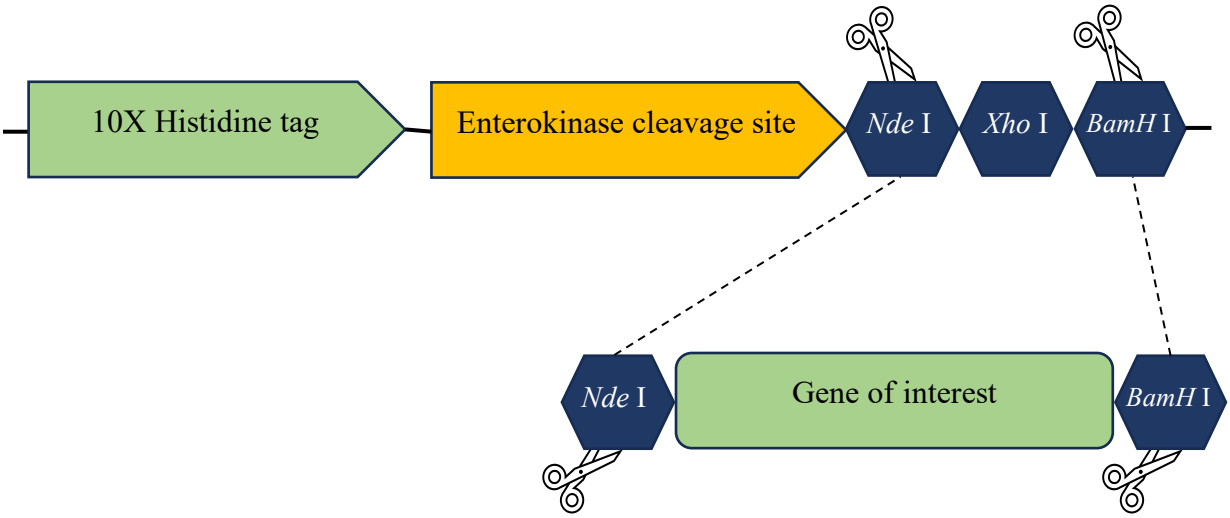


Figure 2.1 Insersion of the gene of interest into pET19b plasmid

Following cloning, the plasmids were transformed into competent *E. coli* cell BL21(DE3) pLysS for endolysins or into BL21(DE3) and SHuffle T7 for the spore proteins (GenScript, Netherlands).

2.2.12 Plasmid preparation

Following cloning of codon optimized versions of CdeC and CdeM into the pET-19b expression vector by GenScript, the plasmid constructs were delivered in vials as a lyophilized powder. In order to reconstitute plasmid DNA, vials were centrifuged at 6000 xg for 1 minute at 4 °C. Then, 20 µl of sterile molecular biology grade water (Nuclease-free) (Fisher Scientific, UK) was added to the vials, which were vortexed for 1 minute. The DNA concentration of the suspension was determined in triplicate using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, UK).

2.2.13 Protein expression and purification.

Four recombinant proteins were expressed and purified. Since there are variations in the method of expression and purification, each technique is described in details.

2.2.13.1 Expression of recombinant endolysins

A 10 µl aliquot of LysCD6356 and EAD of LysCD6356 from freshly prepared culture or a colony from a freshly cultured agar slant were sub-cultured in 10 mL LBG media supplemented with ampicillin (100 µg/mL) and incubated for 14 hours at 37 ºC in a shaker incubator (300 rpm). Following the incubation, 5 mL of overnight culture were sub-cultured into 100 mL of fresh LBG media supplemented with antibiotic (ampicillin 100 µg/mL) and the cultures were incubated at 37 ºC and 300 rpm shaking until cultures reached an absorbance of OD_{600} 0.6. Cells were then induced with 1 mM IPTG and incubated in a shaker incubator (300 rpm) at 27 ºC for four hours. Following incubation, cells were centrifuged at 5000 rpm for 5 min at 4 ºC using a Heraeus biofuge primo R centrifuge (Thermo Scientific, UK). The supernatant was discarded, and the resulting pellets were stored at -20 ºC prior to purification (Alyousef, 2013).

2.2.13.2 Initial purification method of recombinant endolysins

Frozen pellets were thawed for 15 minutes on ice before being resuspended in 30 mL native lysis buffer (50 mM sodium phosphate solution, 300 mM NaCl, 10 mM imidazole, 1 mg/mL Benzonase^{R} nuclease, and 1 mg/mL lysozyme). Cells were then incubated on ice for 30 minutes with mixing three times throughout the incubation period. Pellets were then sonicated five times; each sonication step lasted 40 seconds at an amplitude of 14 on ice using a Soniprep 150 ultrasonic disintegrator (Sanyo, Leicester, UK). The sonication steps were separated by a 30 second incubation on ice. Following cell digestion, the lysate was then centrifuged at 14000 rpm for 30 minutes using an Eppendorf centrifuge 5417 R and the supernatant was retained. The Ni-NTA resin in the fast start column kit was resuspended by inverting the column several times. The storage buffer in the column was removed by cutting the outlet of column followed by the removal of the column's cap. Bacterial lysate was loaded on the column, and the flow through fraction was collected. The column was washed five times with 4 mL washing buffer (50 mM sodium phosphate solution, 300 mM NaCl and 20 mM imidazole), and the washing fraction was collected. Finally, proteins were eluted using 4 mL elution buffer, containing 250mM imidazole, and elution fractions were collected and centrifuged at 8000 g for 15 min at 4 ºC and repeated until the removal of most of the carrier solution. 12 mL of PBS was added to the concentrator tube and centrifuged as

described before. Proteins were collected in 1 mL of the new carrier (PBS), and aliquots were stored at 4 ºC.

2.2.13.3 Modified purification method for the recovery of recombinant endolysins

Frozen pellets were thawed for 15 minutes on ice before being resuspended in 25 mL native lysis buffer (50 mM sodium phosphate solution, 300 mM NaCl, 10 mM imidazole, 1 mg/mL Benzonase^{R} nuclease, and 1 mg/mL lysozyme). Cells were then incubated on ice for 30 minutes with mixing three times throughout the incubation period. Pellets were then sonicated five times; each sonication step lasted 40 seconds at an amplitude of 14 on ice using a MSE Soniprep 150 ultrasonic disintegrator (Sanyo, Leicester, UK). The sonication steps were separated by a 30 second incubation on ice. Following cell digestion, the lysate was then centrifuged at 14000 rpm for 30 minutes using the Eppendorf centrifuge 5417 R and the supernatant was retained.

3 mL of Ni-NTA resin (1.5 mL bed volume) were added to a 50 mL falcon disposable tube after being gently mixed to resuspend the resin. Ni-NTA resin was centrifuged at 5000 rpm for 5 minutes at 4 ºC and then the supernatant was discarded and replaced with 12 mL of lysis buffer, which was mixed with the precipitated resin. The suspension was centrifuged, and the supernatant was discarded and replaced with bacterial lysate, which was mixed with resin and placed on a roller mixer at 4 ºC for 1 hour. Lysates were then poured into previously washed column filters, and the flow through fraction was collected for analysis. Columns were washed five times with 4 mL washing buffer each (50 mM sodium phosphate solution, 300 mM NaCl, and 50 mM imidazole), and the washing fraction was collected for analysis. Finally, the target protein was eluted with a 4 mL elution fraction (50 mM sodium phosphate solution, 300 mM NaCl, and 250 mM imidazole). The elution fraction was collected for analysis and experiments following the removal of imidazole.

2.2.13.4 Expression of recombinant CdeC and CdeM

CdeC and CdeM were expressed as described previously (Brito-Silva et al., 2018).

10 µl of BL21(DE3) + pET-19bCdeC or SHuffle T7 + pET-19bCdeC cells taken from -80 °C were sub-cultured in 10 mL LB media supplemented with antibiotic (ampicillin 100 µg/mL) and incubated for 16 hours at 37 °C in a shaker incubator. Following incubation, 5 mL of fresh LB were added to each tube, the cells were induced with 0.5 mM IPTG, and the culture was incubated at 21° C for a further 3 hours. Cells were then centrifuged at 5000 rpm using Heraeus biofuge primo R centrifuge (Thermo Scientific, UK) for 5 minutes at 4 °C. The supernatant was discarded, and the resulting pellets were stored at -20 ºC prior to purification.

2.2.13.5 Purification of recombinant CdeC and CdeM

Frozen pellets were thawed for 15 minutes on ice before being resuspended in 500 µl soluble lysis buffer, consisting of 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 40 mM imidazole, 0.1% Triton x-100, 2.5 mM β-mercaptoethanol and 30% glycerol. Cells were then sonicated using a MSE soniprep 150 for a total of 6 repeats at 50% output (30 W) for 15 seconds bursts separated by 30 seconds incubation in ice cold water. Following sonication, cells were centrifuged at 14000 rpm for 30 minutes at 4° C using Eppendorf centrifuge 5417 R and the supernatant was retained (Brito-Silva et al., 2018).

3 mL of Ni-NTA resin (1.5 mL bed volume) were added to a 50 mL falcon disposable tube after being gently mixed to resuspend the resin. Ni-NTA resin was centrifuged at 5000 rpm for 5 minutes at 4 ºC and then the supernatant was discarded and replaced with 12 mL of lysis buffer, which was mixed with the precipitated resin. The suspension was centrifuged, and the supernatant was discarded and replaced with bacterial lysate, which was mixed with resin and placed on a roller mixer at 4 ºC for 1 hour. Lysates were then poured into previously washed column filters, and the flow through fraction was collected for analysis. Columns were washed five times with 4 mL washing buffer each (50 mM sodium phosphate solution, 300 mM NaCl, and 50 mM imidazole), and the washing fraction was collected for analysis. Finally, the target protein was eluted with a 4 mL elution fraction (50 mM sodium phosphate solution, 300 mM NaCl, and 250 mM imidazole). The elution fraction was collected for analysis and experiments following the removal of imidazole.

2.2.14 Removal of imidazole from eluted protein preparation

A Pierce™ protein concentrator with a 10,000 molecular weight cutoff was used to concentrate the protein and remove the carrier buffer. Based on the manufacturer's instructions, elution fractions were loaded into the Pierce™ protein concentrator PES 10K MWCO, and the weight of a counterbalance tube was adjusted to match that of the sample tube using a Sartorius CP 6201 scale. The tubes were then placed facing each other in a centrifuge, which was run several times at 8000 g for 15 minutes at 4 ºC until the majority of the carrier solution had been removed. 4 mL of PBS was added to the concentrator tube and centrifuged as described before. The protein fraction was collected in 2 mL of PBS, and aliquots were stored at 4 ºC.

2.2.15 Determination of protein concentration.

The BCA assay was used to determine protein concentration. All reagents were supplied by Thermo Scientific, UK. Based on the manufacturer's instructions, bovine serum albumin (BSA) was diluted to a range of concentrations with PBS and used to construct a standard curve (0-2000 µg/mL). The working reagent was prepared by adding 50 parts of reagent A to one part of reagent B (50:1 ratio (A:B)). 25 µl of each standard concentration, BSA dilutions, and unknown protein concentration were loaded into specific wells of a COSTAR transparent 96-well plate. 200 µl of working reagent were then added to each well, followed by 30 minutes of incubation at 37 ºC using an incubator. The 96-well plate was then left for 5 minutes at room temperature before the absorbance was measured at 562 nm using a plate reader (TECAN, Austria).

2.2.16 Estimating the percentage of other bands.

Using the method described on the Bio-Rad website, the percentage of monomeric form of endolysins was estimated. Briefly, following the opening of the image in Bio-Rad software, the lane and band tool was selected. Each lane in the image was defined automatically, after which the lane outlines were adjusted to cover the whole bands. Bands were then identified either automatically or manually. The lane profile was used to adjust band boundaries and to subtract the background of each lane in order to increase the accuracy of identifying protein concentration.

2.2.17 Sodium dodecyl sulfate polyacrylamide gel electrophoresis.

A 20 µg sample of a protein in PBS solution was mixed with 950 µl of Laemmli sample buffer, (containing 65.8 mM Tris-HCl, pH 6.8, 26.3% (w/v) glycerol, 2.1% SDS, and 0.01% bromophenol blue, plus 50 µl of 2-mercaptoethanol) in a 1:1 ratio and heated at 95 ºC for 5 minutes using a Dri-block® DB·2D (Techne, UK). Protein fractions were then ready to be loaded into the SDS-PAGE gel.

A 12% Mini-PROTEAN® TGX™ precast polyacrylamide protein gel (Bio-Rad, UK) was prepared by removing the blocking tape and the comb. The gel was then placed in the electrode assembly module facing the dam and transferred to the tetra cell tank. The electrode assembly was first submerged with 1X tris glycine SDS buffer (TGS), followed by filling the tetra cell tank to 2 gel mark with 1X TGS. Protein fractions were then loaded gently into the individual wells of the polyacrylamide gel. They included a pre-induction fraction, a post-induction fraction, a flow-through fraction, a wash fraction, an elution fraction, and PageRuler™, which is a molecular weight ladder range from 10 to 250 kDa (Plus Prestained Protein Ladder, Fisher Scientific Ltd., UK). Once loaded, the gel was run using a PowerPac™ Basic at 100 V, 95 mA and 30 W for 75 minutes (Bio-Rad, UK).

2.2.18 Coomassie brilliant blue staining

Following electrophoresis, the polyacrylamide gel was removed from the electrophoresis tank and stained with Coomassie brilliant blue staining solution (100 mL methanol, 25 mL acetic acid, 250 mg Coomassie brilliant blue R-250, and 125 mL filtered water) for 30 minutes at room temperature on a Mini Rocker-Shaker. The stain was removed and replaced by Coomassie destaining solution (200 mL methanol, 100 mL Acetic acid, and 700 mL filtered water), after which it was incubated at room temperature for a further one hour on a Mini Rocker-Shaker. The solution was then replaced with fresh Coomassie destaining solution and left overnight on a Mini Rocker-Shaker at room temperature.

The gel was visualized using a white transillumination light in the Bio-Rad Universal Hood II Gel Doc System (Bio-Rad, UK) and analyzed using Image Lab software (ChemiDoc™ XRS+ 6.0).

2.2.19 Membrane transfer and Western blot

Following the SDS-PAGE run, the proteins in the polyacrylamide gel were transferred into a PVDF membrane (Bio-Rad, UK). The PVDF membrane and the gel were covered top and bottom with filter paper (Bio-Rad, UK). Air bubbles were removed using a roller, after which the sandwich was transferred into the Trans-Blot Turbo Transfer system (Bio-Rad, UK). In order to run the system, from the Bio-Rad list of protocols, one mini gel option was selected, after which mixed molecular weight was chosen since it covers a broad range of proteins' molecular weight (from 5 to 150 kD). The settings of the run were performed at 1.3A and 25V for 7 minutes. Since Trans-Blot Turbo Transfer system has two cassettes, identifying the cassette that holding the gel was performed in order to run the protocol.

Following the transfer of proteins to the PVDF membrane, the membrane was transferred to a small container and stained with 0.1% Ponceau S solution for 5 minutes to confirm that the transfer of proteins to the PVDF membrane had occurred. In addition, an image of the membrane was taken to indicate the protein band's location and their size.

The Ponceau S stain was then removed by washing the membrane three times with distilled water. The membrane was then washed with Tris buffered saline (TBS) for 10 minutes on a Mini Rocker Shaker. The membrane was then blocked with 3% BSA in TBS on a mini rocker shaker at room temperature for one hour. After removal of the blocking buffer, 100 µl of THE™ His tag antibody, a murine antibody specific for the histidine tag, was mixed with 10 mL of 3% BSA in TBS (1:2000 dilution). This antibody solution added to the membrane and incubated overnight at room temperature on a roller mixer. The next day, the membrane was washed three times for 5 minutes each time at room temperature with TBS-Tween-20 (TBST). Goat anti-mouse IgG horseradish peroxidase conjugate diluted with 2% reconstituted milk powder with TBST to reach a final

dilution of 1:10,000 to 1:200,000 was added, and the membrane was incubated for one hour at room temperature. The membrane was then washed three times (five minutes per wash) with TBST. Individual bands were detected by the addition of SuperSignalTM west femto substrate solution (1:1) and incubation for 5 minutes at room temperature. The membrane was visualized using the Bio-Rad Universal Hood II Gel Doc System (Bio-Rad, UK) and analyzed using Image lab software (ChemiDoc^{TM} XRS+ 6.0).

2.2.20 Native-PAGE

Following recombinant protein purification and BCA analysis, recombinant protein was mixed with native sample buffer (62.5 mM Tris-HCl, pH 6.8, 40% glycerol, 0.01% bromophenol blue) (BioRad laboratories, UK) at a 1:2 ratio. The tank was filled with Tris/glycine buffer (25 mM Tris, 192 mM glycine, pH 8.3) (Bio-Rad laboratories, UK), and samples were loaded in 12% mini protein TGX precast protein gel (Bio-Rad laboratories, UK). The gel ran at 100 V, 95 mA, and 30 W for 90 minutes using PowerPac^{TM} Basic (Bio-Rad, UK).

2.2.21 Far western blot analysis

Following the transfer of proteins to PVDF membrane, the membrane was blocked with 5% reconstituted milk powder in PBS-tween-20 (PBST) for two hours on a roller mixer at room temperature. The blocking buffer was removed, and the blocking step was repeated as described before. PVDF was rinsed with PBS and then incubated in PBS on a roller mixer for 10 minutes at room temperature. Following the removal of PBS, endolysin-Alexa 488 was then added to the tube containing the PVDF membrane and incubated for two hours at room temperature. PVDF was washed four times with 200 mL PBS for 5 minutes and then transferred to Bio-Rad Universal Hood II Gel Doc System (Bio-Rad, UK) for visualization using Alexa 488 option (Edmondson and Roth 2001). A range of dilutions of endolysin-Alexa 488 were added to the membrane as a positive control.

2.2.22 Estimated size of the recombinant endolysins

The size of the recombinant endolysins and spore proteins was determined *in silico* using the following online programs:

https://www.bioinformatics.org/sms/prot_mw.html and https://web.expasy.org/compute_pi/

Table 2.5 The estimated sizes of recombinant proteins as determined using online prediction software.

	Recombinant protein size (kDa)			
	CdeC	CdeM	LysCD6356	EAD of LysCD6356
Expasy (compute pI/Mw)	47.532	22.023	32.66811	22.11504
Bioinformatics.org	47.54	22.03	32.67	22.12

2.3 Activity assessment

2.3.1 Assessing the activity of recombinantly produced EAD

Lytic activity against vegetative *C. difficile* isolates was determined using a turbidity reduction assay (Mayer et al., 2008a; Alyousef 2013). Isolates were cultured in 45 mL of pre-reduced BHI broth at 37 ºC under anaerobic conditions to mid-logarithmic phase (OD_{600} 0.4 - 0.6). Following centrifugation (5000 X g for 5 minutes), and the pellets were resuspended in phosphate buffer saline (pH 7.4) and adjusted to OD_{600} 0.8 -1.2. A 300 µl aliquot of cell suspension was mixed with recombinant endolysin in PBS in a Costar 96 flat transparent well plate. As a negative control, endolysin was replaced by PBS. Plates were incubated under aerobic conditions at 37 ºC, and changes in OD_{600} over time were determined using a TECAN infinite F200 pro at fixed time points. To normalize results, the initial value (OD_{600} at time 0) was adjusted to an OD reading of 1.0. Experiments were performed in triplicate for each test condition.

2.3.2 The impact of pH on EAD activity

Vegetative *C. difficile* grown to mid-log phase (OD_{600} 0.4 - 0.6) was centrifuged at 5000 xg for 5 minutes and resuspended in PBS at pH levels ranging from 4 to 9. It is worth mentioning that some of the pH levels were outside the controllable ranges of the buffer. 60 µg/mL EAD was added to

300 µl of the various suspensions in a Costar 96 flat transparent well plate. Changes in OD_{600} over time were determined as described above.

2.3.3 The impact of calcium on EAD activity

Vegetative *C. difficile* grown to mid-log phase (OD_{600} 0.4 - 0.6) was centrifuged at 5000 xg for 5 minutes and resuspended in different concentrations of $CaCl_2$ in SDW (0, 12.5, 25, 50, and 100 mM). 60 µg/mL EAD was added to 300 µl of the various suspensions in a Costar 96 flat transparent well plate. Changes in OD_{600} over time were determined as described above.

2.3.4 The contribution of divalent metal cations to EAD activity

EAD (60 µg/mL) was incubated with 250 mM ethylenediaminetetraacetic acid (EDTA) at 37 ºC for 30 minutes. EDTA was then removed using a 10,000 MWCO Pierce™ protein concentrator, and the treated protein was resuspended in molecular water. The lytic activity of the EDTA treated EAD was determined as described above and compared to that of the untreated EAD.

To identify the cations involved in EAD structure, 60 µg/mL of EDTA-treated EAD was purified as described previously, and was added to 300 µL of resuspended vegetative *C. difficile* R20291 in the presence of 1 mM Mg^{2+}, Mn^{2+}, or Zn^{2+} solution. Lytic activity was determined as described above and compared to that of untreated EAD.

2.3.5 Impact of spore germinants and calcium chloride on the lytic activity of the EAD

The impact of the spore germinants in the presence and absence of $CaCl_2$ on the ability of EAD to lyse vegetative cells of *C. difficile* R20291 was determined. Mid log phase cells (OD_{600} 0.4 - 0.6) were centrifuged at 5000xg for 5 minutes and resuspended in 300µl of the following: sterile distilled water (SDW), SDW + 0.1% taurocholate (Tc) + 50 mM Glycine; SDW + 0.1% Tc + 50 mM Glycine + 12.5 mM $CaCl_2$; SDW + 60 µg/mL EAD; SDW + 0.1% Tc + 50 mM Glycine + 60 µg/mL EAD; SDW + 0.1% Tc + 50 mM Glycine + 12.5 mM $CaCl_2$ + 60 µg/mL EAD. Plates were incubated under aerobic conditions at 37 ºC and changes in OD_{600} over time were determined using a TECAN infinite F200 pro plate reader at fixed time points. To normalize results, the initial value (OD_{600} at time 0) was adjusted to an OD reading of 1.0. Experiments were performed in triplicate for each test condition.

2.3.6 *C. difficile* spore germination

Spore germination was assessed in the presence of 0.1% sodium taurocholate and 50 mM glycine, with and without 12.5 mM calcium chloride. Purified *C. difficile* spores were adjusted to start at ~ 1 OD_{600}. Each germination experiment was performed in a Costar 96 flat transparent well plate. Germination was monitored by measuring the change in the OD_{600} reading at 37 ºC using a TECAN infinite F200 pro plate reader. The experiment was performed in triplicate.

2.3.7 Assessing the sensitivity of newly germinated spores to EAD

2.3.7.1 Simultinous exposure to spore germinants and EAD

C. difficile R20291 spores at a starting OD_{600} of ~ 1.0 were induced to germinate by the addition of the following germinants 0.1% Tc and 50 mM glycine with and without 12.5 mM $CaCl_2$ in SDW. Simultaneously, 60 µg/mL of EAD was added to each test group. R20291 spores were re-suspended in SDW without germinants as a negative control.

2.3.7.2 Sequential exposure to spore germinants and EAD

C. difficile spores at a starting OD_{600} of ~ 0.1 were incubated anaerobically with germinants: 0.1% Tc and 50 mM glycine for 150 minutes in SDW or BHI at 37 ºC. Treated suspensions were then centrifuged and resuspended in PBS to an OD_{600} of 1.0. A 300 µl sample of each test group was loaded into a Costar 96 flat transparent well plate and 60 µg/mL EAD was added. Positive and negative controls were included by suspending vegetative *C. difficile* R20291 in PBS alone or in the presence of 60 µg/mL EAD.

2.4 Fluorescent labelling of recombinant endolysin

2.4.1 Conjugation of Alexa 488 to endolysins

The procedure was performed as instructed by the manufacturer (ThermoFisher Scientific, UK). Alexa Fluor™ 488 NHS ester, succinimidyl ester (ThermoFisher Scientific, UK) (Alexa 488) were dissolved in 200 µl of anhydrous dimethyl sulfoxide (Sigma-Aldrich, UK). Freshly expressed and

purified recombinant endolysin was added to Alexa 488 at a 1:6 ratio and incubated in the dark at room temperature for one hour with continuous stirring.

2.4.2 Removal of unconjugated Alexa 488

Zeba^{TM} spin desalting columns, 7000 MWCO, (ThermoFisher Scientific, UK) were used to remove unattached Alexa 488. The bottom of the column was first removed, and the cap was loosened. The column was then placed in a collection tube, a 50 mL centrifuge tube. The column was centrifuged at 1000 xg for two minutes to remove the storage solution, then washed with 5 mL of the carrier solution (PBS) and centrifuged at 1000 xg for 2 minutes. The washing step was repeated three times, and the PBS was discarded. The column was then placed into a new collection tube, and 700 µl of endolysin-Alexa 488 mixture, including unconjugated Alexa 488, was loaded into the centre of the resin. An additional 200 µl PBS, a stacker, was then added to the column. The column was centrifuged at 1000 xg for two minutes, and sample was collected into a 1.5 mL centrifuge tube.

2.4.3 Determining the degree of labeling (DOL)

The approximate DOL was determined using the Beer-Lambert law as shown below.

$$A = \varepsilon C L$$

Where A is absorbance, ε is extinction coefficient, C is concentration, and L is path length.

The absorbance of samples was read using UV visible spectroscopy (Agilent technologies Cary 60 UV-Vis) where protein was read at 280 nm and Alexa 488 was read at 495 nm. After identifying the absorbance, the extinction coefficient was identified for the protein by uploading the protein sequence to the Expasy.com website to be calculated. The extinction coefficient of Alexa 488 was provided by the supplier (71,000 $cm^{-1}M^{-1}$). The absorbance of the fluorophore was then corrected at 280 nm by multiplying the absorbance at OD495 by 0.11, then subtracting this value from the absorbance at OD280. The path length of the cuvette was 1 cm.

2.4.4 Attachment of endolysins to *C. difficile* spore surface

Using the method described by Mayer et. al, 2011, 7.5 µM of the endolysin-Alexa 488 were mixed with spores and incubated for 20 minutes at 37 ºC. Spores were then centrifuged at 5000 xg for

five minutes and resuspended in PBS. The washing step was repeated twice to remove unattached endolysin-Alexa 488 (Mayer et al., 2011). Spores were then observed under a microscope (Leica DMIRB Inverted Leica Modulation Contrast Microscope).

2.4.5 Attachment of endolysins to the surface of vegetative *C. difficile*

Vegetative cells were mixed with 7.5 μM of EAD-Alexa488 and LysCD6356-Alexa488. Mixtures were incubated for 20 minutes at 37 ºC. Following the incubation, cells were centrifuged at 5000 xg for five minutes and resuspended in PBS. The washing step was repeated twice to eliminate any traces of unattached EAD-Alexa488 or LysCD6356-Alexa488. Cells were then observed under a microscope (Leica DMIRB Inverted Leica Modulation Contrast Microscope).

2.5 Imaging anlysis tools

2.5.1 Scanning electron microscope (SEM)

A *C. difficile* suspension was mixed with a fixative solution (2% glutaraldehyde in 0.1M cacodylate buffer, pH 7.4) (Generon, UK) at a 1:9 ratio and incubated for 2 hours at room temperature. Following fixation, the cells were resuspended in 0.1M sodium cacodylate buffer (pH 7.4) and incubated at room temperature for 3 minutes, after which cells were resuspended in increasing concentrations of ethanol (50%, 70%, 80%, 95%, and 100%). The treated cells were recovered by filtration using a 0.2 μm pore polycarbonate filter membrane. Following the removal of the solution, the membrane was cut into 1 cm square and mounted on stainless steel studs. The membrane was sputter coated with 7 nm of an 80/20 Au/Pd mixture and transferred to the electron microscope vacuum chamber. Cells were then visualized using a 5 kV electron beam.

2.5.2 Phase contrast microscopy

Imaging using phase contrast microscopy was performed as previously described (Alabdali et al., 2021). Briefly, 5 μl of bacterial samples were mounted using 80% glycerol solution on a microscopic slide, and a cover slip was carefully placed over the sample. Cells were then observed under Leica DMIRB Inverted Leica Modulation Contrast Microscope. Images were analyzed and a scale bar was applied using ImageJ software.

2.6 Statistical Analysis

All statistical analysis was carried out using GraphPad Prism 9 software. For analyzing differences between two groups, a t-test analysis was used. A one-way ANOVA was used for differences among more than two groups.

2.7 Bioinformatics tools

To investigate the specificity and conserved levels of CdeC and CdeM and to identify possible endolysin binding regions, bioinformatics analysis tools were utilized.

A search of the NCBI database (https://www.ncbi.nlm.nih.gov/) for the genomic sequence of LysCD6356 (NCBI reference sequence NC_015262.1) was performed in 2020. The database was also searched for homologs to the protein sequences of CdeC and CdeM that are identified in the reference strain *C. difficile* CD630 (NCBI reference numbers WP_003437985 and WP_003436550, respectively). Sequences identified by this process were downloaded in FASTA format to facilitate further analysis.

NCBI-BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to search the NCBI database and identify protein or nucleotide sequences similar to the query sequence. An issue encountered following NCBI-BLAST search was the availability of partial sequences. For this reason, only complete sequences that matched the query sequence were used for further analysis.

The Clustal Omega (version: Clustalo (1.2.4)) (https://www.ebi.ac.uk/Tools/msa/clustalo/), a multiple sequence alignment tool, was used to determine the degree of sequence conservation of CdeC and CdeM homologs and to investigate the homology with non *C. difficile* protein sequences.

NCBI-Identified Protein Groups (NCBI-IPG) (https://www.ncbi.nlm.nih.gov/ipg) is an NCBI tool used to search a range of databases including GenBank, SwissProt, and PDB, in order to organize data entries based on homologous protein sequences, where each identified protein group is characterized by a unique protein sequence. Thus, each protein may have several IPGs depending on the level of conservation of that protein. NCBI-IPG was used to retrieve relevant data about each protein sequence in terms of the number of strains and submitted sequences having the exact sequence.

The crystal structures of CdeC, CdeM, and endolysin have not been resolved, and so the amino acid sequences of each protein were analyzed using the I-TASSER software (version 5.1) (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) to generate an *in-silico* prediction of tertiary protein structure. The structures were then aligned, and images were generated using the Molecular Operating Environment (MOE) program.

Chapter 3 Recombinant endolysin protein expression

3.1 Introduction

Endolysins are enzymes that degrade the cell wall peptidoglycan of vegetative bacteria, leading to osmotic stress induced cell lysis (Vasina et al., 2021). They are able to mediate bacterial lysis when expressed internally as part of the life cycle of a lytic bacteriophage or when applied externally as a recombinant protein. The feasibility of employing recombinant lysins to target the vegetative form of *C. difficile* has been demonstrated across a number of *in vitro* studies since the first example, CD27L, was reported in 2008 (Mayer et al., 2008). This lysin, which is derived from the *C. difficile* specific bacteriophage φCD27, is an N-acetylmuramoyl-L-alanine amidase (amidase 3) class enzyme that is able to lyse peptidoglycan across a wide range of pH values. While lysin is active against a range of *C. difficile* ribotypes it has no activity against the commensal microflora present in the gut. This specificity offers a considerable advantage over broad spectrum antibiotics.

Like CD27L, the majority of recombinantly expressed lysins are derived from bacteriophages, with recent examples including LysCD6356, LysC2, LysCD119, and LysCD38 (Alyousef 2013). Of these, LysCD6356 was found to be the most active in our laboratory against a panel of *C. difficile* clinical isolates, which included a diverse range of ribotypes (001, 027, 106, 014, 002, 005, 078, 020, 010, and 023).

Phage lysins that target Gram positive bacteria are comprised of two domains: an enzymatically active catalytic domain (EAD) and a cell wall binding domain (CBD), which confer target specificity (Murray et al., 2021). Studies have shown that the EAD of some endolysins is more efficient at lysing vegetative bacteria than the full length endolysin and have a wider range of activity (Low et al., 2005; Mayer et al., 2011; Wang et al., 2015). Full length lysins such as CD27L form dimers, which enable them to bind to the cell wall of the target bacteria, after which the protein undergoes autocleavage to release the EAD, which allows it to penetrate further into the bacterial cell wall which may act as a sieve (Dunne et al., 2014). Recently, it has been proposed that in addition to mediating binding to the cell wall, CBD also reduces enzyme diffusion and thereby limits cell lysis of the neighbouring bacteria, providing new hosts for the newly released bacteriophage progeny (Phothichaisri et al., 2022).

In an *ex-vivo* study of the infected mouse colon, the EAD of PlyCD significantly reduced *C. difficile* colonization and in a subsequent *in vivo* mouse model, decreased morbidity and mortality (Wang et al., 2015). In a further animal study, LCD, an enzymatic active domain of *C. difficile* phage (phiC2) endolysin linked to the human defensin protein HD_5, protected infected mice (Peng et al., 2019).

One of the complications facing the application of recombinant endolysin is the delivery to the site of action. Digestive proteases lower the chance of delivering endolysins. In animal studies, gavage was used to deliver and assess the activity of endolysin (Peng et al., 2019). One approach that has been adopted to address this issue is to clone and express CD27L from the bacterium *Lactococcus lactis* to facilitate *in situ* expression (Mayer et al., 2008).

In this chapter, we describe the development of an *E. coli* based expression system for the production of LysCD6356, the lysin that in our earlier study, was found to have the widest range of activity against clinical isolates of *C. difficile*. We will also develop an *E. coli* based EAD expression system to determine if removal of the cell binding domain increases lytic activity.

3.2 Aims and objectives

The aim of this chapter is to clone, express, and purify full length LysCD6356 endolysin and the EAD of LysCD6356 from an *E. coli* host to support future characterization studies.

The objectives are to:

- To clone, express, and purify a codon optimized version of LysCD6356 from an *E. coli* based expression system.
- To clone, express, and purify a codon optimized version of EAD of LysCD6356 from an *E. coli* based expression system.

3.3 Results

3.3.1 Agarose gel electrophoresis of PCR products to confirm the size of the cloned sequence

The predicted size of the PCR products that would be generated if the gene sequences encoding the full length lysin and EAD fragment were present were 1004 bp and 719 bp, respectively. As can be seen from the Figure below, PCR products of approximately the correct size were obtained, suggesting that cloning had been successful (Figure 3.1).

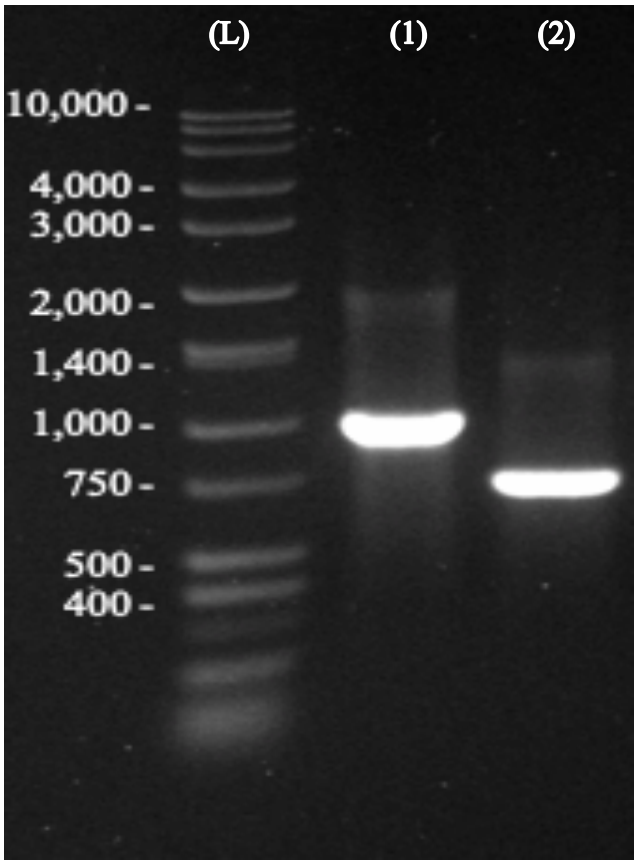


Figure 3.1 A 1% agar gel showing the migration distances of PCR products generated using template DNA derived from a construct containing the full length LysCD6356 (1) and the EAD domain of LysCD6356 (2). (L) Direct Load TM (Merck, UK).

These results confirm the quality control data provided by Genscript following the cloning of the inserts.

3.3.2 Recombinant protein expression

E. coli BL21(DE3)pLysS cells carrying either the LysCD6356 or its EAD gene construct were cultured to produce recombinant endolysins. Following expression, cells were lysed, and his-tagged proteins were purified using a Ni-NTA resin (Qiagen, UK). Samples collected through the purification process were used to optimize the expression and purification processes. The predicted sizes of the recombinant endolysins were estimated using two online tools as between 32.67 kDa for full length LysCD6356 and 22.12 kDa for the EAD fragment (https://www.bioinformatics.org/sms/prot_mw.html and https://web.expasy.org/compute_pi/)

3.3.3 Recombinant protein expression of full length Lys CD6356

As can be seen in the Coomassie stained PAGE gel below, a number of bands were present in the elution fraction, including a band corresponding to the predicted size of full length LysCD6356 (Figure 3.2).

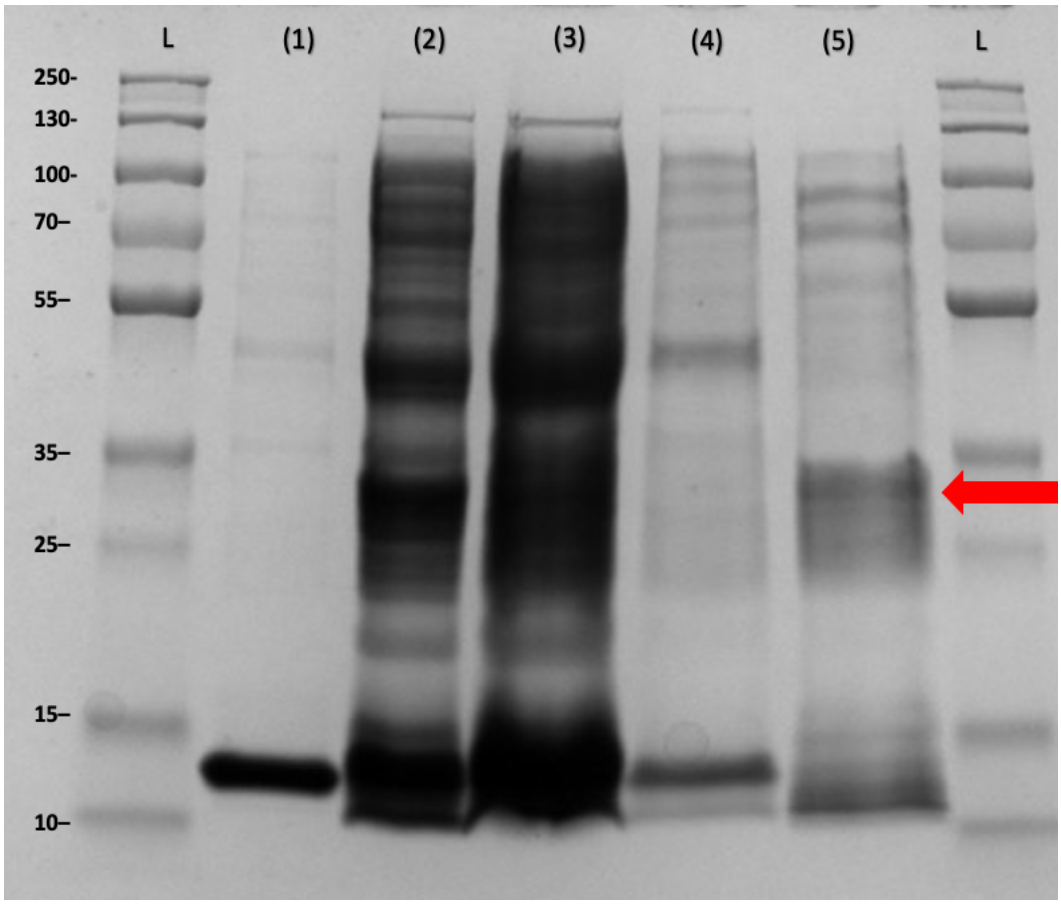


Figure 3.2 An SDS-PAGE gel showing different fraction following the induction and purification of a culture of *E. coli* BL21(DE3) pLysS lysCD6356 expressing recombinant full length LysCD6356.

The gel was stained with Coomassie brilliant blue to visualize protein bands. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) Elution fraction containing 250 mM imidazole, (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa). The red arrow indicated the predicted size of the recombinant LysCD6356 band (32.67 kDa).

The relatively large number of protein bands in the elution fraction indicates that "other" proteins are adhering to the QIAgen column. The column works on the premise that the amino acid histidine will bind to nickel, which is incorporated into the column. Other proteins also contain histidine, but not as a multimeric N-terminal tag. This may explain their presence in the elution fraction. In order to confirm that a band with an N-terminal histidine tag has been expressed, we repeated the analysis and performed a Western blot in which we probed the membrane with an antibody which specifically binds to the histidine tag region.

3.3.4 Western blot

The LysCD6356 recombinant protein has a histidine tag attached to the N terminus of the protein. In addition to providing a purification tag, it also provides a means of identifying the protein in a complex mixture. To determine if the recombinant protein was present, we first separated the proteins on an SDS polyacrylamide gel, transferred the proteins to a PVDF membrane, and then probed them with a Penta-His antibody. As can be seen in Figure 3.3, the elution fraction included a number of bands, including a band of size that corresponded to that of full length LysCD6356. The smaller band, approx. 14 kDa, could represent a breakdown product of the full length LysCD6356, although its presence in the pre-induction sample suggests that this is not the case and that this band is unrelated to LysCD6356. The larger bands between 55 and 100 kDa in size could represent multimeric forms of the protein, or like the 14 kDa band, they could be culture derived proteins unrelated to LysCD6356.

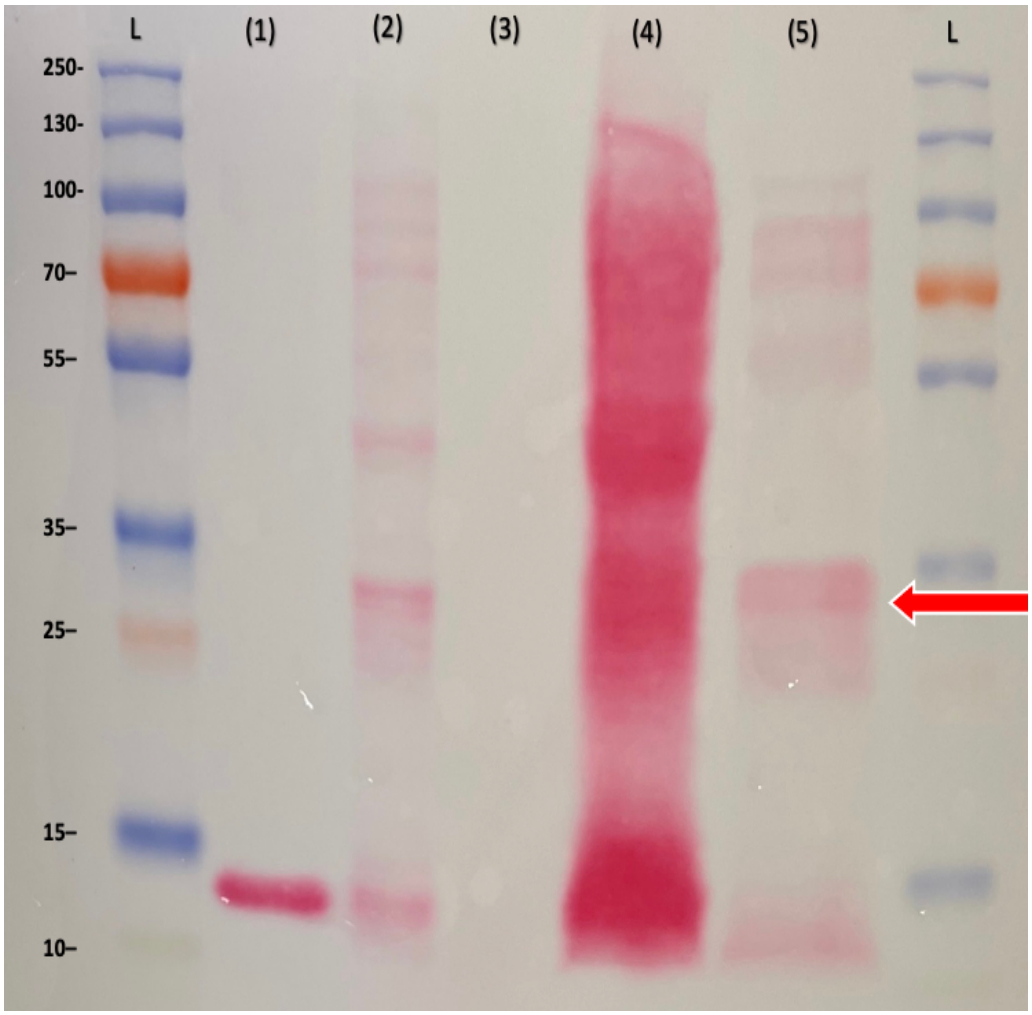


Figure 3.3 A PVDF membrane with different fraction of recombinant LysCD6356 collected during and after the expression and purification process

The membrane was stained with 0.1% Ponceau S solution. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) Elution fraction, (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder), on PVDF membrane that was stained with ponceau S solution to confirm the transformation of bands from polyacrylamide gel to the membrane. Red arrow indicates the expected location of full length LysCD6356 on the basis of its predicted size (32.67 kDa).

To determine if the protein band at around 32 kDa was full length LysCD6356, we probed the membrane with a pent-His antibody (THE™ His tag antibody) murine antibody specific for the histidine tag. The secondary antibody was used at a concentration of 1:10,000. As can be seen from Figure 3.4, the antibody complex bound to a large band at 32 kDa, which corresponds in size to full length LysCD6356. This band was present in the elution fraction but was absent from the pre-induction sample (Figure 3.3).

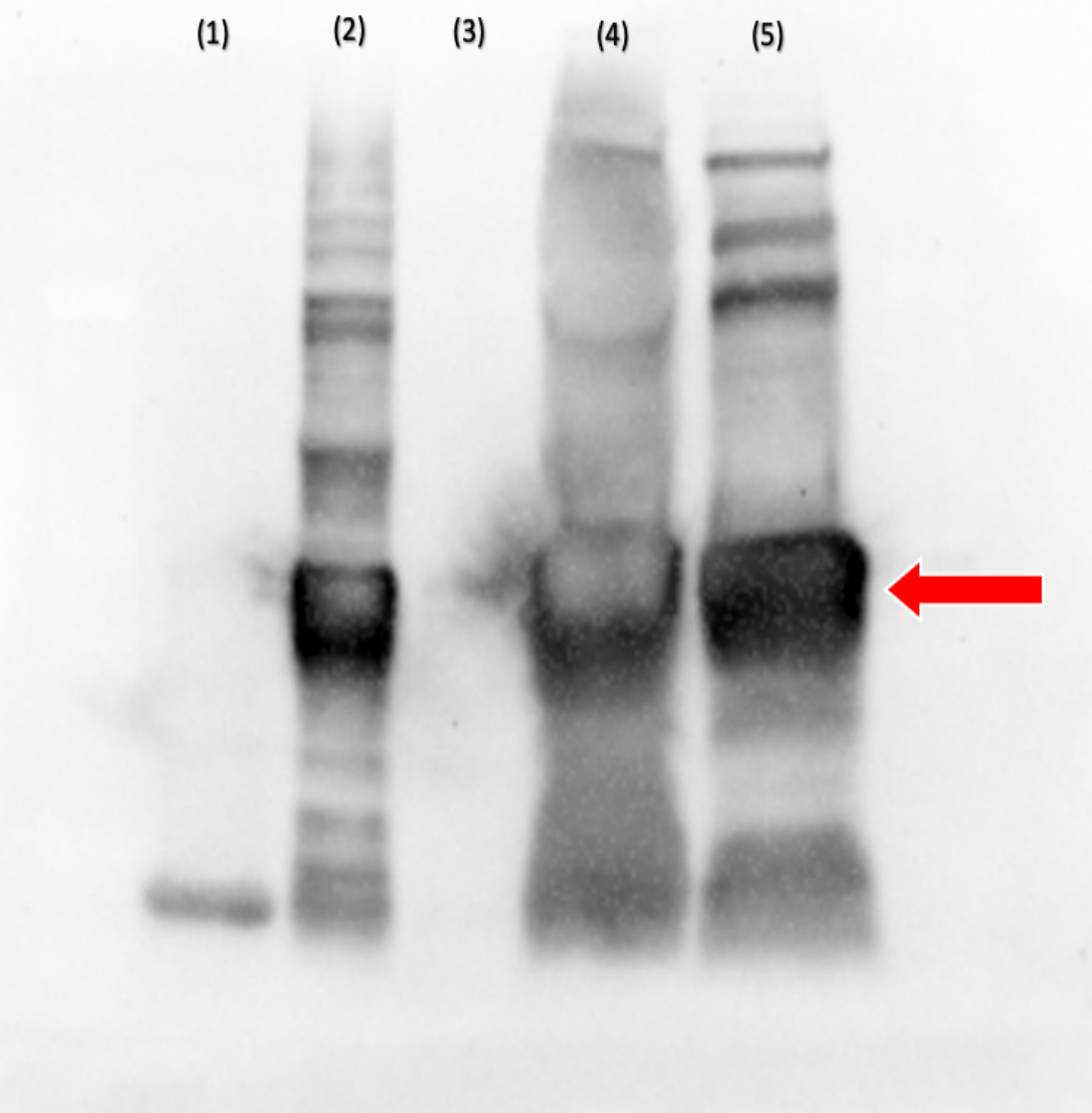


Figure 3.4 A Western blot of a PVDF membrane with different fraction of recombinant LysCD6356 collected during and after the expression and purification process.

The membrane has been stained with a histidine specific antibody at a concentration of 1:10,000 and a secondary goat anti-mouse detection antibody at a concentration of 1:10,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) Elution fraction. The red arrow represents the predicted location of LysCD6356 ~(32.67KDa).

The antibody also recognized a number of smaller and larger bands, suggesting the presence of multimeric forms of LysCD6356, degraded forms of the protein, or a high degree of non-specific binding by the antibody complex. The secondary antibody was used at a concentration of 1:10,000

as described before (Alyousef, 2013) and as listed as the highest concentration tested by the manufacturer. In subsequent studies, this antibody was used at a dilution of 1:200,000 as the lowest concentration tested by the manufacturer (ThermoFisher Scientific, UK).

3.3.5 Optimization of recombinant protein production and purification

To enhance the recovery of the recombinant protein, we increased the volume of the initial culture from 250 mL to 1 L and included a sonication step in the protein purification process. We also increased the binding capacity of the Ni-NTA columns from 300 µg to up to 75 mg of histidine tagged protein by using 1.5 mL bed volume Ni-NTA resin (binding capacity is up to 50 mg/mL) instead of a prepacked column (binding capacity is 300 µg/column). Fresh lysis, washing, and elution buffers were made from their constituent chemicals to address the possibility that there may be a problem with the commercially supplied buffers. We also increased the number of washing steps from two to five, as recommended in the Qiagen handbook, in an attempt to reduce non-specific protein binding to the column.

3.3.6 Modifying the concentration of imidazole in the elution buffer to enhance the recovery of full size protein.

The presence of additional bands in the elution fraction is undesirable. In an effort to remove these additional bands, we investigated the impact of modifying the concentration of imidazole in the elution buffer. The role of imidazole is to displace the histidine tagged protein from the Ni-NTA column. We hypothesized that the histidine tagged protein would adhere more strongly to the Ni-NTA column than proteins that lacked the presence of a 10 x histidine tag.

For proprietary reasons, Qiagen does not publish the concentration of imidazole in their elution buffer.

In order to assess the effect of using different concentrations of imidazole in the elution buffer, proteins were eluted sequentially using 3 mL of elution buffer containing increasing concentrations of imidazole (50, 100, 150, and 250 mM). The eluted samples were then processed to remove

imidazole from the sample and replace it with PBS. This was done because high imidazole concentration (more than 50 mM) is incompatible with protein quantification using BCA assay as demonstrated by the manufacturer (ThermoFisher Scientific, UK). Given that the recombinant protein we are expressing is an enzyme which will be used in future studies to assess its ability to digest the cell wall of *C. difficile,* identifying the exact protein concentration is important.

To remove imidazole, individual eluted fractions were transferred to the tubes provided as part of the Pierce Protein Concentrator PES, MWCO 10kDa system (Fisher Scientific, UK). Following the addition of the elution fraction, tubes were centrifuged at 8000 g at 4 ºC until the solution had passed through the membrane. 4 mL of PBS were then added to the concentrator tube and centrifuged as described before. The proteins were recovered from the filter in 1 mL of the new carrier (PBS). The protein composition of these samples was determined by SDS-PAGE and Western blot analysis.

As can be seen from Figure 3.5, imidazole at concentrations of 150 mM and above markedly reduced the number of bands in the elution fraction, with the major band corresponding in size to that of the predicted size of full length LysCD6356. An additional faint band could also be seen at around 70 kDa. To determine the nature of these additional faint bands, the membrane was probed with the histidine tagged antibody.

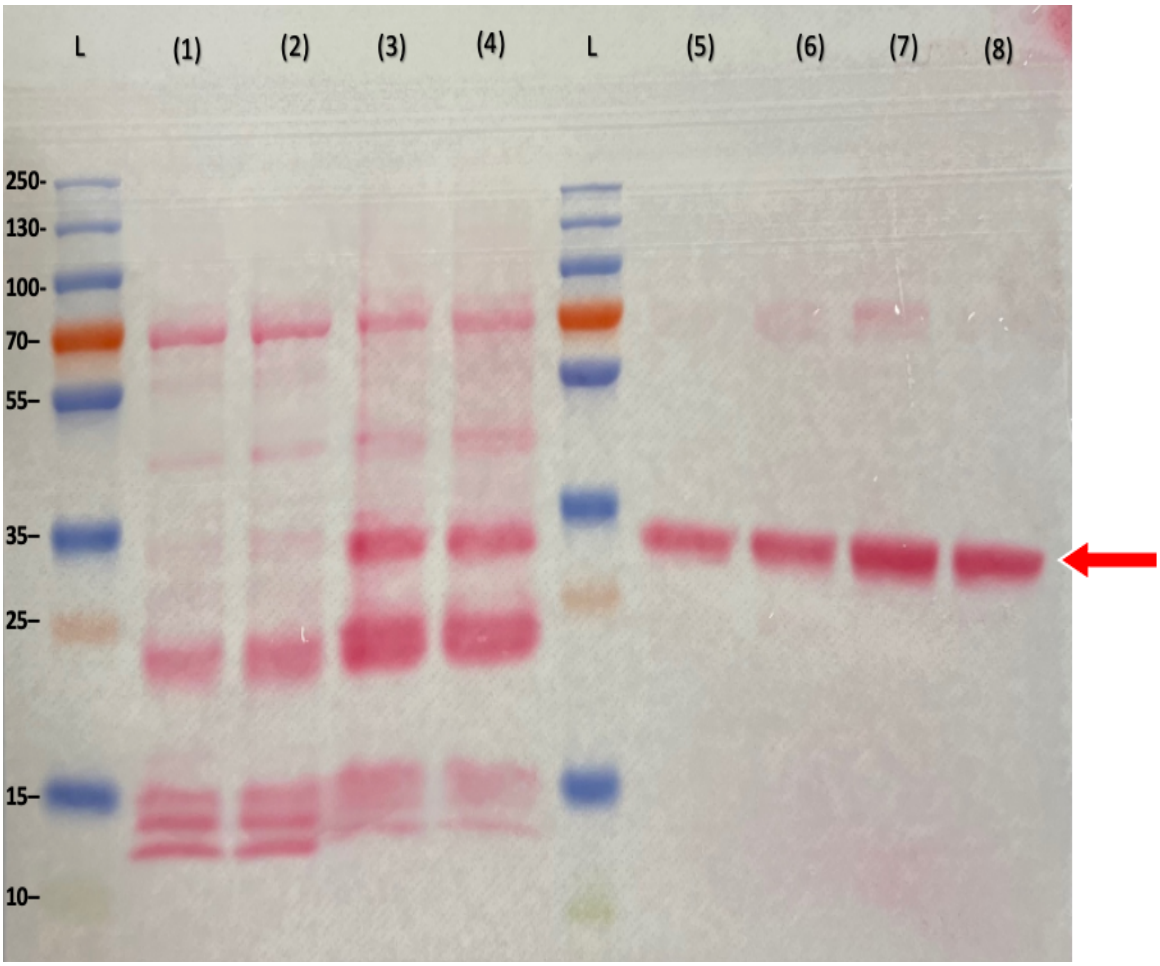


Figure 3.5 PVDF membrane of different fraction of recombinant LysCD6356 following elution for a Ni-NTA column using different concentrations of imidazole.

The membrane has been stained with 0.1% Ponceau S solution. (1) and (2) elution fraction containing 50 mM imidazole, (3) and (4) elution fraction containing 100 mM imidazole, (5) and (6) elution fraction containing 150 mM imidazole, (7) and (8) elution fraction containing 250 mM imidazole, (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant LysCD6356 band (32.67 kDa).

As can be seen in Figure 3.6, the histidine tag specific antibody generated a strong signal with a band of similar size to LysCD6356. Additional fainter bands were detected possibly representing multimeric forms of the protein, breakdown products, or minor cross reacting contaminants in the sample.

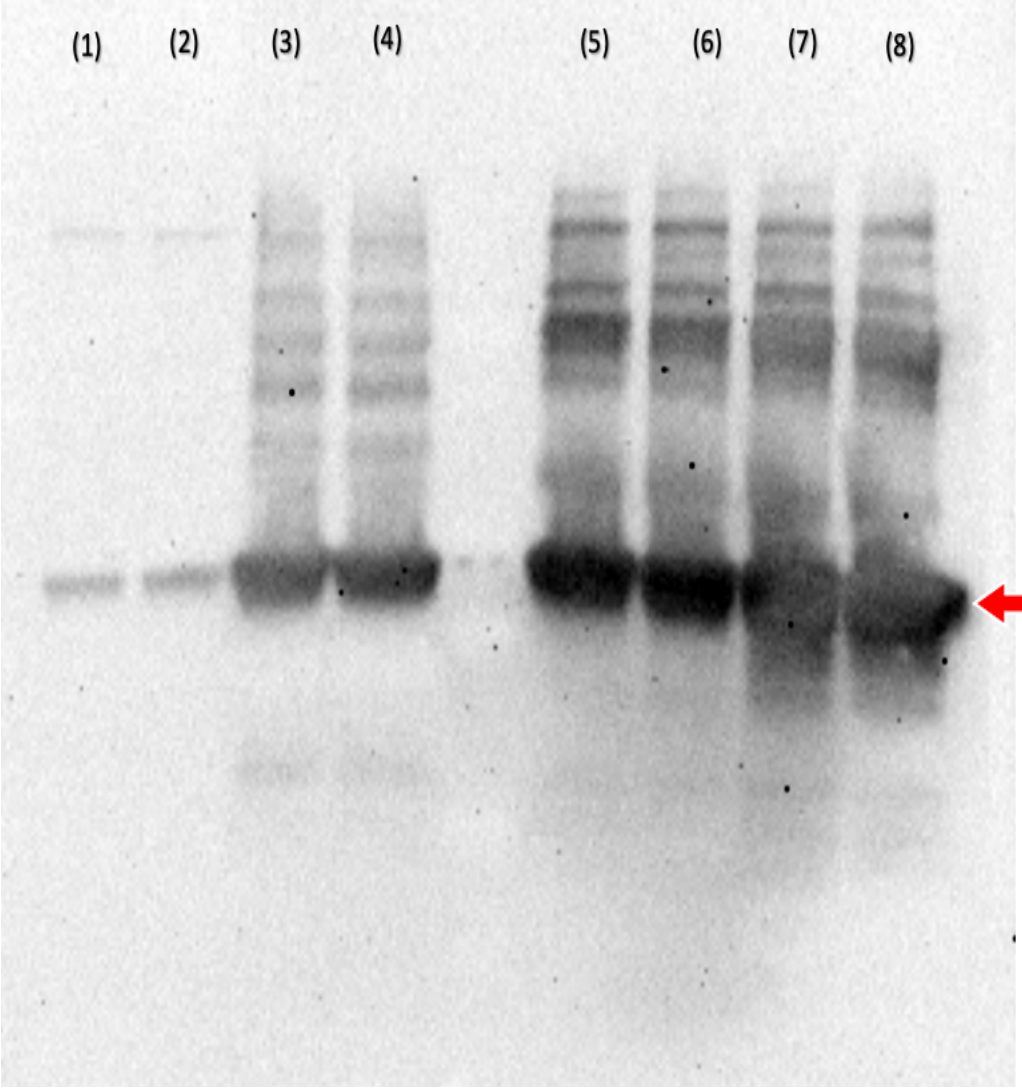


Figure 3.6 A Western blot of a PVDF membrane with different fraction of recombinant LysCD6356 following elution for a Ni-NTA column with different concentrations of imidazole.

The membrane has been stained with a histidine specific antibody at a concentration of 1:10,000 and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (1) and (2) elution fraction containing 50 mM imidazole, (3) and (4) elution fraction containing 100 mM imidazole, (5) and (6) elution fraction containing 150 mM imidazole, (7) and (8) elution fraction containing 250 mM imidazole, (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant LysCD6356 band (32.67 kDa).

3.3.7 Production and purification of LysCD6356 to support lysin activity studies

Based on the results generated up to this point, we employed the following strategy to produce recombinant lysin: The Ni-NTA column was treated with 150 mM imidazole to eliminate the majority of the non-specifically bound proteins, and the full length LysCD6356 was recovered by washing the column with 250 mM imidazole. As can be seen in Figure 3.7, the bands of LysCD6356 in the post-imidazole samples (lanes 5 and 6) are stronger than those in the pre-imidazole samples.

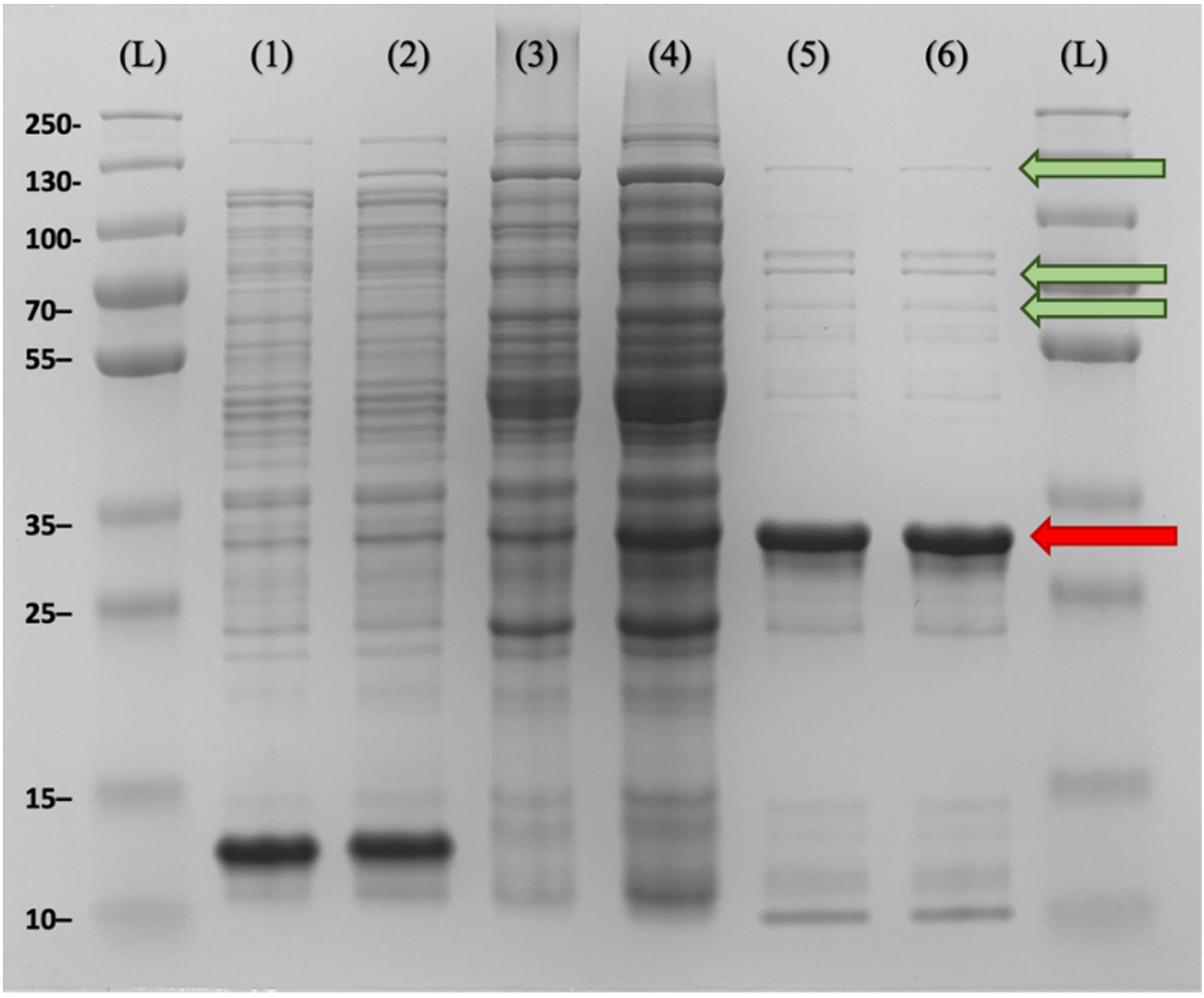


Figure 3.7 SDS-PAGE analysis of different fractions of recombinant LysCD6356 following elution through an Ni-NTA column and the removal of imidazole.

Coomassie Blue staining of an SDS-PAGE. (1) pre-induction fraction, (2) post-induction fraction, (3) flow through fraction, (4) wash fraction, (5) and (6) elution fraction following removal of imidazole and (L) molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). The red

arrows represent the predicted location of LysCD6356 ~(32.67kDa) while green arrows represent multimeric forms of LysCD6356.

The fractions were then run on SDS-PAGE for western blot analysis. To confirm protein transfer, the PVDF membrane was stained with a 0.1% Ponceau S solution. As can be seen in Figure 3.8, the full length LysCD6356 was the strongest band in the elution fraction (lanes 7 and 8) compared to the samples prior to the removal of imidazole. This is because we reduced the volume of solution in which the recovered protein was resuspended.

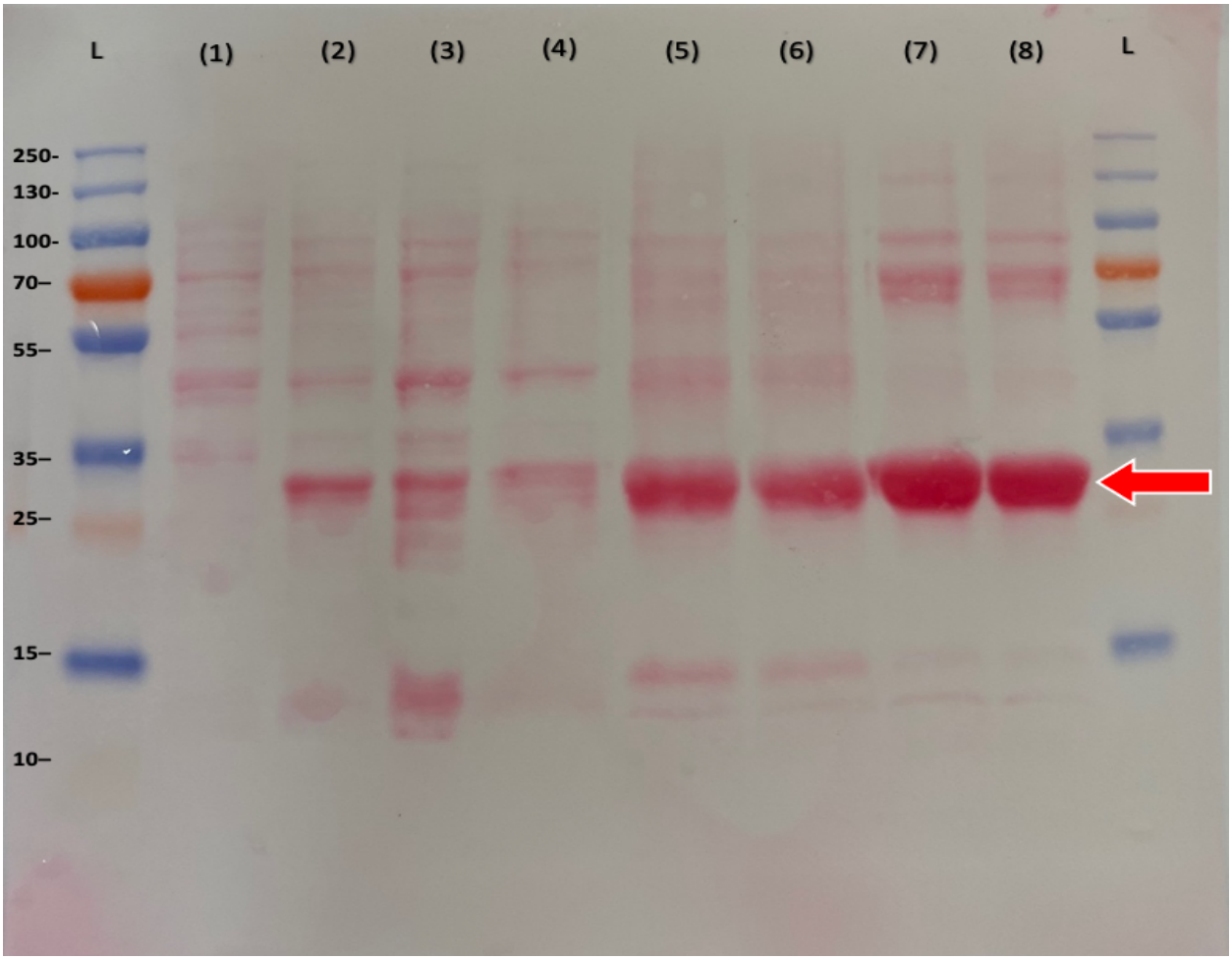


Figure 3.8 PVDF membrane of different fractions of recombinant LysCD6356 following elution for a Ni-NTA column and the removal of imidazole.

The membrane has been stained with 0.1% Ponceau S solution. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) and (6) elution fraction before the removal of imidazole (250mM), (7) and (8) elution fraction following removal of imidazole, (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). The red arrow represents the predicted location of LysCD6356 ~(32.67 kDa).

To confirm that the major band represents the recombinant protein, the membrane was probed with a histidine tag specific monoclonal antibody. As can be seen from Figure 3.9, the major band, which corresponds in size to LysCD6356, gave a strong signal with the antibody, suggesting a high concentration of this protein in the samples. Minor bands were also recognized in the concentrated sample, suggesting that break down products and multimeric forms of LysCD6356 may be present.

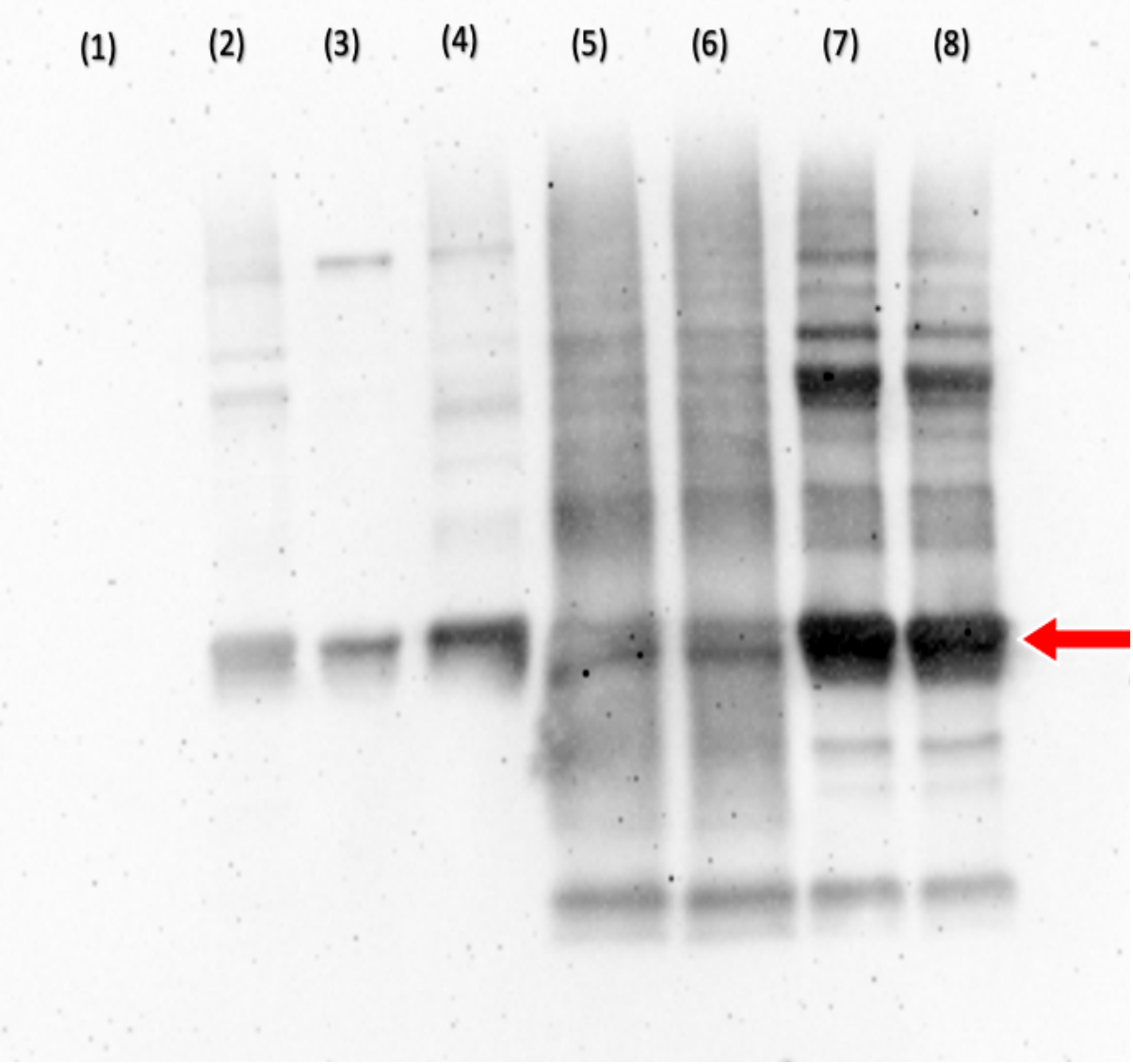


Figure 3.9 A Western blot of a PVDF membrane with different fraction of recombinant LysCD6356 following elution for a Ni-NTA column and the removal of imidazole.

The membrane has been stained with a histidine specific antibody at a concentration of 1:10,000 and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) and (6) elution fraction before removal of imidazole; (7) and (8) elution fraction following removal of imidazole. The red arrow represents the predicted location of LysCD6356 ~(32.67 kDa).

3.3.8 Quantifying of the percentage of full length LysCD6365 in the recombination protein preparation using Bio-Rad software

To determine the proportion of full length recombinant LysCD6356 in our extract, we used Bio-Rad software to analyze the image density of individual bands. Full length LysCD6356 represents approximately 88.1%, of which 82.2% was monomeric in the elution fraction. Using the initial purification method, the final protein yield of LysCD6356 was approximately 350 μg/L of the original 1L culture. The protein concentration was estimated using the BCA assay as described in Chapter 2 section 2.2.15.

3.3.9 Recombinant protein expression of the EAD of LysCD6356

In addition to expressing the full length lysin from an *E. coli* expression system, we also expressed the EAD from the following strain BL21(DE3) pLysS pET-19b EAD using the same culture and induction conditions. As can be seen in Figure 3.10, a band corresponding to the predicted size of EAD (approx. 22 kDa) can be seen in the elution fraction.

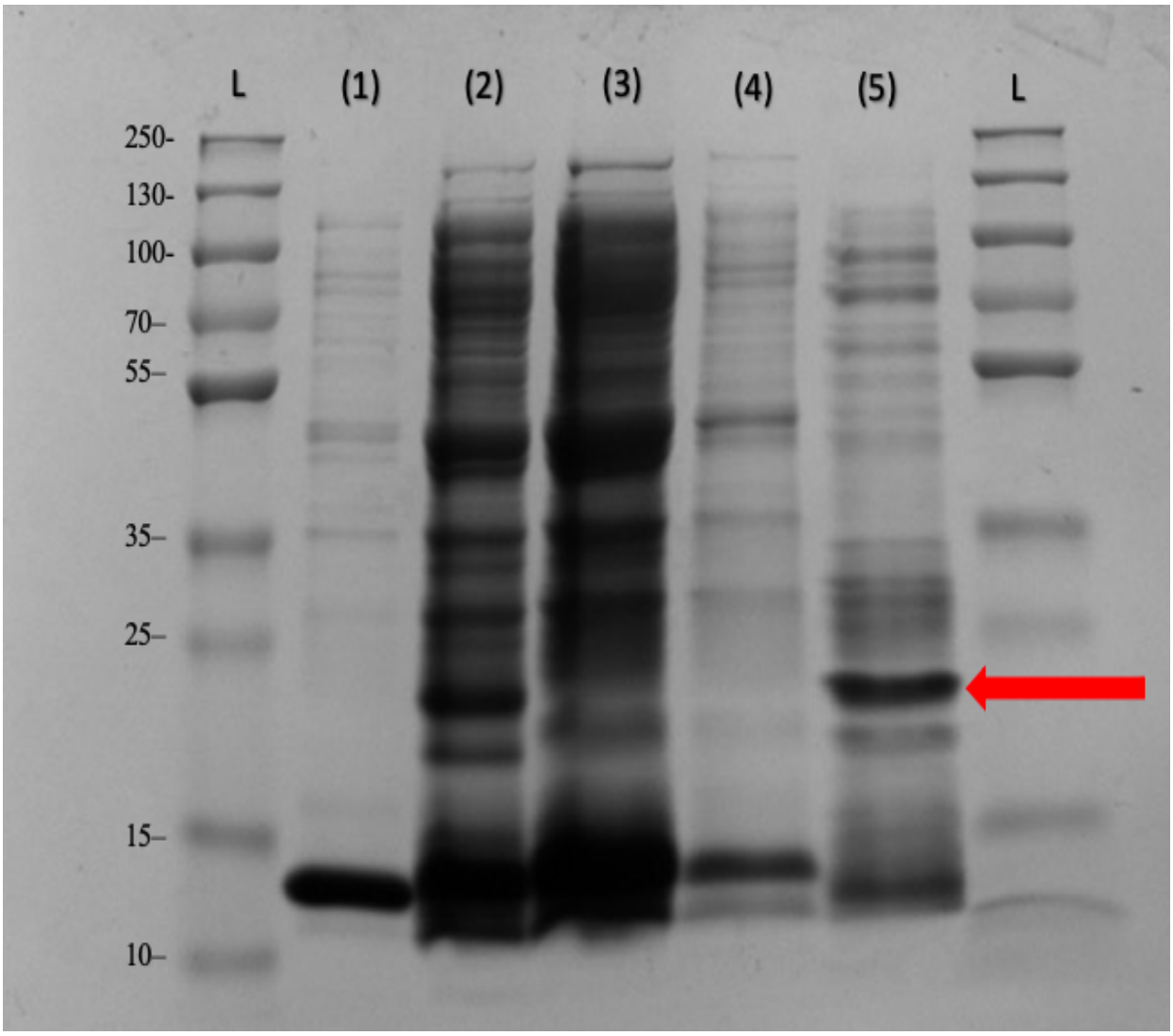


Figure 3.10 An SDS-PAGE gel showing different fractions following the induction and purification of a culture of *E. coli* BL21(DE3) pLysS EAD of LysCD6356 expressing recombinant EAD of LysCD6356.

The gel was stained with Coomassie brilliant blue to visualize protein bands. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) Elution fraction, (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant EAD of LysCD6356 (22.12 kDa).

As was the case for the full length lysin, the relatively large number of bands in the elution fraction indicates that "other" proteins are also sticking to the Qiagen column. Further optimization of the expression and purification conditions needs to be undertaken.

3.3.10 Optimization of recombinant protein production and purification

To increase the recovery of recombinant EAD from BL21(DE3) pLysS pET-19b EAD, we applied the same optimization approaches as were used to increase the level of expression and recovery of LysCD6356 and obtained similar results.

3.3.11 Modifying the concentration of imidazole in the elution buffer to enhance the recovery of EAD fraction.

As was the case for full length CD6356, we explored the impact of increasing the concentration of imidazole in the elution buffer on the recovery of EAD from the Ni-NTA column. The imidazole was removed from the sample and replaced with PBS using Pierce Protein Concentrator PES, MWCO 10 kDa system (Fisher Scientific, UK). The purity of the elution fractions was then assessed by SDS-PAGE and western blot.

As can be clearly seen in Figure 3.11, the protein fraction was eluted using 250 mM imidazole contained fewer minor bands compared to the samples sequentially eluted with increasing concentrations of imidazole (50, 100, and 150 mM). The major band had an estimated molecular weight of 22 kDa, which corresponds to the predicted size of recombinant EAD (22.12 kDa). Additional minor bands appear around 45 kDa. In order to investigate the nature of these minor bands, the PVDF membrane was probed with the histidine tagged antibody.

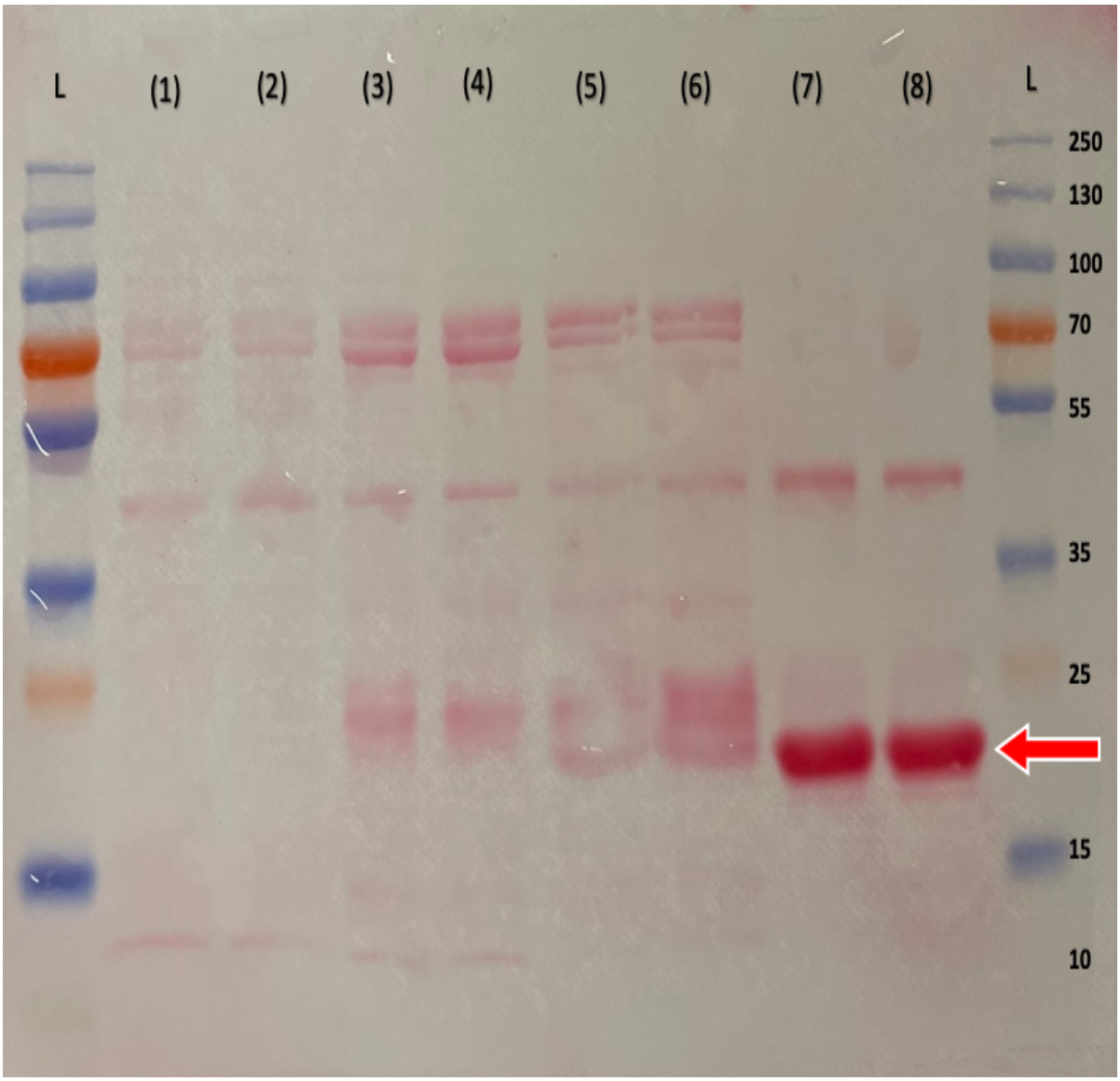


Figure 3.11 PVDF membrane of different fraction of recombinant EAD of LysCD6356 following elution for a Ni-NTA column using different concentrations of imidazole.

The membrane was stained with 0.1% Ponceau S solution. (1) and (2) elution fraction containing 50 mM imidazole, (3) and (4) elution fraction containing 100 mM imidazole, (5) and (6) elution fraction containing 150 mM imidazole, (7) and (8) elution fraction containing 250 mM imidazole, (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant EAD of LysCD6356 band (22.12 kDa).

Probing with the antibody revealed the presence of a major band corresponding in size to full size EAD with a second, fainter band with an approximate size of 45 kDa (Figure 3.12). A third faint band is located around 70 kDa. On the basis of their sizes, these bands could represent multimeric forms of EAD, although this hypothesis would need to be confirmed.

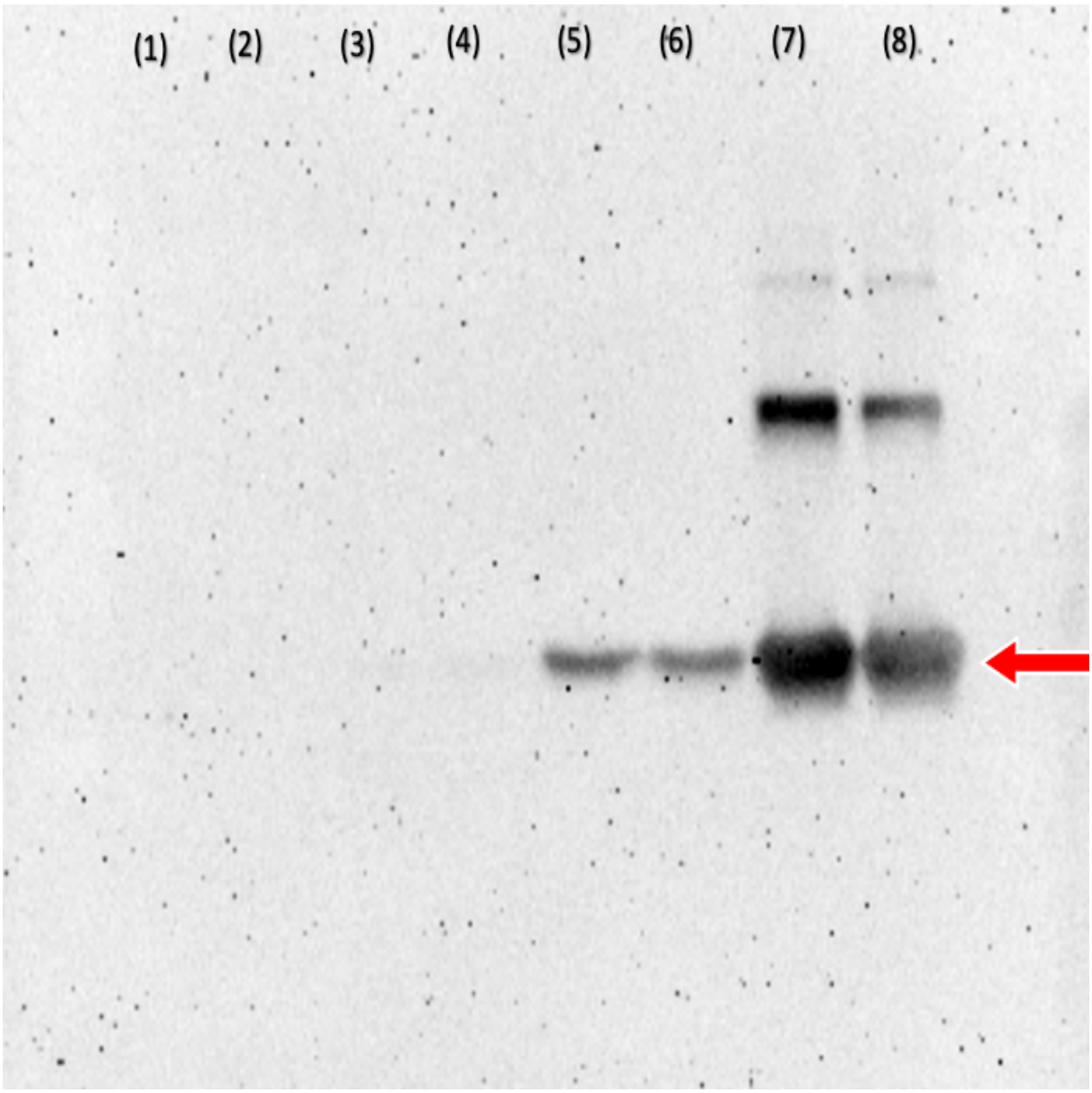


Figure 3.12 Western blot of a PVDF membrane with different fraction of recombinant EAD of LysCD6356 following elution for a Ni-NTA column with different concentrations of imidazole.

The membrane was stained with a histidine specific antibody at a concentration of 1:10,000 and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (1) and (2) elution fraction containing 50 mM imidazole, (3) and (4) elution fraction containing 100 mM imidazole, (5) and (6) elution fraction containing 150 mM imidazole, (7) and (8) elution fraction containing 250 mM imidazole, (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant EAD of LysCD6356 band (22.12 kDa).

To determine if we could further purify the EAD fraction, we examined the effect of treating the Ni-NTA column with 200 mM of imidazole. A similar pattern to that obtained using 150 mM imidazole was obtained (Figure 3.13).

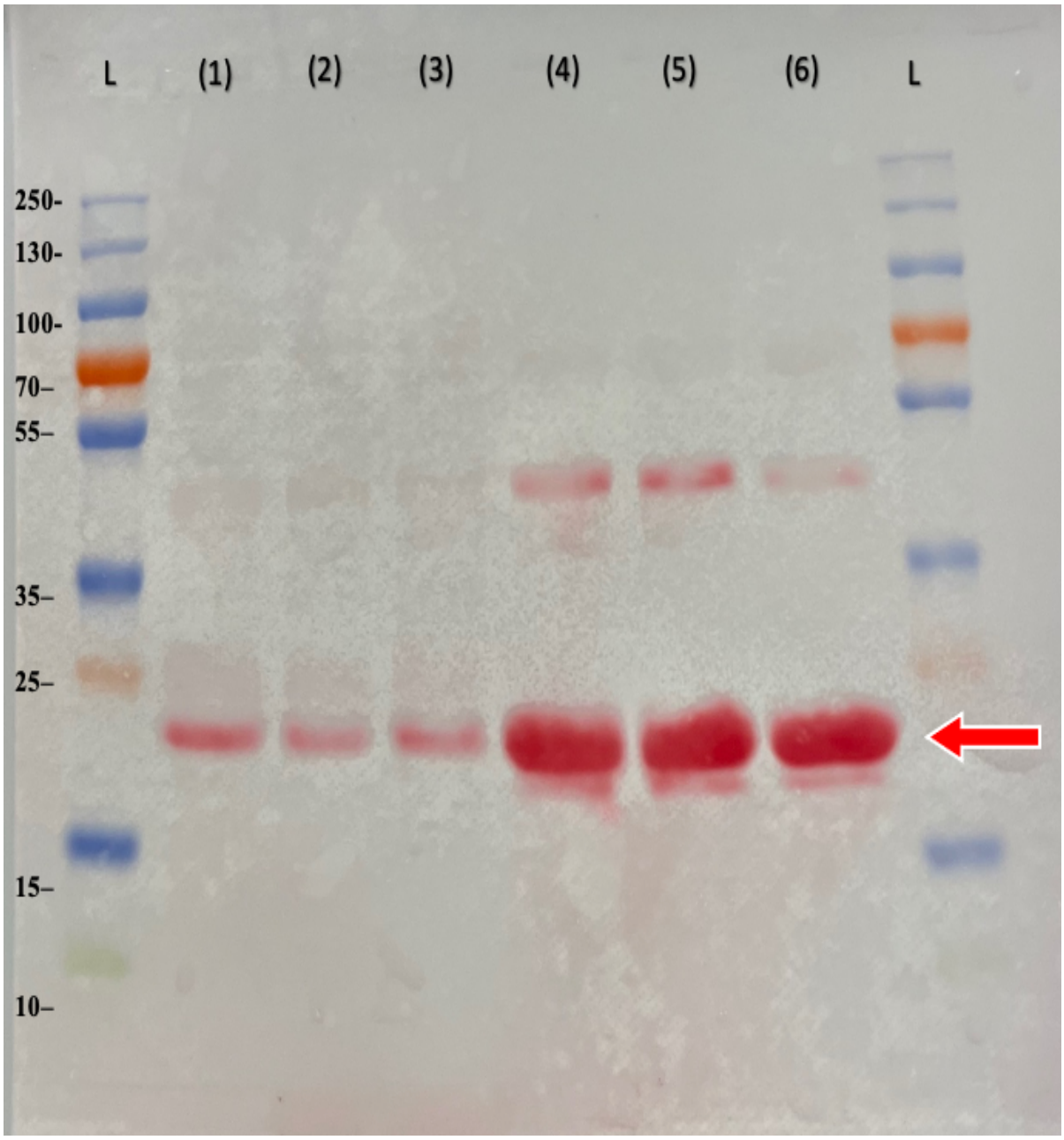


Figure 3.13 PVDF membrane of different fraction of recombinant EAD of LysCD6356 following elution for a Ni-NTA column using different concentrations of imidazole.

The membrane was stained with 0.1% Ponceau S solution. (1), (2) and (3) are elution fraction containing 200 mM imidazole, (4), (5) and (6) are elution fraction containing 250 mM imidazole,

(L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant EAD of LysCD6356 band (22.12 kDa).

3.3.12 Production and purification of EAD of LysCD6356 to support lysin activity studies

We originally planned to produce a pure recombinant EAD of LysCD6356. Unfortunately, the impact of COVID forced us to modify our approach and move forward with a partially purified product.

Based on the results to date, we adopted a similar approach to that used for LysCD6356 and treated the Ni-NTA column with 200 mM imidazole to eliminate the majority of the non-specifically bound proteins. The remaining EAD of LysCD6356 was recovered by washing the column with imidazole at 250 mM.

As can be seen in Figure 3.14, the bands of EAD in the post-imidazole sample (lane 5) are stronger than those in the pre-imidazole samples.

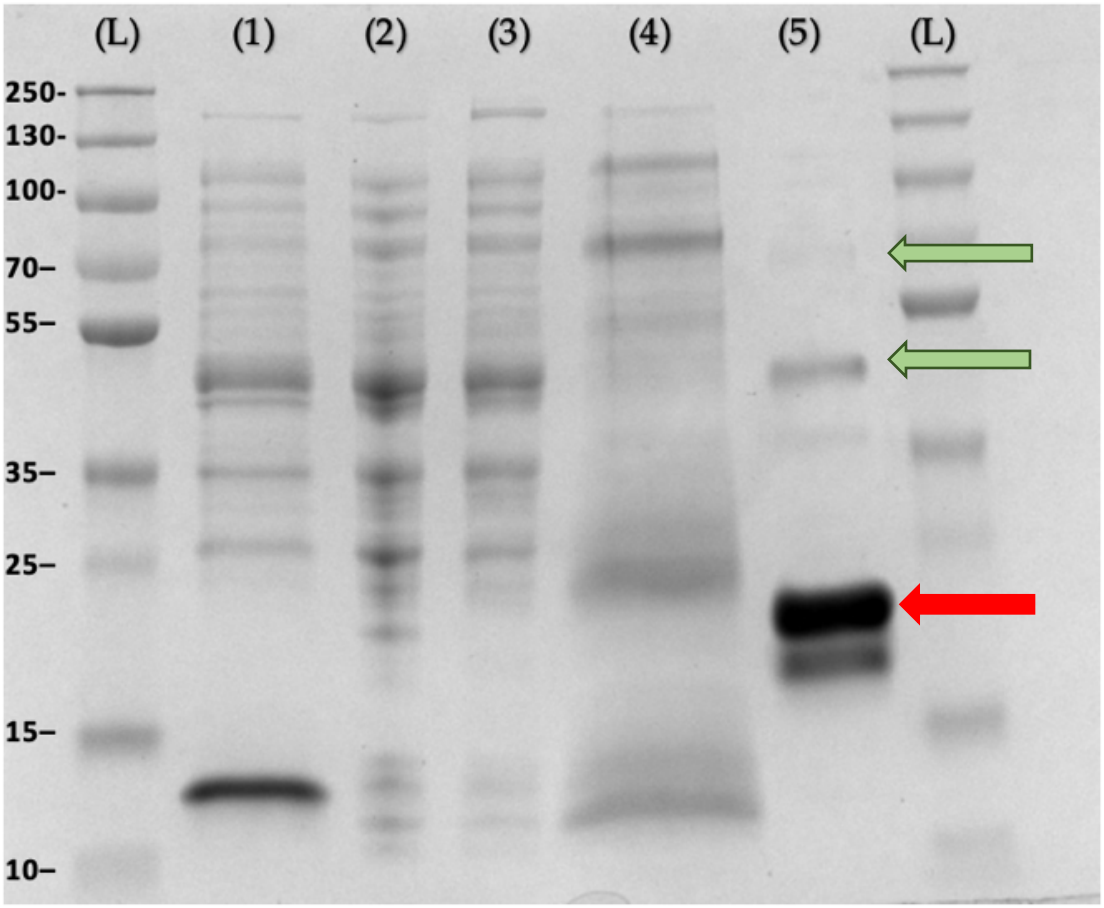


Figure 3.14 SDS-PAGE analysis of different fractions of recombinant EAD of LysCD6356 following elution through an Ni-NTA column and the removal of imidazole.

Protein band detection using Coomassie Blue staining on SDS-PAGE. (1) pre-induction fraction, (2) post-induction fraction, (3) flow through fraction, (4) wash fraction, (5) elution fraction after the removal of imidazole (250mM), (L) molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow represents the predicted location of EAD of LysCD6356 band (22.12 kDa) while green arrows represent multimeric forms of EAD.

The fractions were run on SDS-PAGE for western blot analysis. To confirm protein transfer, the PVDF membrane was stained with a 0.1% Ponceau S solution. As can be seen in Figure 3.15, proteins appear to be transferred, and EAD appears stronger in the elution fraction compared to other fractions.

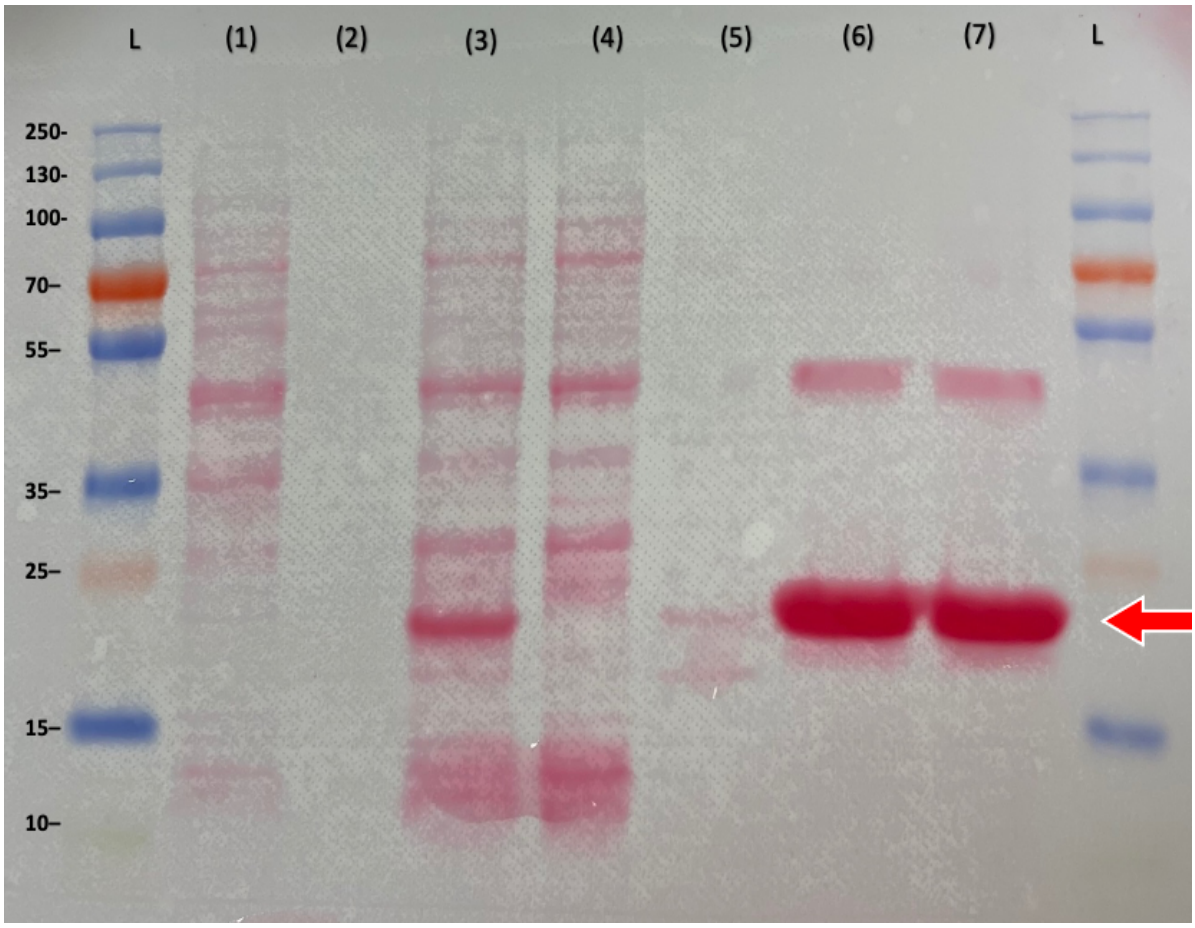


Figure 3.15 PVDF membrane of different fractions of recombinant EAD of LysCD6356 following elution for a Ni-NTA column and the removal of imidazole.

The membrane has been stained with 0.1% Ponceau S solution. (1) Pre-induction fraction, (2) Blank, (3) Post-induction fraction, (4) Flow through fraction, (5) Wash fraction, (6) and (7) elution fraction after the removal of imidazole (250mM), (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow represents the predicted location of EAD of LysCD6356 band (22.12 kDa).

To confirm that high-density bands represent EAD recombinant protein, the PVDF membrane was treated with primary and secondary antibodies, and the attachment of antibodies was detected using a chemiluminescence substrate. As can be seen in Figure 3.16, the major bands pointed by the red arrow represent EAD recombinant protein in addition to other band around 45 kDa, suggesting a multimeric form of EAD may be present.

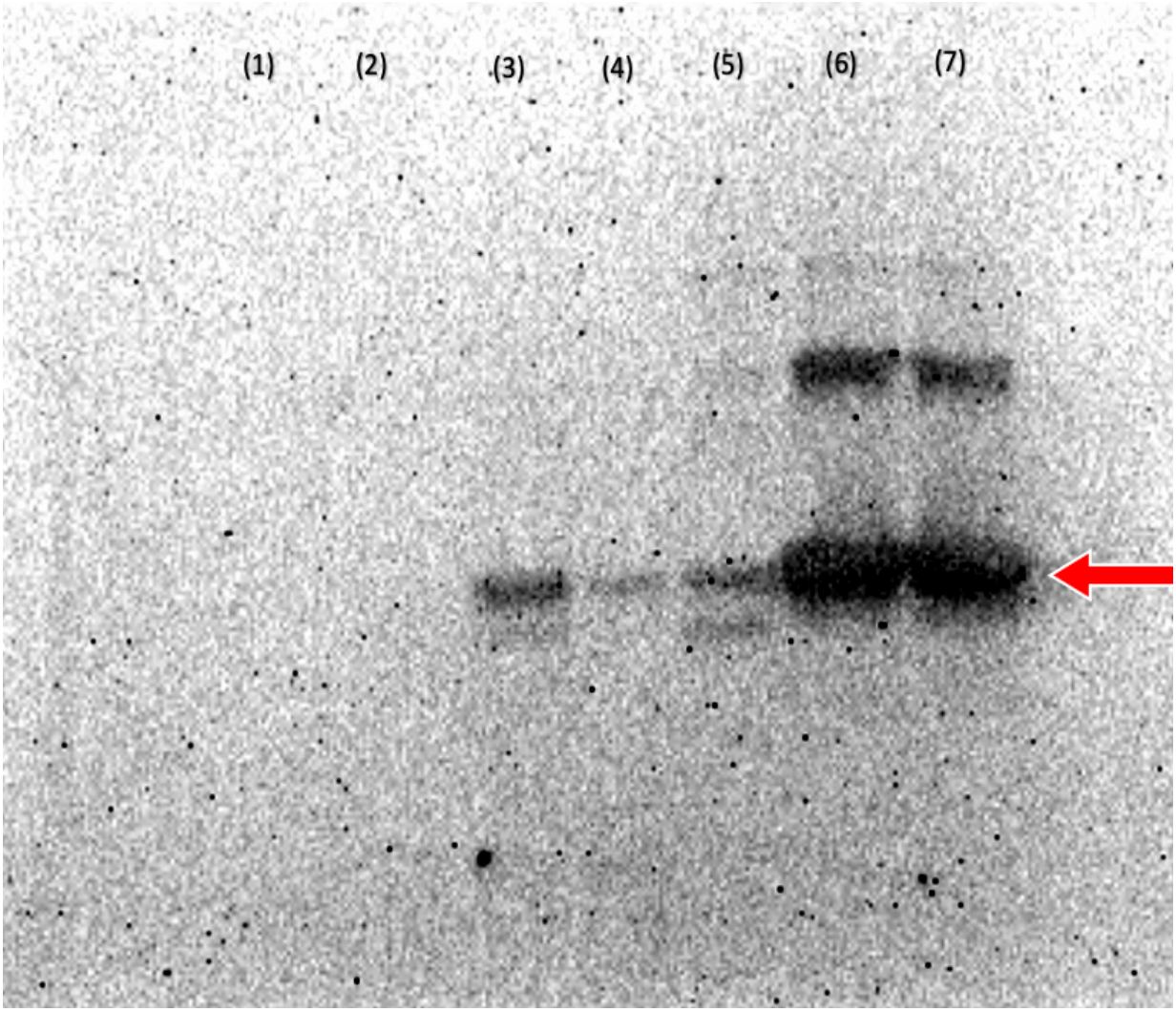


Figure 3.16 A Western blot of a PVDF membrane with different fraction of recombinant EAD of LysCD6356 following elution for a Ni-NTA column and the removal of imidazole.

The membrane was stained with a histidine specific antibody at a concentration of 1:10,000 and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (1) Pre-induction fraction, (2) Blank, (3) Post-induction fraction, (4) Flow through fraction, (5) Wash fraction, (6) and (7) elution fraction after the removal of imidazole (250mM), (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow represents the predicted location of EAD of LysCD6356 ~ (22.12kDa).

3.3.13 Quantifying of the percentage of EAD of LysCD6356 in our recombination protein preparation using Bio-Rad software.

To determine the proportion of the recombinant EAD of LysCD6356 in our extract, we used Bio-Rad software to analyze the image density of individual bands. Using Bio-Rad software, the proportion of recombinant EAD in the purified sample was estimated to be 86.1%, of which 73.8% was monomeric. Using the initial purification method, the final protein yield of EAD of LysCD6356 was approximately 300 μg/L.

3.3.14 Enhancing the purity and recovery of the EAD of LysCD6356

In an effort to improve the purity and recovery of the monomeric form of EAD, we modified the purification method as follows: The post induction bacterial lysate was suspended in free Ni-NTA resin and incubated on a roller mixer for one hour at 4 ºC instead of running the lysate through a column packed with Ni-NTA resin (section 3.3.11).

As can be seen in Figure 3.17, strong bands of the expected size of the EAD of LysCD6356 (22.12 kDa) were seen in the 200 mM and 250 mM imidazole elution fractions (lanes 6 and 7). The faint molecular weight band (around 50 kDa) seen in the 200 mM imidazole fraction was missing from the 250 mM imidazole elution fraction.

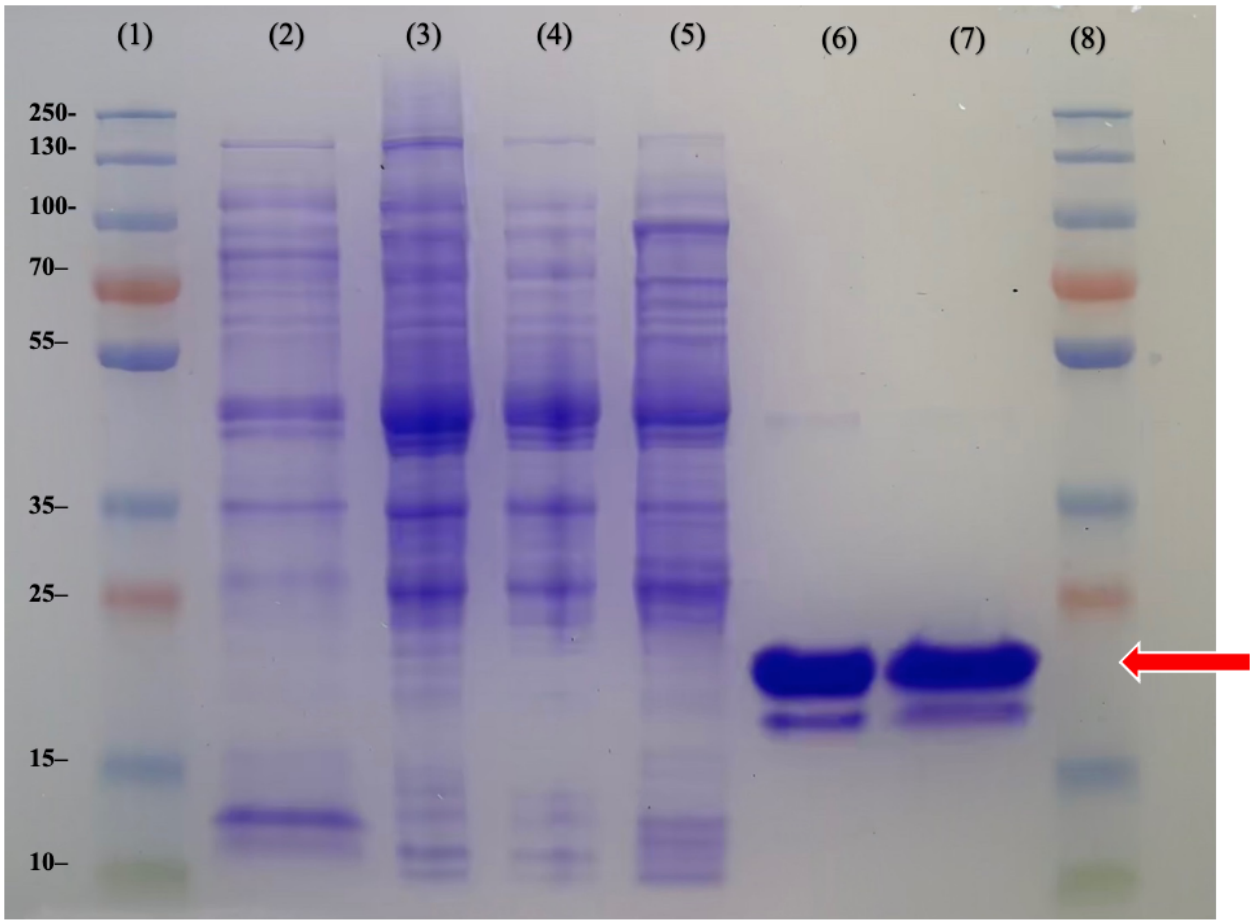


Figure 3.17 An SDS-PAGE gel showing different fraction following the induction and purification of a culture of *E. coli* BL21(DE3)pLysS pET-19b EAD of LysCD6356 expressing recombinant EAD of LysCD6356.

The gel was stained with Coomassie brilliant blue. (2) Pre-induction fraction, (3) post-induction fraction, (4) Flow through fraction, (5) Wash fraction, (6) 1st elution fraction (200 mM imidazole), (7) 2nd elution fraction (250mM imidazole), (1) and (8) molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant EAD of LysCD6356 (22.12 kDa). The purification process was performed using the modified purification method.

In addition to the strong EAD band, a second, less intense staining band was observed just below that of EAD. Since both proteins are highly expressed and are almost similar in size, it is possible that the lower band represents protein degradation or an oxidized version of EAD. In order to investigate this hypothesis, we replaced 2-mercaptoethanol, which is the reducing agent recommended by Bio-Rad, with dithiothreitol (DTT), a more potent reducing agent. When DTT treated samples were analyzed by SDS-PAGE, the lower molecular weight band could no longer be seen (Figure 3.18).

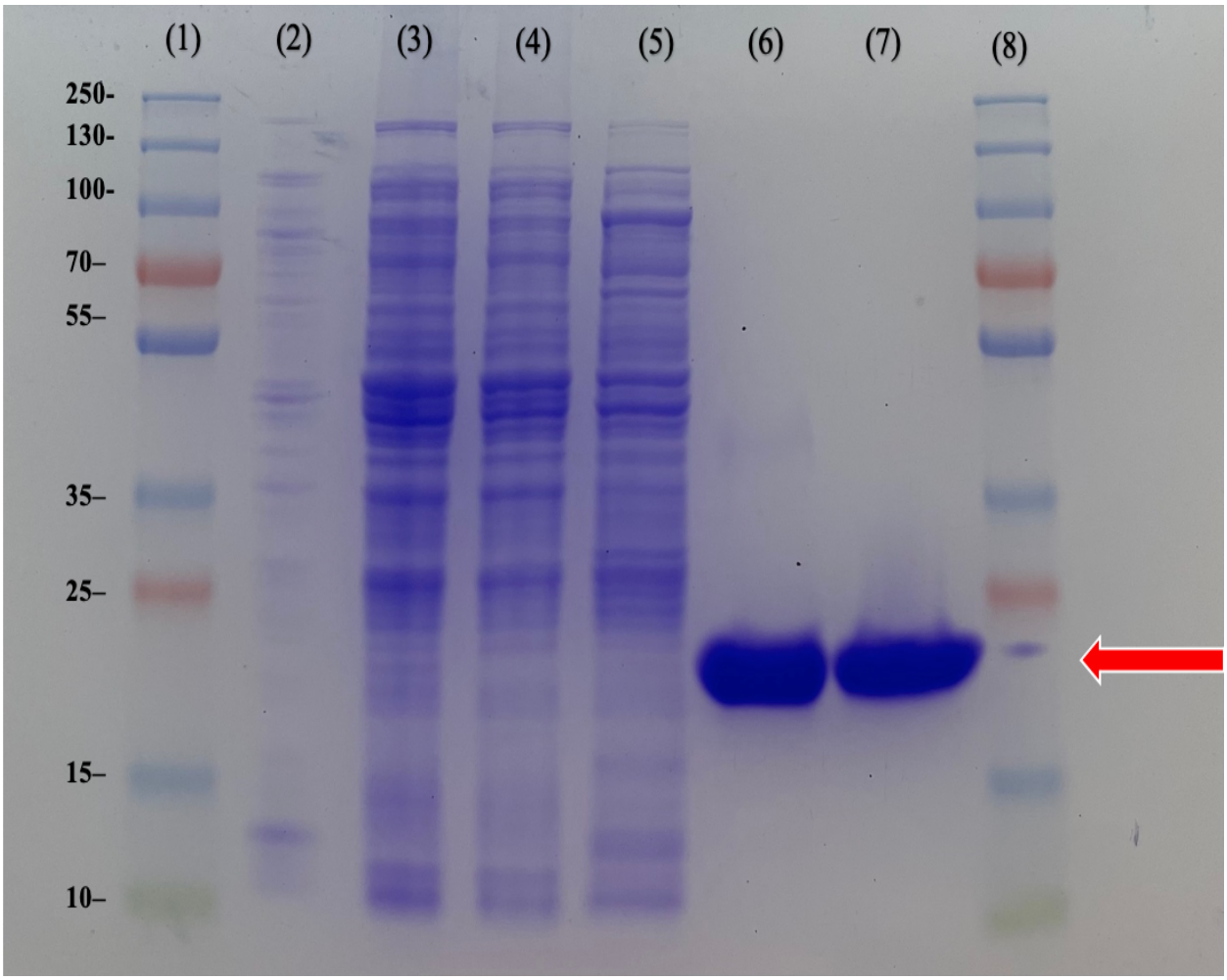


Figure 3.18 An SDS-PAGE gel showing different fraction of *E. coli* BL21(DE3)pLysS pET-19b EAD of LysCD6356 expressing recombinant EAD of LysCD6356 following using of DTT as reducing agent.

The gel was stained with Coomassie brilliant blue to visualize protein bands. (2) Pre-induction fraction, (3) post-induction fraction, (4) Flow through fraction, (5) Wash fraction, (6) 1st elution fraction (200 mM imidazole), (7) 2nd elution fraction (250mM imidazole), (1) and (8) molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant EAD of LysCD6356 (22.12 kDa).

To investigate the identity of the 50 kDa band in the 1st elution fraction (200 mM imidazole), a western blot was performed. As can be seen in Figure 3.19, two strong bands of the expected size of EAD (22.12 kDa) were seen in columns 6 and 7. A band of the same size was also seen in column number 3, which represents the post induction fraction. No band corresponding to the 50 kDa protein in column 6 was seen in the western blot, suggesting that this is not a multimeric form of the protein.

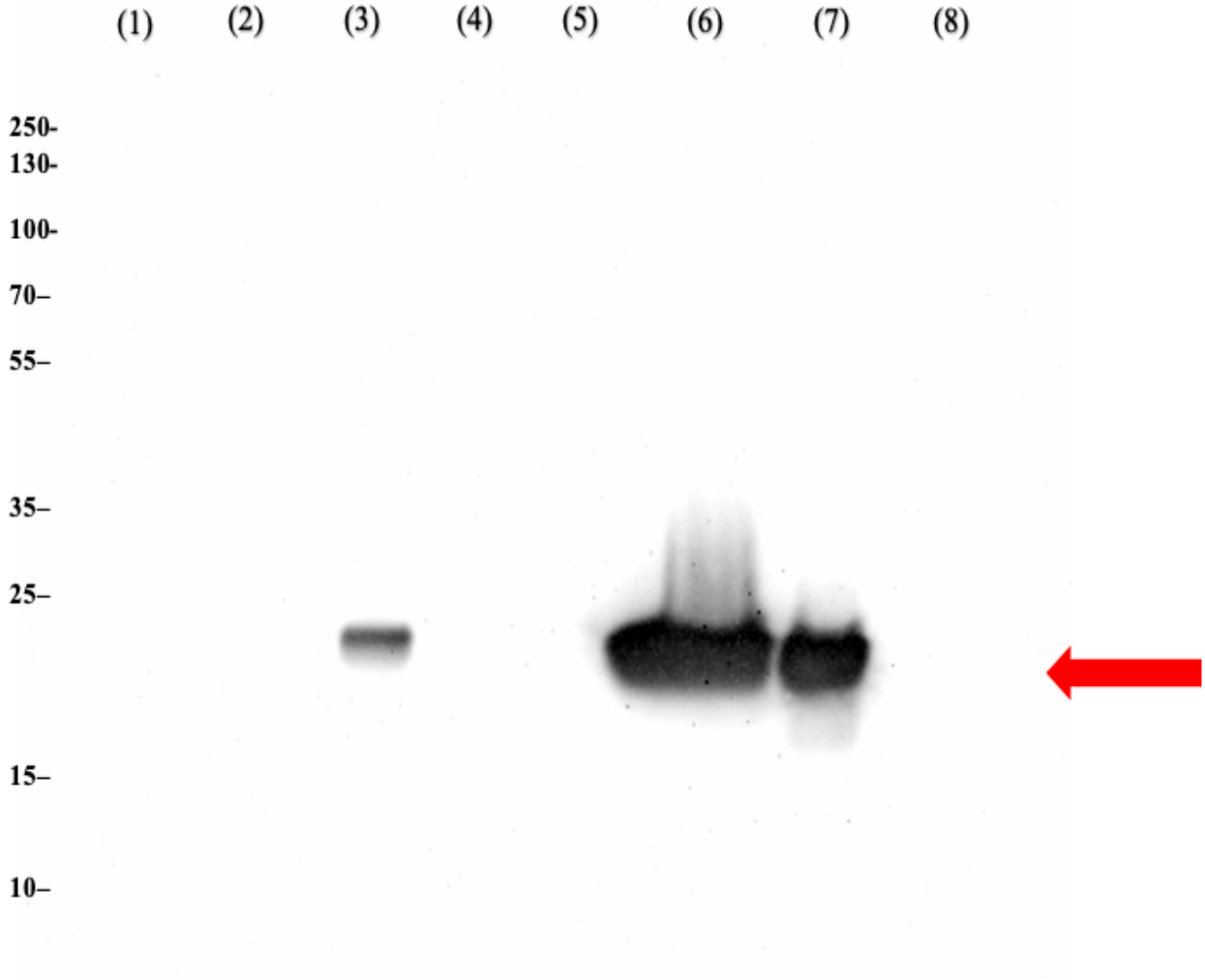


Figure 3.19 Western blot of a PVDF membrane with different fraction of purification process of recombinant EAD of LysCD6356 including two elution fractions (200mM and 250mM imidazole).

The membrane was stained with a histidine specific antibody at a concentration of 1:10,000 and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (1) and (8) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder), (2) Pre-induction fraction, (3) post-induction fraction, (4) Flow through fraction, (5) Wash fraction, (6) 1st Elution fraction (200mM imidazole) (7) 2nd Elution fraction (250mM imidazole). Red arrow indicated the predicted size of the recombinant EAD of LysCD6356 band (22.12 kDa).

Using the modified purification method, the final protein yield of monomeric EAD was approximately 275 µg/L, which was slightly lower in yield than the initial purification method (300 µg/L).

3.3.15 Enhancing the initial expression of the EAD of LysCD6356

In an attempt to enhance the protein yield of the EAD of Lys CD6356, we explored the impact of increasing the expression time post IPTG induction from 4 to 10 hours. As a result, the final protein yield of the EAD of LysCD6356 increased approximately 10 folds from 275 to 2,700 µg/L with a purity of 100% as determined by Bio-Rad software analysis (Figure 3.20).

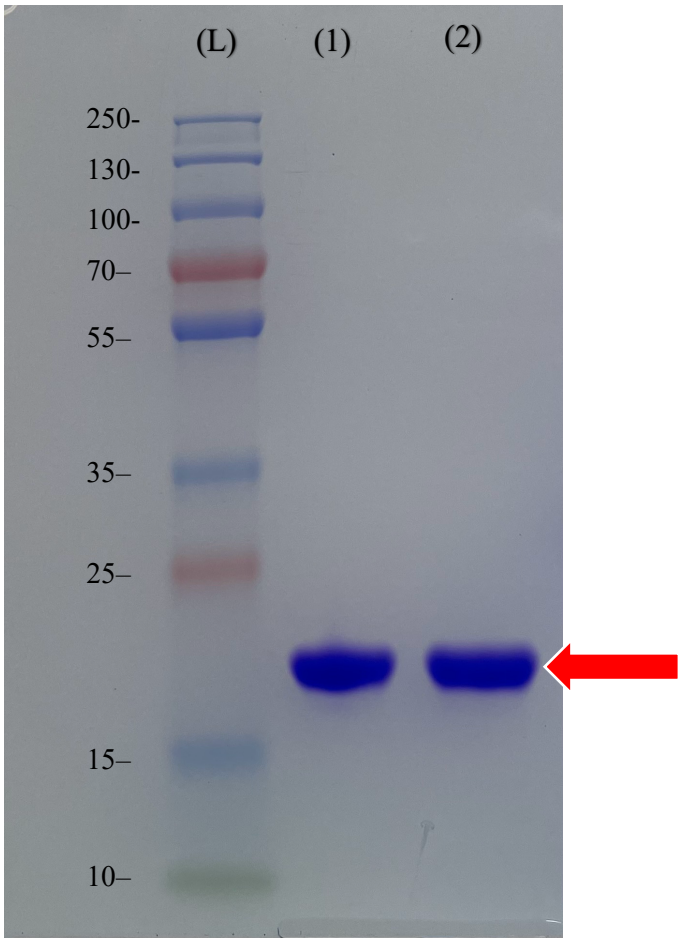


Figure 3.20 An SDS-PAGE gel showing elution fractions following increasing the induction time of a culture of *E. coli* BL21(DE3)pLysS pET-19b EAD of LysCD6356 expressing recombinant EAD of LysCD6356.

The gel was stained with Coomassie brilliant blue to visualize protein bands. (1) 1st elution fraction (200 mM imidazole), (2) 2nd elution fraction (250mM imidazole), (L) molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant EAD of LysCD6356 (22.12 kDa).

3.3.16 Enhancing the purity and the yield of full length LysCD6356

Following the same expression and purification methods used to enhance the purity and recovery of EAD, the full length protein was produced and purified, as demonstrated in Figure 3.21.

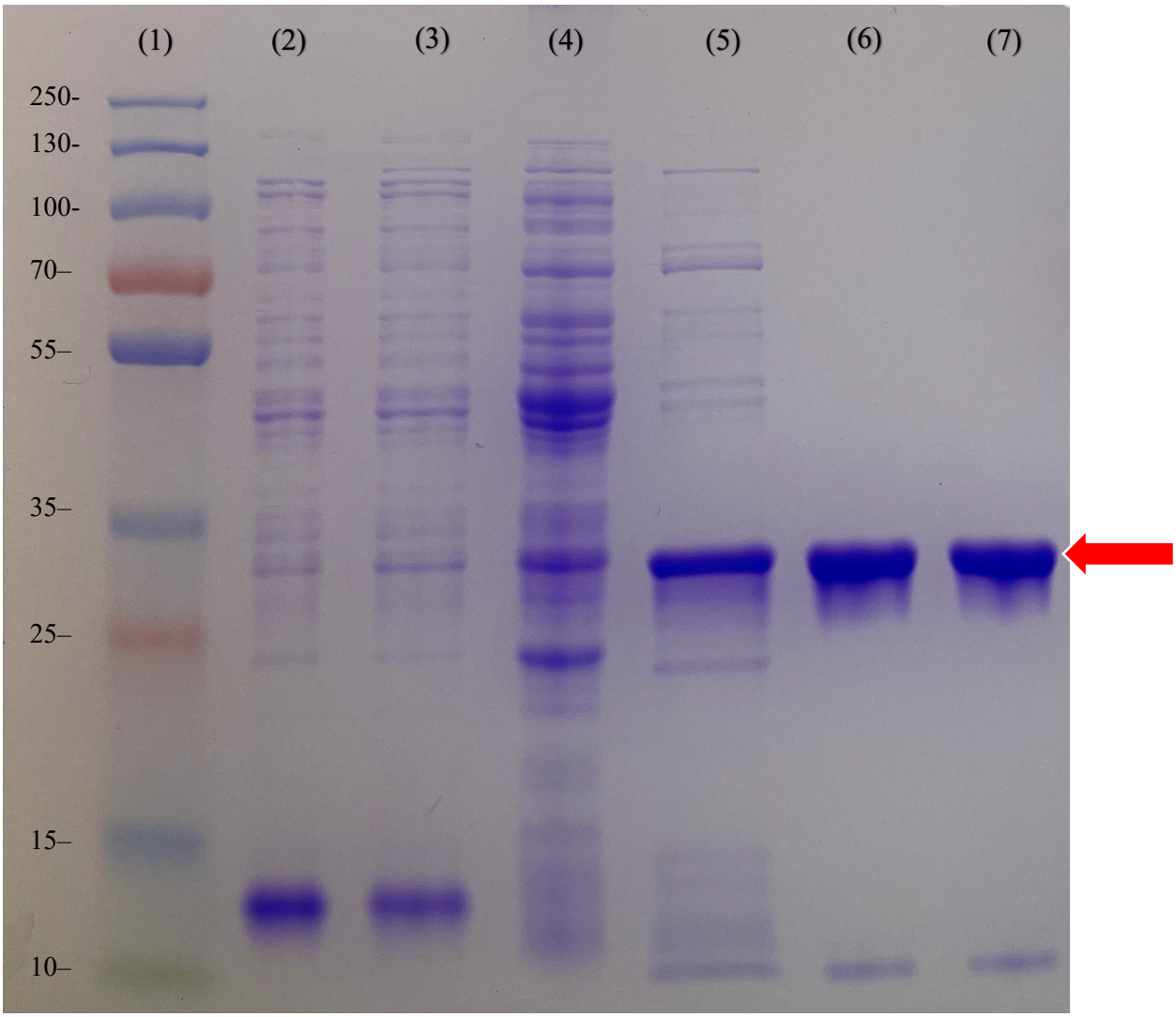


Figure 3.21 An SDS-PAGE gel showing different fraction of *E. coli* BL21(DE3)pLysS pET-19b LysCD6356 expressing recombinant LysCD6356 following using of DTT as reducing agent.

The gel was stained with Coomassie brilliant blue to visualize protein bands. (2) Pre-induction fraction, (3) post-induction fraction, (4) Flow through fraction, (5) Wash fraction, (6) and (7) elution fraction (250mM imidazole), (1) molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant LysCD6356 (32.67 kDa).

To investigate the nature of the low molecular weight band (around 10 kDa), a western blot was performed, as shown in Figure 3.22. Only one high intensity band was seen, which was similar in size to full length LysCD6356 (32.67 kDa).

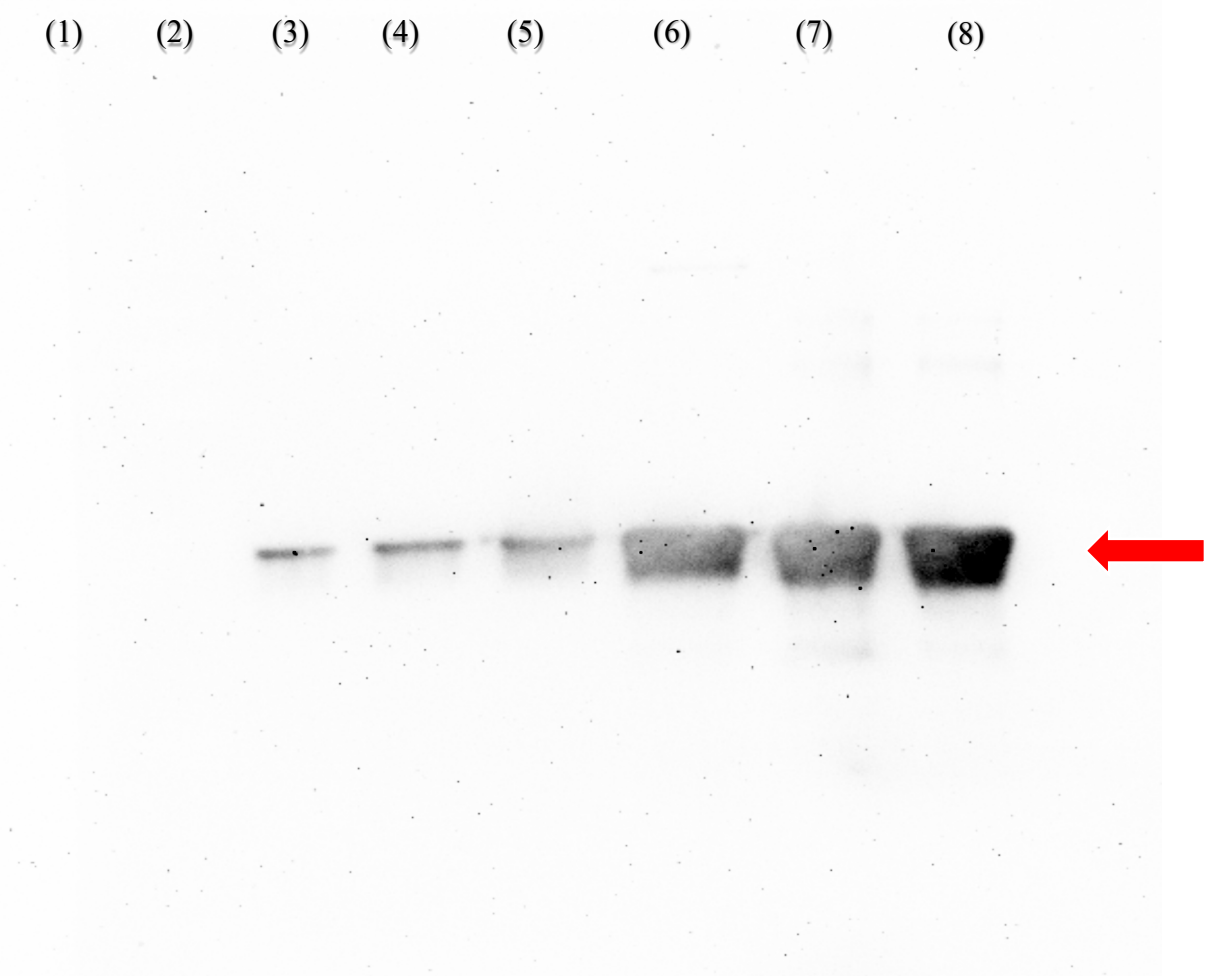


Figure 3.22 Western blot of a PVDF membrane with different fractions of purification process of recombinant LysCD6356.

The membrane was stained with a histidine specific antibody at a concentration of 1:10,000 and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (2) Pre-induction fraction, (3) and (4) post-induction fraction, (5) Flow through fraction, (6) Wash fraction, (7) and (8) elution fraction (250mM imidazole), (1) molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant LysCD6356 (32.67 kDa).

Increasing the expression time following IPTG from 4 to 10 hours when combined with the modified purification method increased the protein yield of LysCD6356 more than 14-fold from 350 µg/L to approximately 5000 µg/L, with the purity increasing from 81% to 94.85% as determined by Bio-Rad software analysis (Figure 3.23).

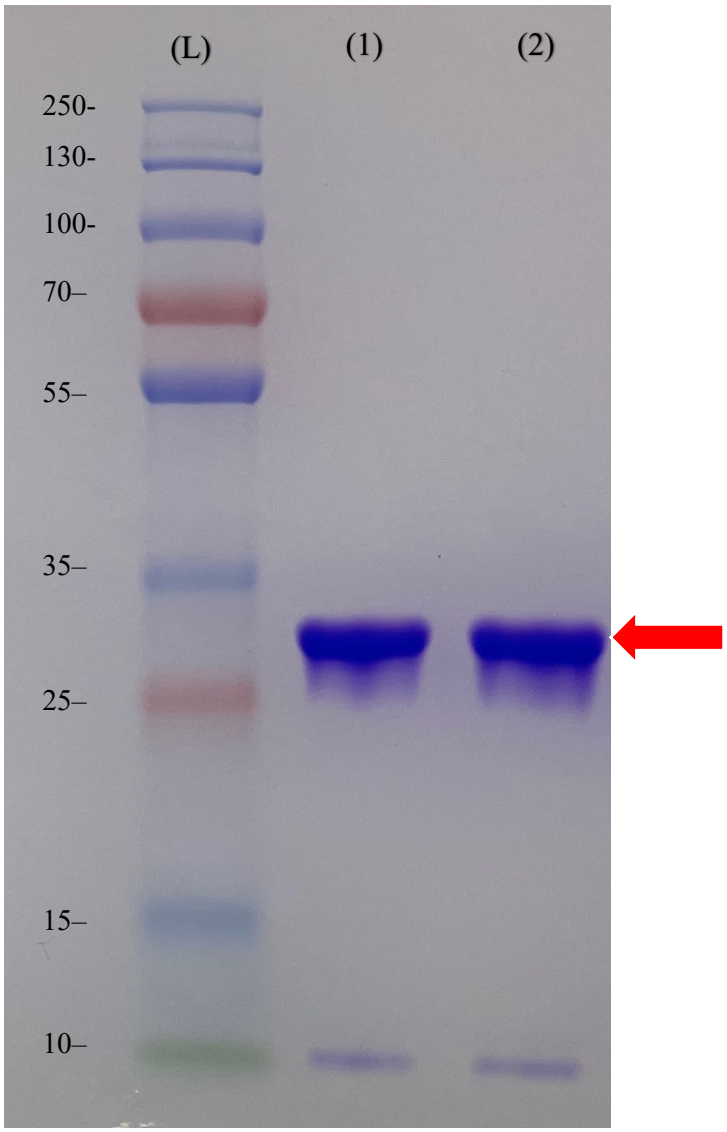


Figure 3.23 An SDS-PAGE gel showing elution fractions following increasing the induction time of a culture of *E. coli* BL21(DE3)pLysS pET-19b LysCD6356 expressing recombinant LysCD6356

The gel was stained with Coomassie brilliant blue to visualize protein bands. (1) 1st elution fraction (150mM imidazole), (2) 2nd elution fraction (250mM imidazole), (L) molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). Red arrow indicated the predicted size of the recombinant LysCD6356 (32.67 kDa).

3.4 Discussion

The level of recombinant protein expression is impacted by a wide range of factors, including the characteristics of the host strain, the codon usage of the gene sequence encoding the protein, and the biological properties of the expressed protein. In this study, we successfully expressed codon optimized versions of LysCD6356 and its EAD from *E. coli* BL21(DE3)pLysS. This strain was selected because it has previously been used to express endolysins from a range of sources (Antonova et al., 2019; Żebrowska et al., 2022).

Using an induction protocol developed in an earlier study, we optimized the expression of LysCD6356 and the EAD from our *E. coli* host (Alyousef 2013). Inclusion of a sonication step in our host cell lysis protocol, combined with modifications to the expression and purification protocols, allowed us to maximize the level of protein recovery.

Employing the optimized expression and purification protocol developed in this study, we were able to produce the monomeric form of the full length recombinant LysCD6356 at a yield of 5000 μg/L and a purity of 94.85%. In the case of EAD, the monomeric yield was 2700 μg/L, and the purity was 100%.

These levels were in line with those previously reported in the literature. Alyousef was able to produce 15 mg/L of highly purified LysCD6356 (Alyousef 2013). Furlon and colleagues reported a yield of 1.5 mg/L of 90% pure EAD (Furlon et al., 2021). While these studies employed gene sequences codon optmized by commercial companies to maximize expression in *E. coli* the algorithms employed are not publicly available, meaning that small variations in the codons selected can have a big impact on the level of protein synethesis (Mauro 2018). The codon optmized genes were also expressed from different *E. coli* hosts, in the Alyousef study, they used *E. coli* strain M15, while Furlon and colleagues expressed EAD from an unspecified *E. coli* host (Alyousef 2013; Furlon et al., 2021).

A considerable amount of effort was devoted to producing recombinant protein preparations free of host derived proteins. The standard packed Ni-NTA column approach proved to be less efficient

that allowing the protein extracts to mix with a suspension of the suspended Ni-NTA resin (Qiagen 2023). The higher protein attachment probably results from the longer contact period with the resin (1 hour) compared to that seen using a flow through column. Additionally, having a long histidine tag (10 histidine residues) allows for the use of harsher washing steps (Grisshammer and Tucker 1997). Consequently, using several washing steps with a higher imidazole concentration (50 mM) in the modified purification method helped to eliminate weakly bound contaminants better than 20 mM imidazole that was used in the initial purification method.

The presence of additional protein bands recognized by the histidine tagged antibody following the initial purification method in our Ni-NTA column elution preparations suggests the presence of multimeric forms of the recombinant lysin that were resistant to SDS degradation.

It has been proposed that lysins such as CD27L form dimers, which enable them to bind to the cell wall of the target bacteria, after which the protein undergoes autocleavage to release the catalytic domain, which allows it to further penetrate the bacterial cell wall, which may act as a sieve (Dunne et al., 2014). Recently, it has been proposed that binding to the cell wall reduces enzyme diffusion and thereby limits cell lysis of the neighbouring bacteria, providing new hosts for the newly released bacteriophage progeny (Phothichaisri et al., 2022).

Increasing the recombinant protein expression period from 4 to 10 hours post induction and the development of a more efficient purification protocol yield sufficient quantities of protein to support future experimental studies.

The ultimate proof that we have successfully expressed recombinant lysins is to demonstrate their ability to lyse vegetative isolates of *C. difficile.* In the next chapter, we will determine the ability of our recombinant proteins, expressed and purified as described in this chapter, to lyse a diverse range of clinical isolates of *C. difficile*.

Chapter 4 An evaluation of the viability of utilizing spore germinants and bacteriophage endolysin for the treatment of recurrent *C. difficile* infection

4.1 Introduction

Clostridioides difficile (*C. difficile*) is a Gram-positive, anaerobic, spore-forming bacillus that causes healthcare-associated infections. The vegetative form of *C. difficile* produces toxins TcdA and TcdB, causing disease, while its spores facilitate spread and recurrence (Voth and Ballard 2005). Hypervirulent strains, like BI/NAP1/027, have caused recent epidemics around the world (Dawson et al., 2011; Frieden 2013; Office for National Statistics 2014). Infections can range from mild diarrhea to fatal pseudomembranous colitis, often triggered by antibiotics disrupting gut microflora. Effective treatments including Vancomycin and Fidaxomicin leading to high recurrence rates as a result of resistant spores. Fecal transplants can restore protective microflora and reduce recurrent CDI. New approaches, such as *C. difficile*-specific bacteriophages and recombinantly expressed bacteriophage-derived endolysins, active agaisnt vegetative bacteria and antibiotic resistant isolates while sparing gut flora (Romero-Calle et al., 2019). Endolysins, have shown effectiveness in reducing *C. difficile* colonization in animal models (Peng et al., 2019; Wang et al., 2015). Despite these advances, eradicating spores remain a challenge for preventing recurrent infections. Interestingly, studies have shown that a combination of spore germinants and a biocidal agent can be used to clean environments contaminated with spore forming bacteria such as *B. anthracis* and *C. difficile* (Celebi et al., 2016a; Nerandzic and Donskey 2016; Nerandzic and Donskey 2017).

4.2 Aims and objectives

In this chapter, the ability of the bacteriophage derived endolysin LysCD6356, and its EAD to target clinical isolates obtained primarily in the UK will be assessed. This lysin, which was isolated from φCD6356, a temperate phage of the *Siphoviridae* family, was previously shown in our hands to be more active than LysCD27 against our collection of isolates (Horgan et al., 2010; Alyousef 2013).

We will undertake an assessment of the ability of EAD in combination with the germinant taurocholate and cogerminants glycine, and calcium to target the spore form of the pathogen with a view to developing a treatment regimen that minimizes the potential for recurrent infections (Budi et al., 2020).

The **objectives** were to:

1. Assess the biological activity of the recombinant LyCD6356, full length lysin, and EAD.
2. Determine the activity of the lysins against a panel of *C. difficile* clinical isolates.
3. Characterize the impact of pH, metal ions, and calcium chloride concentration on the lytic activity of the EAD of LysCD6356.
4. Investigate the impact of spore germinants and calcium on the lysis of vegetative *C. difficile.*
5. Identify a germination / lysin delivery strategy that maximizes the killing of *C. difficile* spores.

4.3 Results

4.3.1 Culture and characterization of vegetative *C. difficile.*

To confirm the purity of our stocks of *C. difficile,* individual isolates were cultured on sheep blood agar plates and incubated at 37 ºC under aerobic and anaerobic conditions. *C. difficile* is a strict anaerobe, and colonies are grayish white in colour, irregular in shape, and range from 4 to 6 mm in diameter (Figure 4.1).

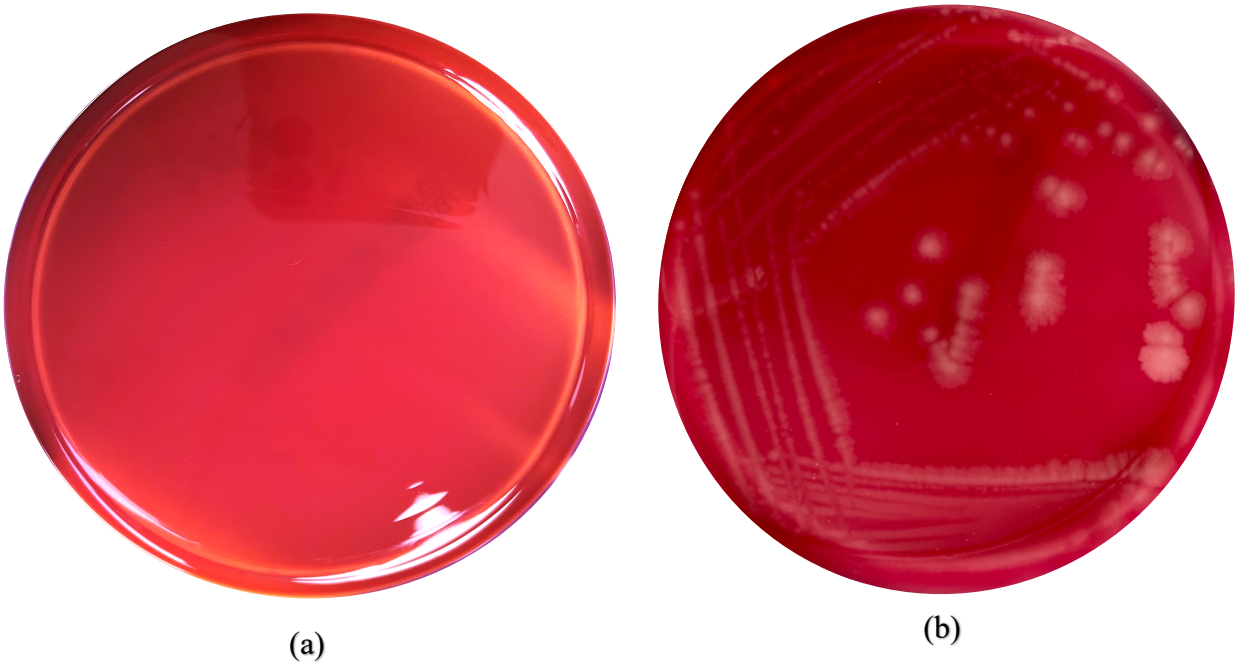
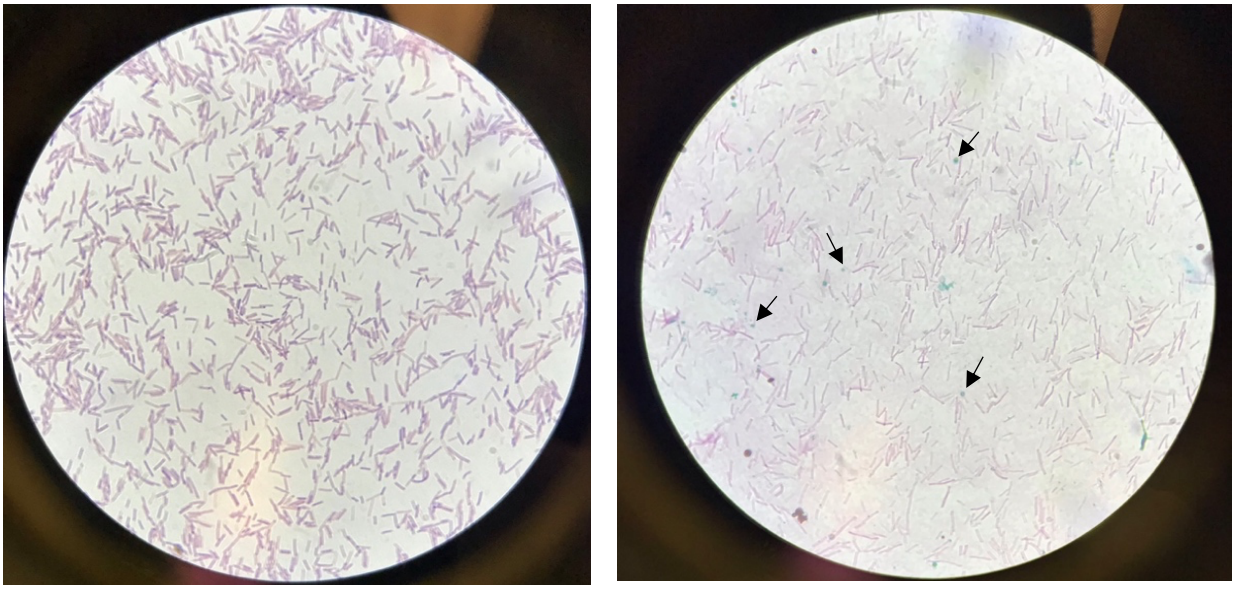


Figure 4.1 The growth of *C. difficile* on sheep blood agar.

C. difficile NCTC 12727 was incubated aerobically (a) and anaerobically (b) on sheep blood agar for 48 hours at 37°C.

Gram and spore stains of material from characteristic colonies confirmed the presence of spore forming Gram-positive bacteria with a rod-shaped appearance (Figure 4.2).



(a) (b)

Figure 4.2 Gram and spore stains of *C. difficile* cultures.

A sample of each bacterial culture was subjected to Gram (a) and Spore (b) staining after which they were observed under light microscope using oil immersion lens (100x) (Leica DM750, Leica Microsystems, UK). Arrows indicate the presence of green staining spores.

4.3.2 Confirmation of the biological activity of the lysins

The turbidity reduction assay, which measures changes in optical density, is widely used to determine the activity of endolysins. It is based on the premise that a reduction in OD_{600} is due to disruption of the cell wall of the bacterium, which results in the release of cellular contents (Nelson et al., 2012). This assay was employed to characterize the activity of the recombinant forms of LysCD6356 and EAD expressed in Chapter 3.

We first identified a working concentration of recombinant protein to support future studies by testing different concentrations (7.5 and 60 µg/mL) of full length lysin and the EAD for lytic activity against three clinical isolates, DS1813 (Ribotype 027), R20291 (Ribotype 027), and DS1787 (Ribotype 106) (Figure 4.3).

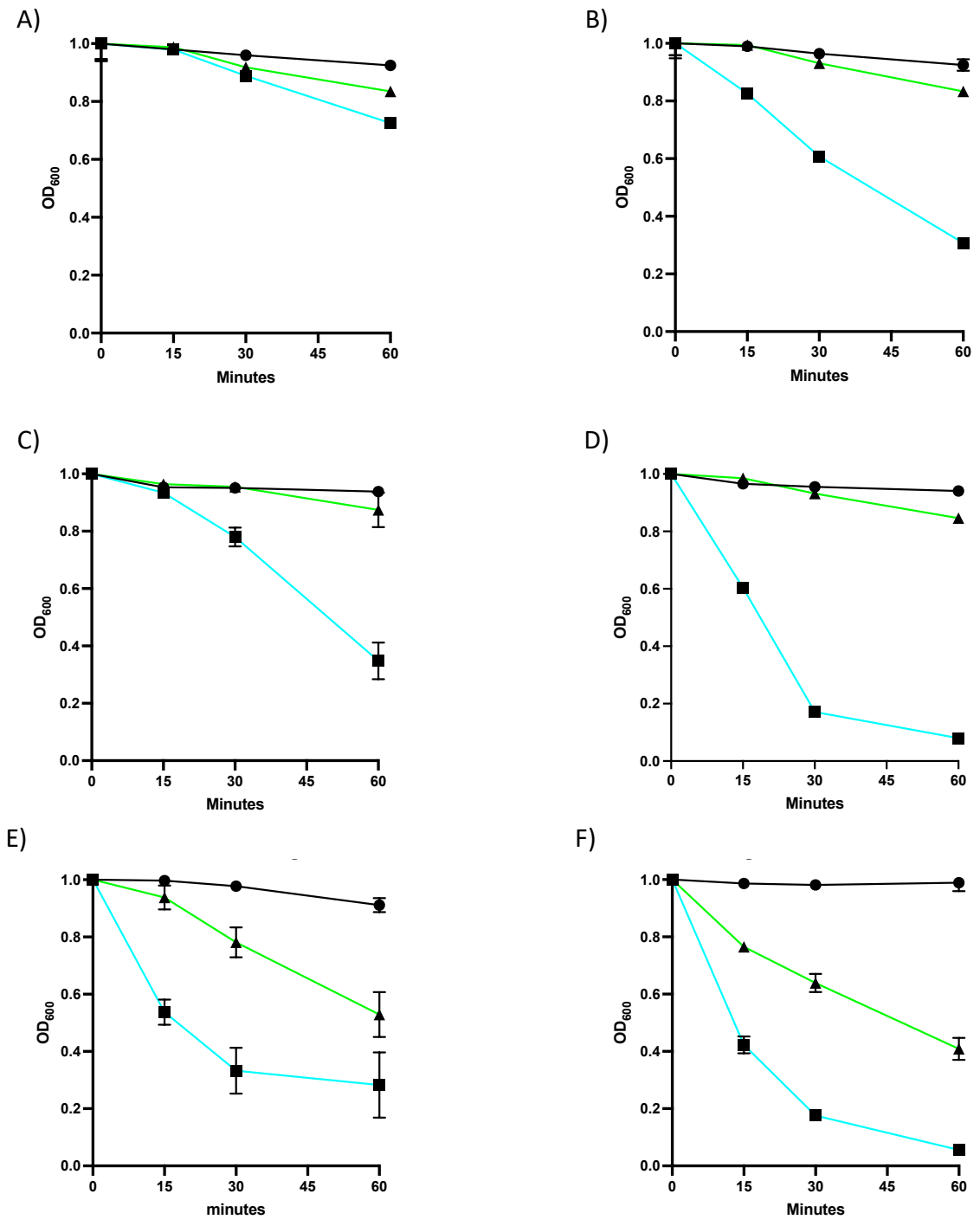


Figure 4.3 Lytic activity of different concentrations of LysCD6356 and its EAD against the vegetative forms of *C. difficile* DS1787, DS1813 and R20291.

C. difficile strains DS1787 (A and B), DS1813 (C and D) and R20291 (E and F) were cultured in BHI broth to mid log phase and resuspended in PBS (pH 7.4) (black line). LysCD6356 (green line) or EAD (blue line) was added to the suspension which was incubated at 37°C for 60 mins. Different concentrations of both endolysins were assessed, where left graphs represent 7.5µg/mL and right graphs represent 60 µg/mL of each endolysin. Changes in OD_{600} were recorded using a TECAN infinite F200 pro plate reader. Experiments were performed in triplicate.

Even though the current experiment demonstrated that EAD is more active than full length endolysin on a mass basis, it does not determine the difference in activity on a molar basis. By comparing the activity against DS1813, R20291 and DS1787, DS1787 was the least sensitive strain. Even though R20291, and DS1813 are from the same ribotype (027), the activity has slight differences, especially at low endolysin concentration (7.5 μg/mL). To determine the difference in activity and to have better representative data, endolysins need to be used in unified molar concentration and to be assessed against a wider range of ribotypes.

4.3.3 Characterizing the activity of EAD against vegetative *C. difficile* using SEM

To visualize the impact of EAD on the cell wall of vegetative bacteria, we examined EAD treated samples of R20291 using SEM. This strain is a hypervirulent 027 ribotype strain isolated from a major outbreak in the UK, which has been characterized across a number of studies. As can be seen in Figure 4.4, the degree of disruption to the cell surface increased with time following the bacteria incubation with the lysin. At 15 minutes of incubation, bumps like structures can be observed on the surface of individual bacteria. At 60 minutes, individual bacteria had lost their shape and had contracted into irregular balls, suggesting a major loss of cellular content.

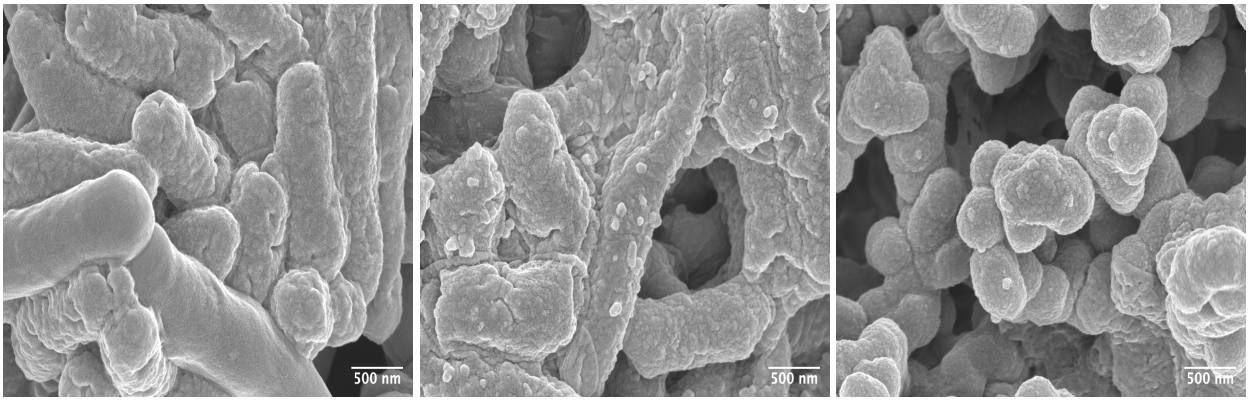


Figure 4.4 SEM images of vegetative *C. difficile* R20291 incubated in the presence of the EAD of LysCD6356.

Vegetative *C. difficile* strain R20291 was incubated in BHI broth to mid log phase after which cells were resuspended in SDW. (a) no EAD, (b) 60 μg/mL EAD for 15 minutes and (c) 60 μg/mL EAD for 30 minutes. Images are X70,000 magnification.

4.3.4 Enzymic Activity of CD6356 and its EAD Fragment Against a Panel of Vegetative Isolates of *C. difficile* Representing Diverse Ribotypes

Once we had confirmed the biological activity of the recombinantly expressed endolysins against R20291, DS1787, and DS1813 strains, we next used the turbidity assay to determine their activity against a diverse panel of clinical isolates following a 60-minute incubation period. The lytic activity of the full length lysin varied depending on the test strain (Figure 4.5).

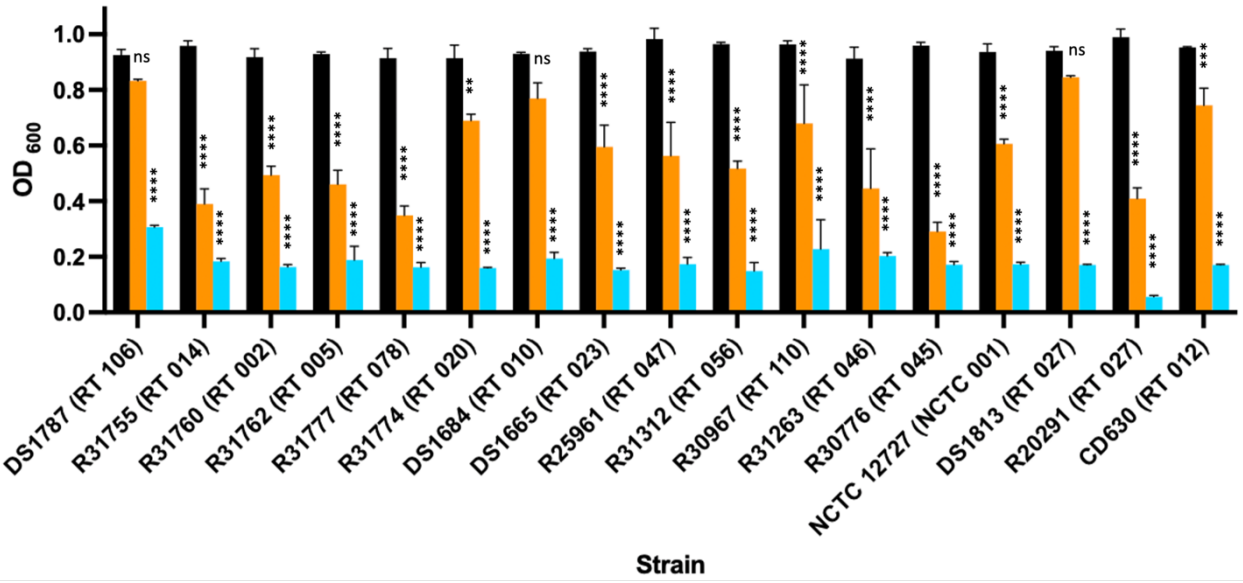


Figure 4.5 Lytic activity of LysCD6356 and its EAD fragment against a collection of clinical isolates of *C. difficile*.

Vegetative forms of the test bacteria were suspended in PBS to an OD_{600} of ~1.0 and incubated with either full length lysin (orange bar) or EAD (blue bar) at a concentration of 60µg/mL for 60 minutes at 37° C at which time the OD_{600} was recorded. Suspended bacteria without treatment was used as a control (black bar). Experiments were performed in triplicate. Error bars represent the standard deviation. P value = < 0.05(*), < 0.01(**), < 0.001 (***), < 0.0001 (****).

Previous studies reported that EADs are more active than the full length lysin (Mayer et al., 2011; Wang et al., 2015). We were able to confirm that EAD was more active than full length endolysin on a mass basis: however, the difference in activity cannot be confirmed without investigating the molar basis.

4.3.5 Effect of pH on Lytic Activity

We determined the ability of the EAD fragment to lyse vegetative bacteria across a range of pH values, as pH is known to affect the activity of enzymes, such as *C. difficile* specific phage derived recombinant lysins (Table 4.1) (Mayer et al., 2011; Alyousef 2013; Wang et al., 2015; Peng et al., 2019; Mondal et al., 2021).

Table 4.1 Comparison on the activity of published endolysins in different pH levels.

Endolysin	Source	pH range	Activity	Reference
PlyCD	φCD630	pH 4 to 8	Active in pH 6, 7 and 8.	(Wang et al., 2015)
$PlyCD_{1-174}$			No activity below pH 5	
LysCD6356	φCD6356	pH 4 to 9	Almost lost activity at pH 9	(Alyousef 2013)
CD27L	φCD27	pH 4.5 to 8.3	Active in all pH range tested	(Mayer et al., 2008)
$CWH_{351-656}$	φMMP01	pH 4 to 8	Active in all pH range tested, pH7 was optimal	(Mondal et al., 2021)
LHD	φC2+HD5	pH 6 to 8	Active in all pH with delay in activity in pH 8	(Peng et al., 2019)

Similar to CD27L, the EAD fragment retained activity against the vegetative form of R20291 across a pH range of 4-9 (Figure 4.6).

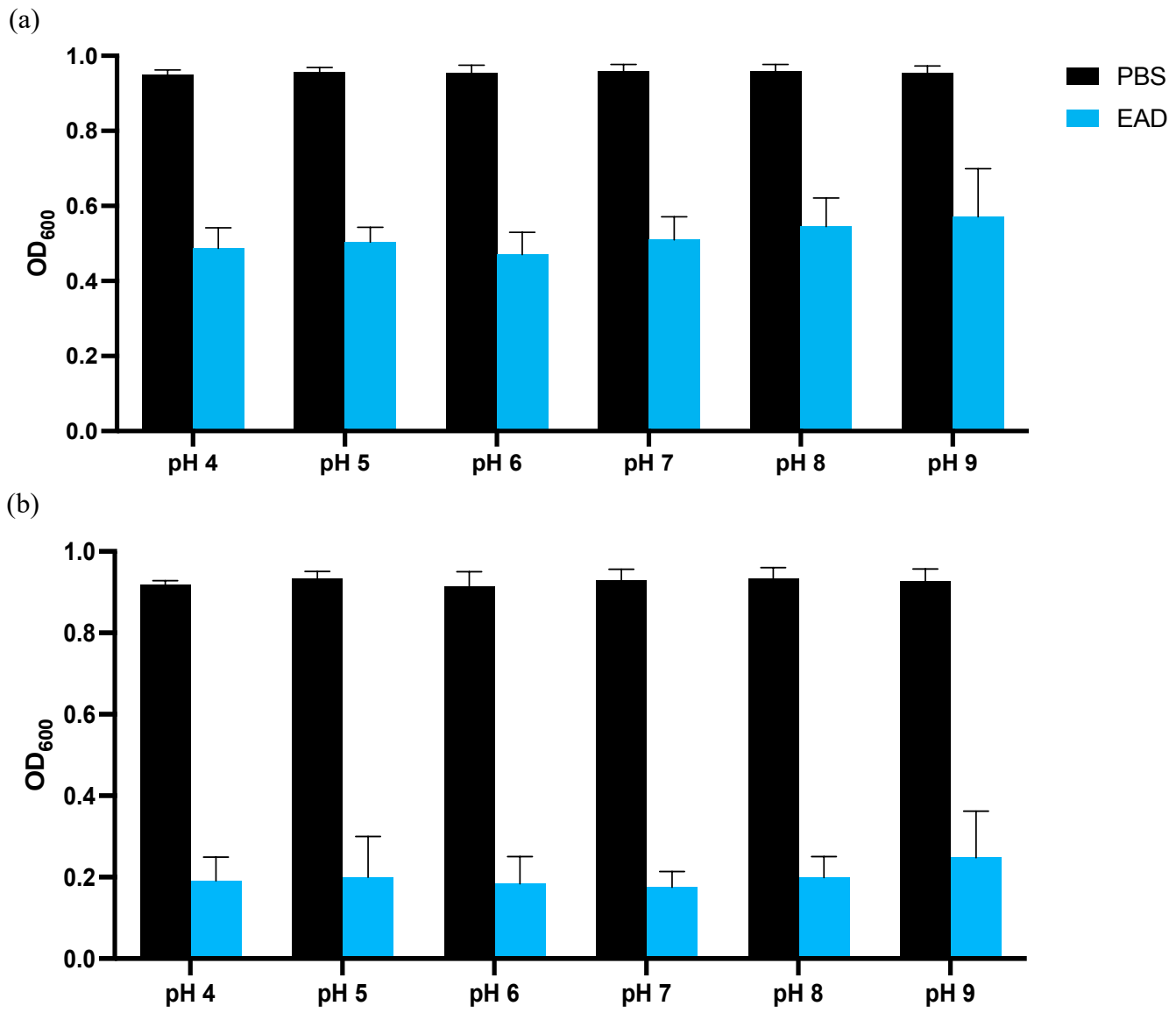


Figure 4.6 The effect of pH on the lytic activity of EAD against vegetative *C. difficile* strain R20291.

C. difficile strain R20291 was grown in BHI broth to mid-log phase. Pellet was then resuspended in PBS with pre-adjusted pH ranging from 4 to 9, with and without 60µg/mL EAD and incubated for (a) 30 and (b) 60 minutes. Each experiment was performed in triplicate.

4.3.6 The contribution of divalent cations to the enzymic activity of EAD.

To investigate the contribution of cations to the enzymic activity of EAD, the protein was incubated with the chelating agent, EDTA, for 30 minutes prior to testing for lytic activity against R20291. As can be

seen in Figure 4.7, the catalytic activity of EAD was lost, indicating the importance of divalent cations in the functionality of the lysin.

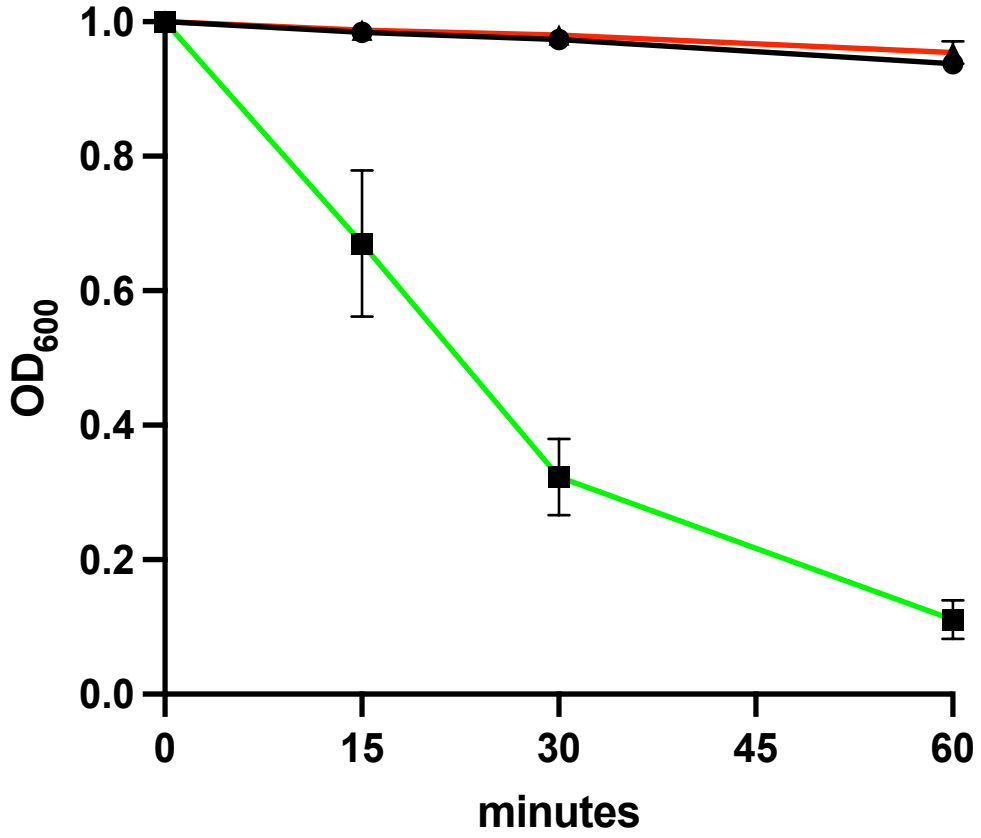


Figure 4.7 The lytic activity of EAD against vegetative *C. difficile* R20291 following treatment of the protein with EDTA.

Following the incubation of EAD with EDTA, the chelating agent was removed using Pierce™ protein concentrator with a 10,000 molecular weight cut off. 60µg/mL of EDTA treated EAD was added to vegetative *C. difficile* strain R20291 (red line) and the changes in OD_{600} recorded throughout 60 minutes. Vegetative R20291 (black line) and vegetative R20291 mixed with 60µg/mL of untreated EAD (green line) were included as controls. Experiments was performed in triplicate. Changes of OD_{600} were normalized by the initial value.

Next, the ability to restore the lytic activity was investigated by the addition of the metal ions Mn^{2+}, Mg^{2+}, and Zn^{2+} to EDTA treated EAD. The addition of Mn^{2+} or Mg^{2+} failed to restore enzymic activity (Figures 4.8 and 4.9).

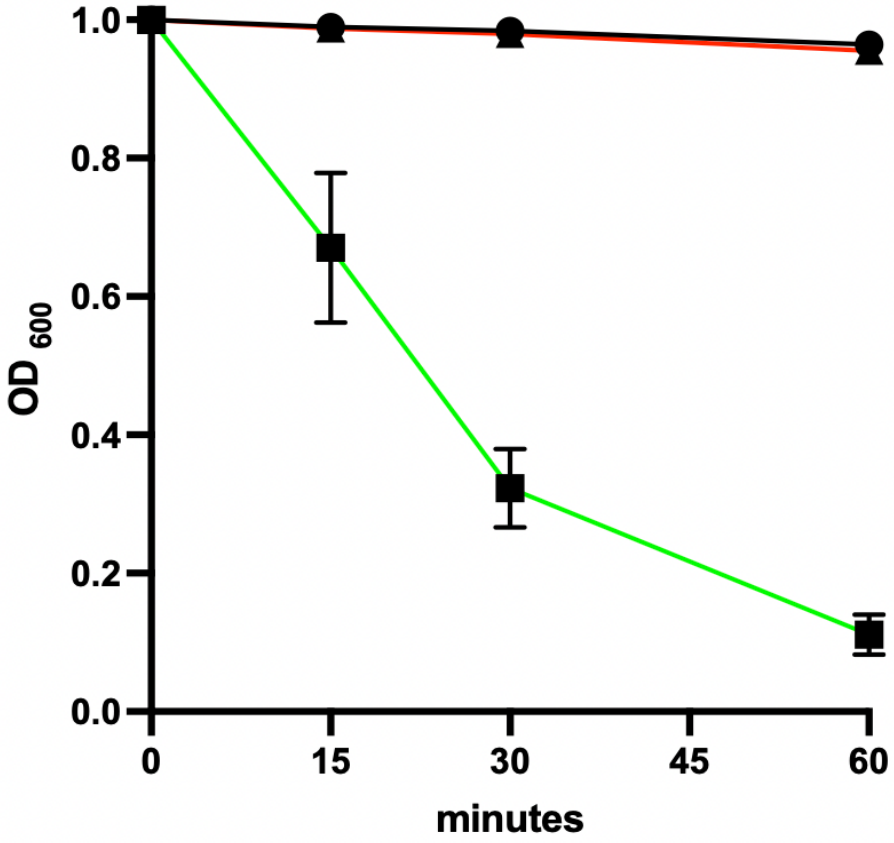


Figure 4.8 Investigating the ability of Manganese to restore the activity of EDTA treated EAD to lyse vegetative R20291.

EDTA treated EAD (60 µg/mL) was added to vegetative *C. difficile* strain R20291 in the presence of 1.0 mM Mn^{2+} solution (red line). Vegetative R20291 in 1.0 mM Mn^{2+} solution (black line) and vegetative R20291 with 60µg/mL of untreated EAD (green line) were included as controls. Experiments were performed in triplicate. Changes of OD_{600} were normalized by the initial value.

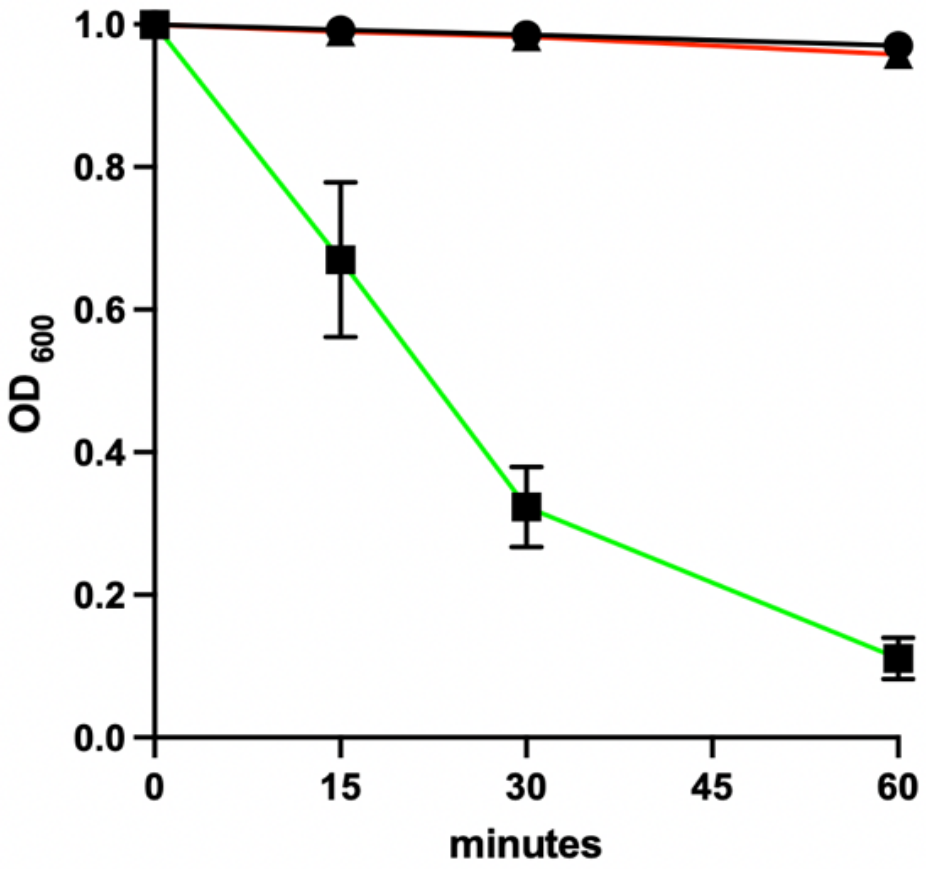


Figure 4.9 investigating the ability of Magnesium to restore the activity of EDTA treated EAD to lyse vegetative R20291.

EDTA treated EAD (60 µg/mL) was added to vegetative *C. difficile* strain R20291 in the presence of 1.0 mM Mg^{2+} solution (red line). Vegetative R20291 in 1.0mM Mg^{2+} solution (black line) and vegetative R20291 with 60 µg/mL of untreated EAD (green line) were included as controls. Experiments were performed in triplicate. Changes of OD_{600} were normalized by the initial value.

In the presence of Zn^{2+}, we observe the restoration of a low level of lytic activity, suggesting EAD might be a zinc dependent enzyme (Figures 4.10).

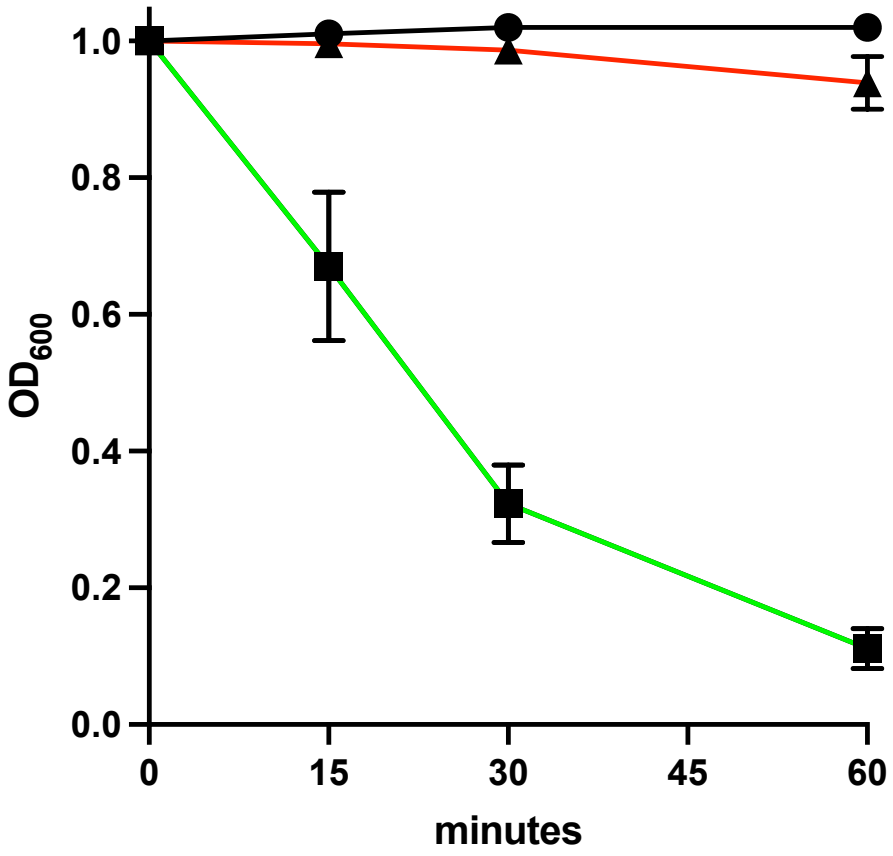


Figure 4.10 Investigating the ability of Zinc to restore the activity of EDTA treated EAD against vegetative R20291 throughout 60 minutes.

EDTA treated EAD (60 µg/mL) was added to vegetative *C. difficile* strain R20291 in the presence of 1.0 mM Zn^{2+} solution (red line). Vegetative R20291 in 1.0mM Zn^{2+} solution (black line) and vegetative R20291 with 60 µg/mL of untreated EAD (green line) were included as controls. Experiments were performed in triplicate. Changes of OD_{600} were normalized by the initial value.

To determine if the restorative effect of zinc on lytic activity was time dependent, we increased the incubation period of the experiment to 150 minutes. As can be seen from the Figure below, extending the incubation period enhanced lytic activity, suggesting that EAD is a zinc dependent enzyme (Figure 4.11). It is interesting to note that $CD27L_{1-179}$, the catalytic domain of the CD27L endolysin, which is being investigated as a potential treatment for CDI, only shares 37.4% amino acid sequence homology with the EAD of LysCD6356 and is also zinc dependent N-acetylmuramoyl-L-alanine amidase (Mayer et al., 2011).

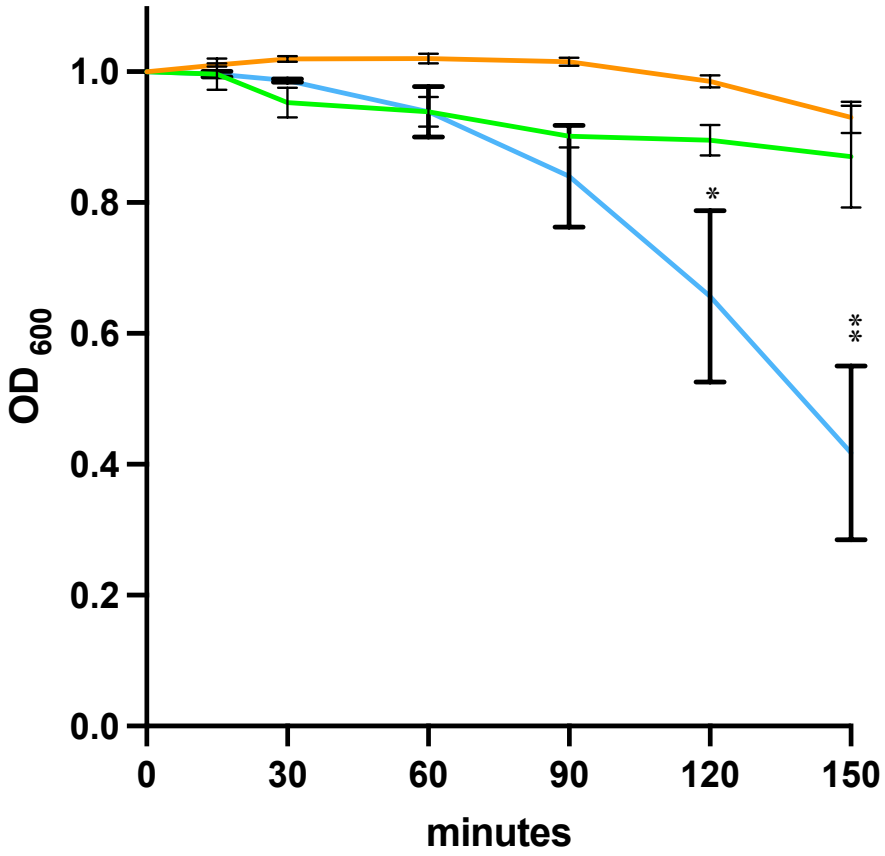


Figure 4.11 Restoration of the lytic activity of EDTA treated EAD by Zn^{2+} ions.
The ability of Zn^{2+} to restore the lytic activity of 60 µg/mL of EDTA treated recombinant EAD following the addition of 1mM Zn^{2+} solution was assessed against vegetative R20291 (blue line). As a control, vegetative R20291 was suspended with EDTA (green line) and with Zn^{2+} (orange line). Experiments were performed in triplicate. Changes of OD_{600} were normalized by the initial value. Statistical analysis was performed for treated and untreated groups. P value = < 0.05(*), < 0.01(**).

4.3.7 The impact of calcium on the enzymic activity of the EAD of LysCD6356.

It has been reported that calcium can adversely affect the lytic activity of lysins such as PlyCD, which shares 77.5% sequence homology with the EAD of LysCD6356 and is currently being developed as a treatment for CDI (Furlon et al., 2021). This is important as calcium is present in the gut as part of the normal diet, with an estimated average concentration of 40 mM passing through the gut daily (Barrett et al., 2015). Estimating the calcium concentration at a particular point in time is a challenge, and for this reason, we have selected a concentration of 12.5 mM calcium, which was based on the level detected in the ileum of a mouse to represent the baseline gut concentration (Kochan et al., 2017). Additional calcium is released from *C. difficile* spores upon the triggering of germination (Wang et al., 2011).

To determine if calcium inhibited the activity of the EAD of LysCD6356, we incubated EAD with increasing concentrations of calcium chloride and assessed the lysing ability to lyse the vegetative form of R20291 following 60 minutes of incubation at 37 ºC. The levels of calcium chloride used in this experiment are based on those used to assess the impact of calcium chloride on the activity of PlyCD (Wang et al., 2015).

As can be seen from Figure 4.12, as the concentration of calcium chloride increased, there was a corresponding reduction in lytic activity. Interestingly, we still saw activity at concentrations above 25 mM (50 and 100 mM), which are likely to be in excess of the levels of the local concentration of calcium seen in the gut. Additionally, by comparing the reduction of different treatments, there was no significant differences between 50 mM and 12.5 mM suggesting that the reduction of EAD activity as a result of increasing calcium concentrations was minimal after 12.5 mM even at concentrations considered excessive to the local calcium concentration (50 mM).

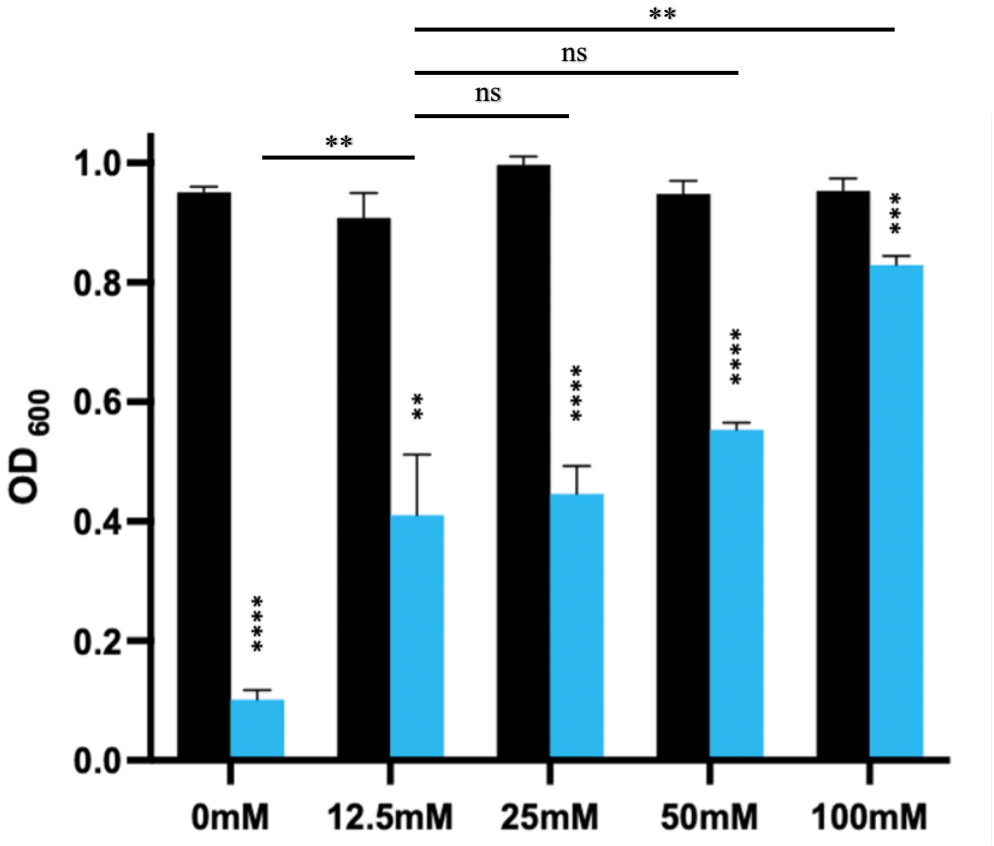


Figure 4.12 The lytic activity of 60 µg/mL EAD against the vegetative form of *C. difficile* R20291 in the presence of increasing concentrations of calcium chloride following 60 mins of incubation.
Vegetative *C. difficile* strain R20291 was used to assess the activity of 60µg/mL EAD of LysCD6356 in the presence of different concentrations of calcium chloride (pH 7 and 37° C). The effect on lysis was determined by measuring the change of the OD_{600} in the presence (blue) and absence (black) of EAD in

different calcium concentrations. Experiment was performed in triplicate**.** P value = <0.05(*), <0.01(**), <0.001 (***), <0.0001 (****).

4.3.8 Identification of compounds which trigger the germination of *C. difficile* spores

Given that the primary aim of this project is to develop a lysin based therapy that targets the spore form of *C. difficile*, we first need to identify the optimum germination mixture for R20291 spores. Studies have reported that the primary bile salt, sodium taurocholate, in combination with the amino acid co-germinant, glycine, are effective germinants for *C. difficile* spores (Sorg and Sonenshein 2008). Concentration levels documented in the literature vary as to the optimal sodium taurocholate (Tc) concentration, and for this reason, we investigated the impact of different concentrations of taurocholate (0.1, 0.2%, 1%, 2%, and 4%) in the presence of a fixed concentration of glycine (50 mM) on the germination efficiency of R20291 spores.

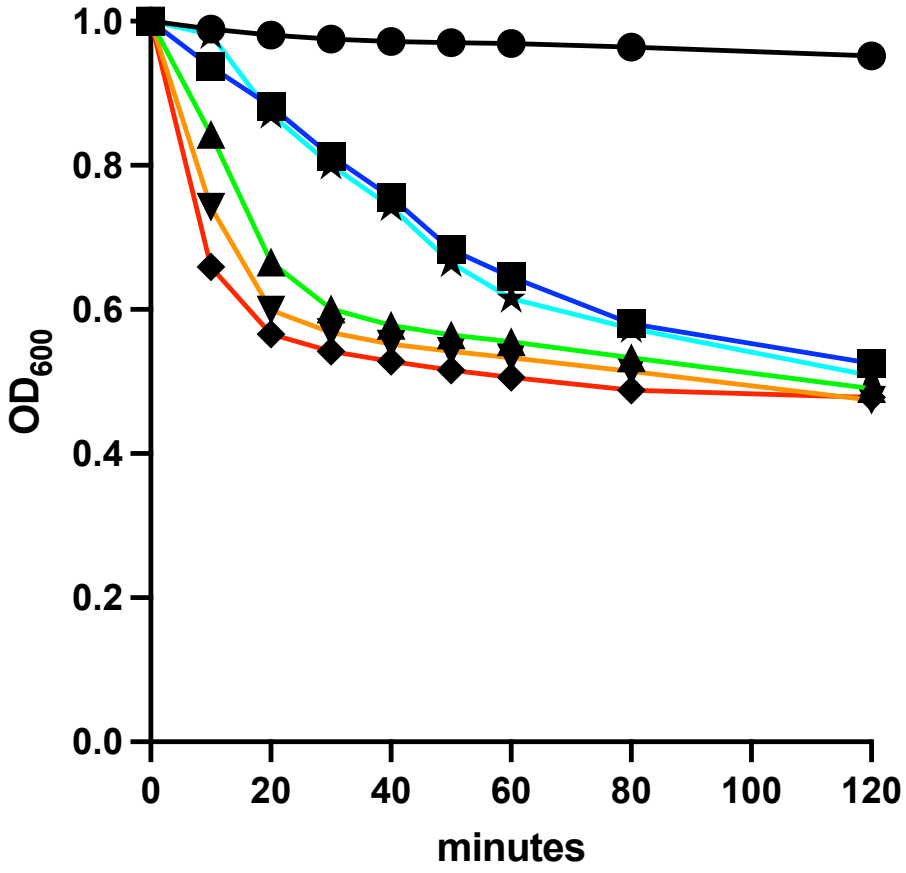
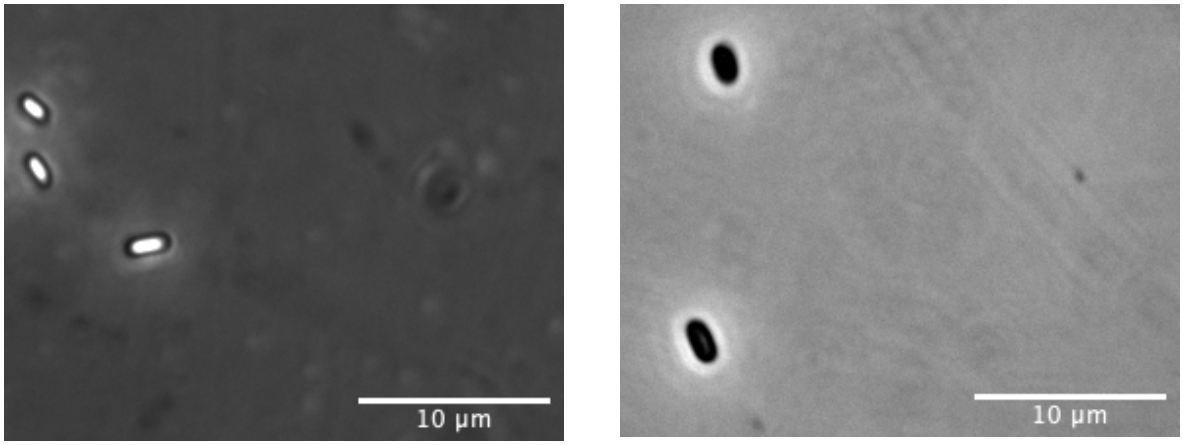


Figure 4.13 Effect of different taurocholate concentrations on germination efficiency.

Freshly produced spores of *C. difficile* strain R20291 were suspended in PBS at 37° C and induced with different concentrations of taurocholate including 0.1% (light blue line), 0.2% (dark blue line), 1% (green line), 2% (orange line) and 4% (red line) with fixed glycine concentration (50mM). Changes in OD_{600} were recorded every 10 minutes for two hours using TECAN infinite F200 pro plate reader. Spores suspended in PBS only (black line) were used as a negative control. Experiment was performed in duplicate and the readings of OD_{600} were normalized by the initial value.

As can be seen from the results presented in Figure 4.13, the rate of reduction in optical density (indicates germination) increased as the concentration of sodium taurocholate increased. Following 60 minutes of incubation, the optical density was similar for all concentrations of taurocholate.

To confirm that the reduction of optical density seen following the addition of germinants was due to germination, phase contrast microscopy was performed (Figures 4.14) (Mayer et al., 2008; Sorg and Sonenshein 2008). As can be seen from the Figure, in the presence of germinants, the conversion of spores from phase bright (Figure 4.14a) to phase dark with no phase bright spores being detected (Figure 4.14b) following 60 minutes of incubation, indicating that germination had occurred.



(a) (b)

Figure 4.14 Phase contrast images of germinating *C. difficile* spores

Purified spores of *C. difficile* strain R20291 were imaged without germinants (a). Spores appear in phase bright which indicate ungerminated spores. Following 60 minutes germination of *C. difficile* spores with 0.1% sodium taurocholate and 50 mM glycine, spores appear to form phase dark spores (b). Images were captured using Leica DMIRB Inverted Leica Modulation Contrast Microscope.

Bile salts such as sodium taurocholate have detergent properties which may damage the cell wall of vegetative bacteria. Additionally, the maximum level reported of taurocholate in human duodenum was 0.125% (Leverrier et al., 2003). Thus, the lowest effective concentrations of germinants, 0.1% sodium taurocholate and 50mM glycine, were employed in subsequent experiments to prevent cell wall damage and to mimic the natural levels of taurocholate at the site of action (Leverrier et al., 2003; Elmi et al., 2018; Sievers et al., 2019).

4.3.9 Effect of Spore Germinants on the Activity of EAD Against Vegetative R20291.

We determined the impact of the germinant, sodium taurocholate, and the co-germinant, glycine, on the ability of EAD to lyse vegetative *C. difficile*. As can be seen from Figure 4.15, the presence of these germinants did not negatively impact the activity of the lysin; in contrast, they have positive effect.

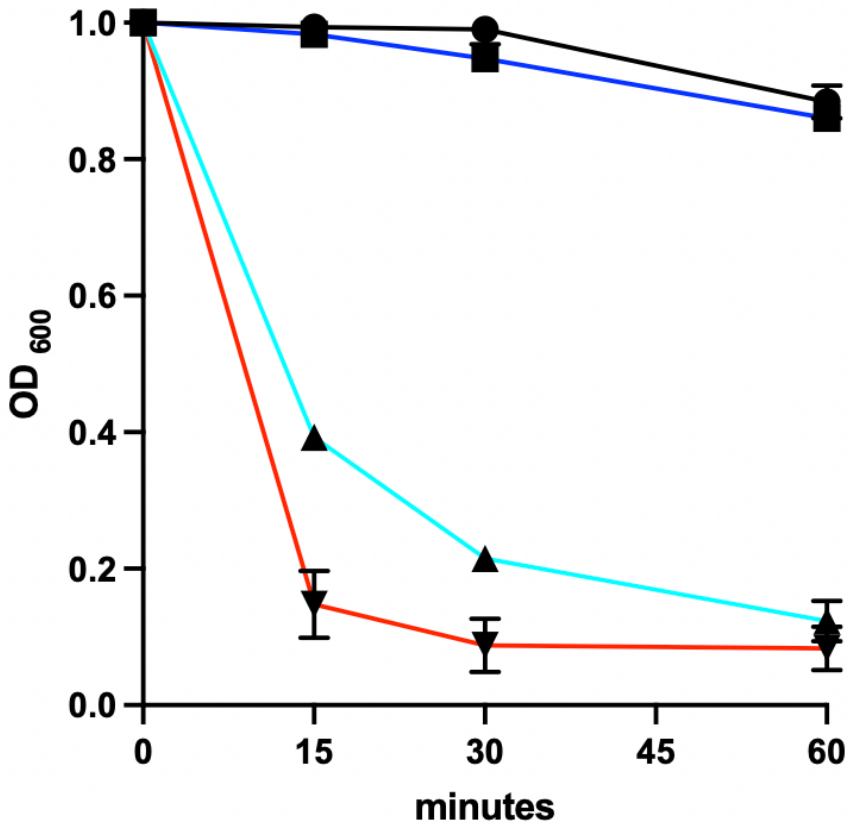


Figure 4.15 Impact of germinants on the activity of EAD against vegetative *C. difficile* R20291.
Vegetative *C. difficile* strain R20291 was grown to mid-log phase in pre-reduced BHI broth. Cells were centrifuged and resuspended in PBS only (black line), PBS and germinants, 0.1% Tc and 50mM glycine (dark blue line), PBS and 60µg/mL EAD (light blue line), and PBS + germinants and 60µg/mL EAD

(red line) for 60 minutes at 37° C. The experiment was performed in triplicate and readings were normalized to initial value.

4.3.10 Effect of Spore Germinants and Calcium on the Activity of EAD Against Vegetative R20291.

We next determined if the presence of calcium chloride at 12.5 mM in addition to germinants had any impact on the lytic activity of EAD against vegetative bacteria. This level of calcium was selected to reflect the level reported to be present in the mouse ileum (Kochan et al., 2017). As can be seen from Figure 4.16, the presence of 12.5 mM calcium chloride reduced the lytic activity of EAD alone and in combination with germinants (35.25% for both groups following 60 minutes of incubation).

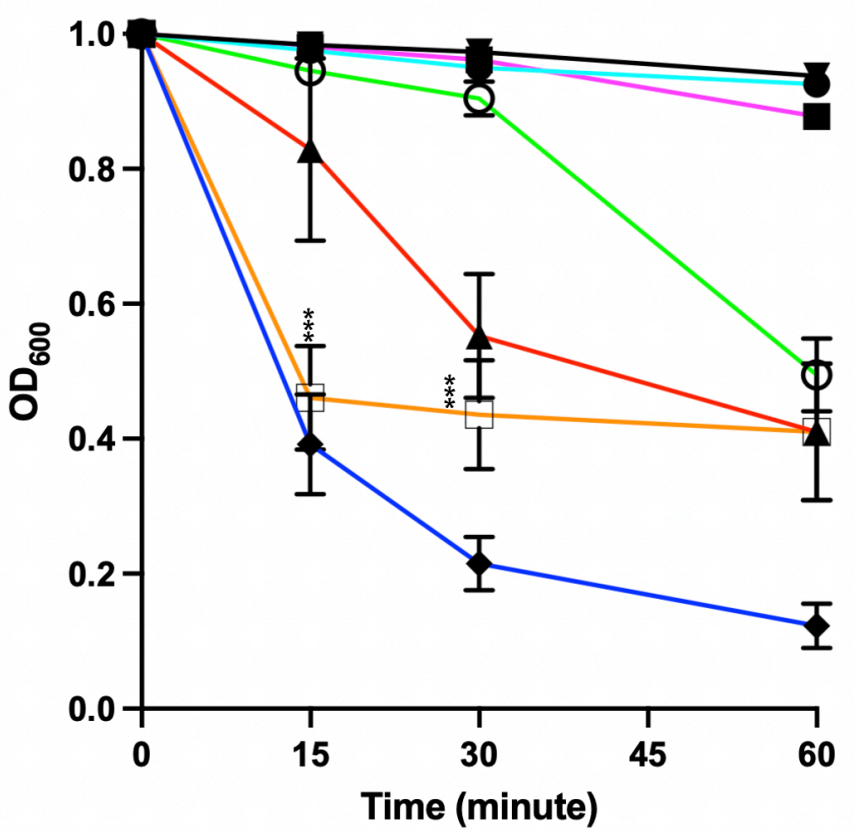


Figure 4.16 The impact of germinants and calcium on the lytic activity of the EAD of LysCD6356 against vegetative R20291.

C. difficile strain R20291 was grown to mid-log phase and resuspended in DIW (black line), DIW with 12.5mM $CaCl_2$ (light blue line), DIW with germinants (0.1%Tc and 50 mM glycine)(purple line), DIW with germinants (0.1%Tc and 50 mM glycine) and 12.5 mM calcium (green line). DIW with 12.5mM $CaCl_2$ and EAD (red line), DIW with germinants, calcium and EAD (orange line) and DIW with EAD (dark blue lines). For all experimental groups changes of OD_{600} were normalized by the initial value and the experiment was performed in triplicate at 37° C. Statistical analysis was performed between the

combination of germinants and $CaCl_2$ with and without EAD against vegetative R20291. P value < 0.001 (***).

Interestingly, we also saw a reduction in the optical density of vegetative bacteria in the absence of EAD but in the presence of Tc, glycine, and calcium, an effect that was not seen in the presence of taurocholate and glycine, or calcium chloride alone. These results suggest that the combination of calcium and germinants may possess some form of direct activity against the vegetative bacteria in the absence of EAD. Similar results were seen with higher concentrations of calcium chloride (25 and 50 mM) in the presence of germinants (0.1% Tc, and 50 mM glycine). The reduction of OD_{600} of these two groups were 50.2% and 50.1% following 60 minutes incubation, respectively (Figure 4.17).

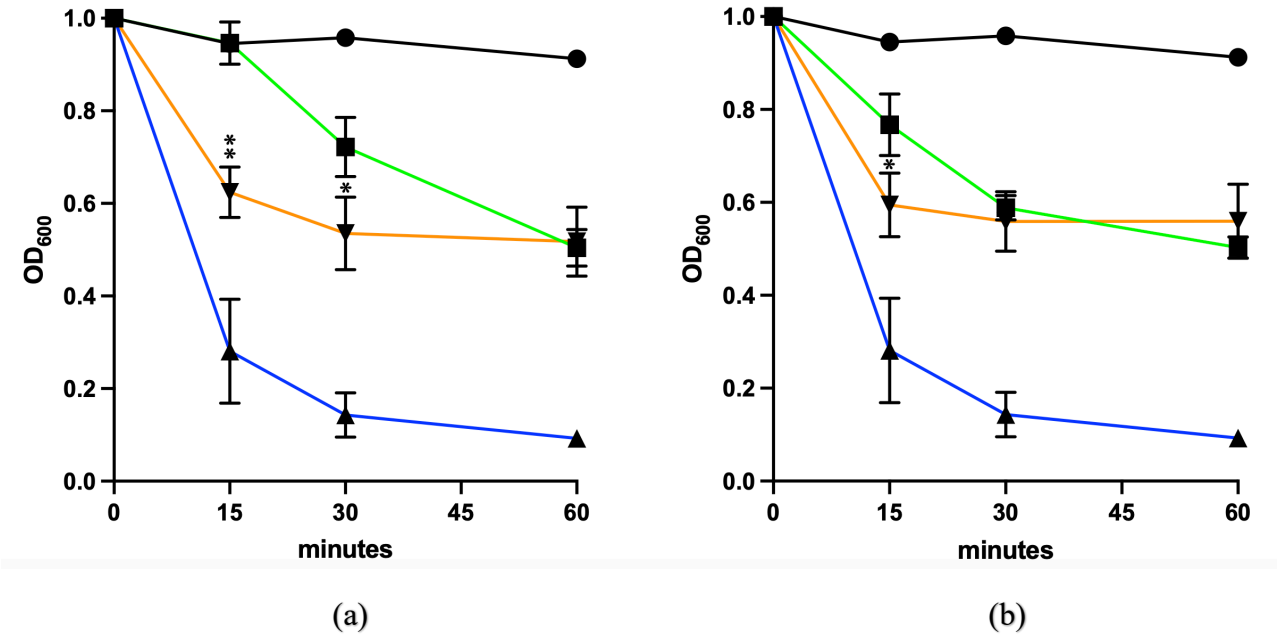


Figure 4.17 Impact of germinants and 25mM (a) and 50mM (b) calcium chloride on the activity of the EAD of LysCD6356 against vegetative R20291.

C. difficile strain R20291 was grown to mid-log phase and resuspended in DIW (black line), DIW with germinants (0.1%Tc and 50 mM glycine) and 25mM (a) or 50mM (b) calcium (green line), DIW with germinants, calcium and EAD (orange line) and DIW with EAD (dark blue lines). For all experimental groups, changes of OD_{600} were normalized by the initial value and the experiment was performed in triplicate at 37° C. Statistical analysis was performed between the combination of germinants and $CaCl_2$ with and without EAD against vegetative R20291. P value = < 0.05(*), < 0.01(**).

4.3.11 Characterization of Bacterial Structure Using SEM

To visualize the impact of the various treatments on the structure of the vegetative bacterium, we examined treated samples using scanning electron microscopy. As demonstrated previously in Figure 4.4, exposure of the vegetative bacteria to the lysin caused disruption of the cell surface, which increased with time.

SEM imaging was also used to investigate the impact of calcium chloride on the activity of EAD in the presence of germinants. Even though calcium chloride was able to lower the impact of EAD on vegetative R20291, the disruption of *C. difficile* cell surface as a result of EAD was still observed, as can be seen in Figure 4.18.

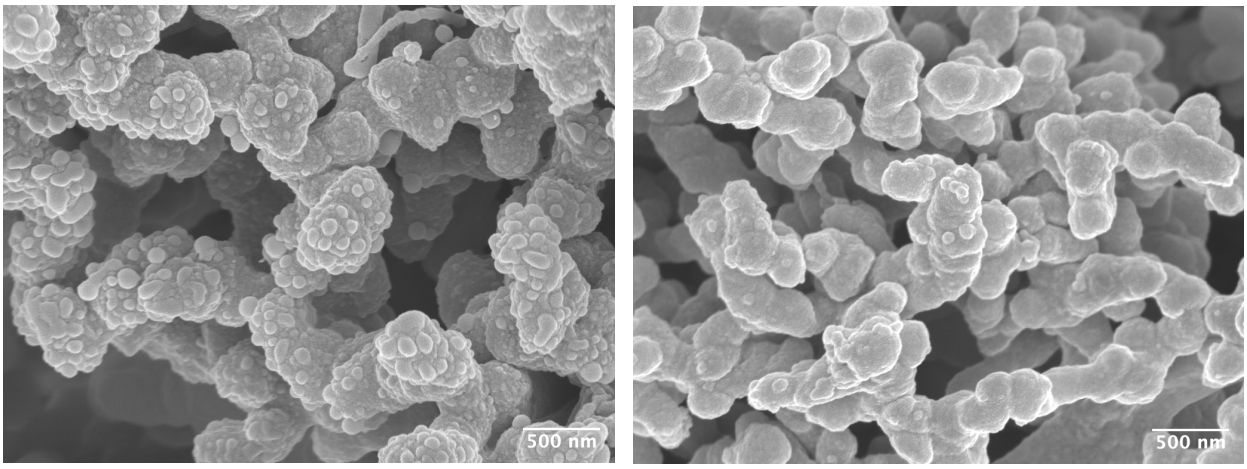


Figure 4.18 SEM images of *C. difficile* R20291 exposed to EAD of LysCD6356 and germinants in presence and absence of calcium

Vegetative *C. difficile* strain R20291 was incubated in BHI broth to mid log phase after which cells were resuspended in DIW and exposed to 60 µg/mL EAD of LysCD6356 and germinants (0.1% taurocholate and 50 mM glycine) in absence (a) and presence (b) of 12.5mM calcium chloride for 30 minutes. Images are X 70,000 magnification.

Even though the combination of germinants and calcium chloride without endolysin against vegetative R20291 clearly reduced the optical density, the visual analysis using SEM demonstrated no obvious cellular disruption (Figure 4.19).

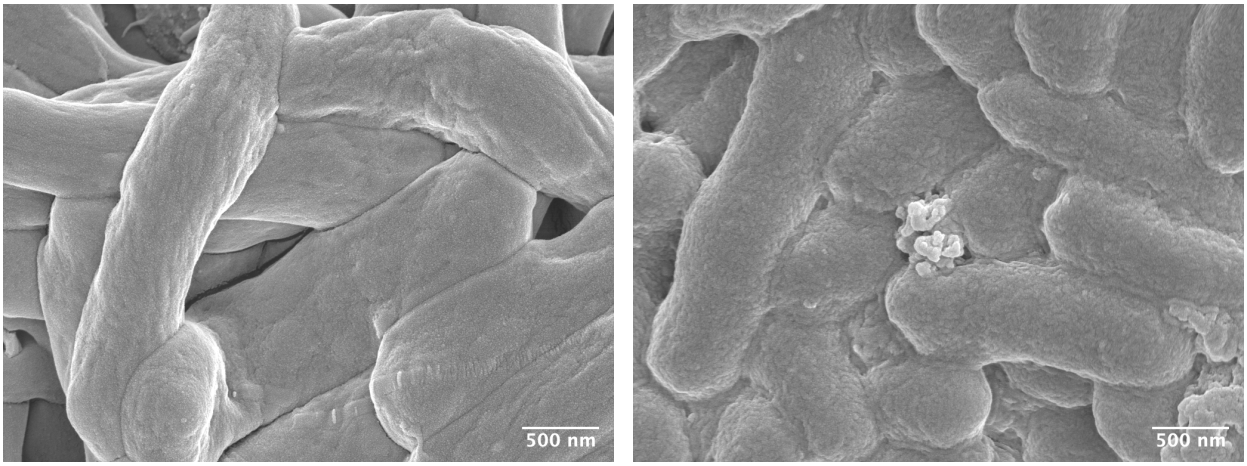


Figure 4.19 SEM images of *C. difficile* R20291 exposed to germinants in presence and absence of calcium.

Vegetative *C. difficile* strain R20291 was incubated in BHI broth to mid log phase after which cells were resuspended in SDW and exposed to (a) germinants (0.1% taurocholate and 50 mM glycine) (b) germinants and 12.5mM calcium chloride for 30 minutes. Images are X 70,000 magnification.

In an effort to understand the reason behind the reduction of the optical density, the length of 30 *C. difficile* cells was investigated using ImageJ software. The analysis revealed that there was a significant reduction in the length of vegetative *C. difficile* exposed to germinants and calcium chloride: 1.73 μm (± 0.28) compared to the germinants only 4.46 μm (± 1.05) (Figure 4.20). Indeed, the change of bacterial size directly changes the OD_{600} by providing a chance for the light to pass through the sample.

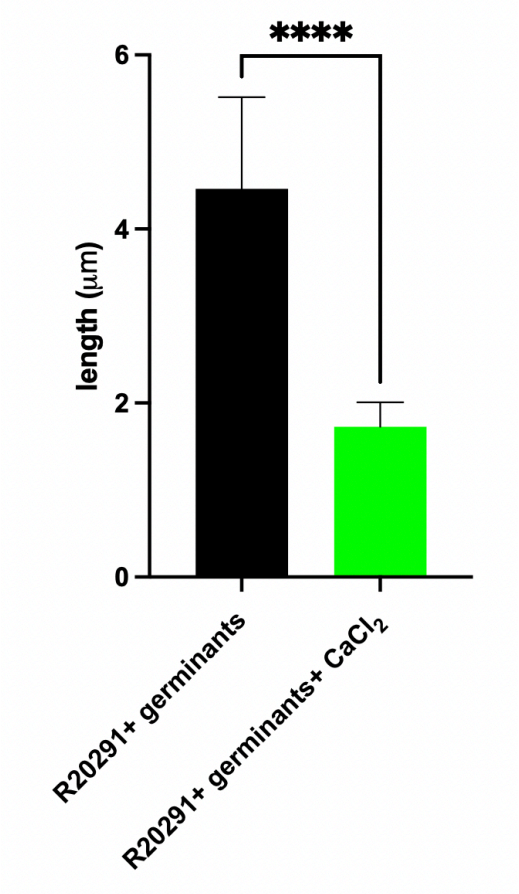


Figure 4.20 Length comparison between vegetative *C. difficile* and germinants with and without calcium.

Vegetative *C. difficile* strain R20291 was incubated in BHI broth to mid log phase after which cells were resuspended in SDW exposing to germinants (0.1% Tc + 50mM glycine) (black bar) or to germinants and $CaCl_2$ (green bar). Cells were then fixed and imaged using SEM. The length of 30 different cells was measured using ImageJ and a statistical analysis was performed. P value = < 0.0001 (****).

4.3.12 Impact of germinants and calcium chloride on *C. difficile* R20291 spore germination

The ability of the germination mixture in either 20 mL DIW or BHI to trigger the germination of R20291 spores in the presence and absence of 12.5 mM calcium chloride was determined.

While a reduction in optical density was observed, indicative of germination, in the presence of 0.1% Tc and 50 mM glycine in DIW, the rate of germination was considerably faster in the remaining germination mixtures, with the majority of spores germinating in approximately 15 minutes (Figure 4.21). These results suggest that the presence of calcium chloride can enhance the rate of germination of R20291 spores, and the presence of BHI supported bacterial outgrowth.

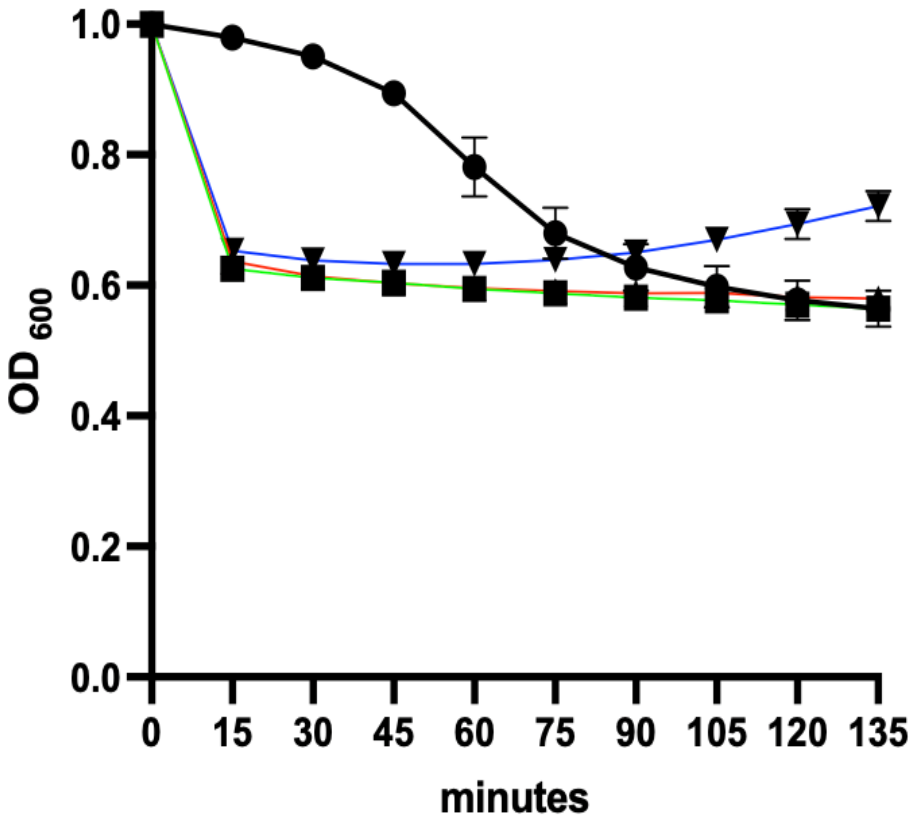
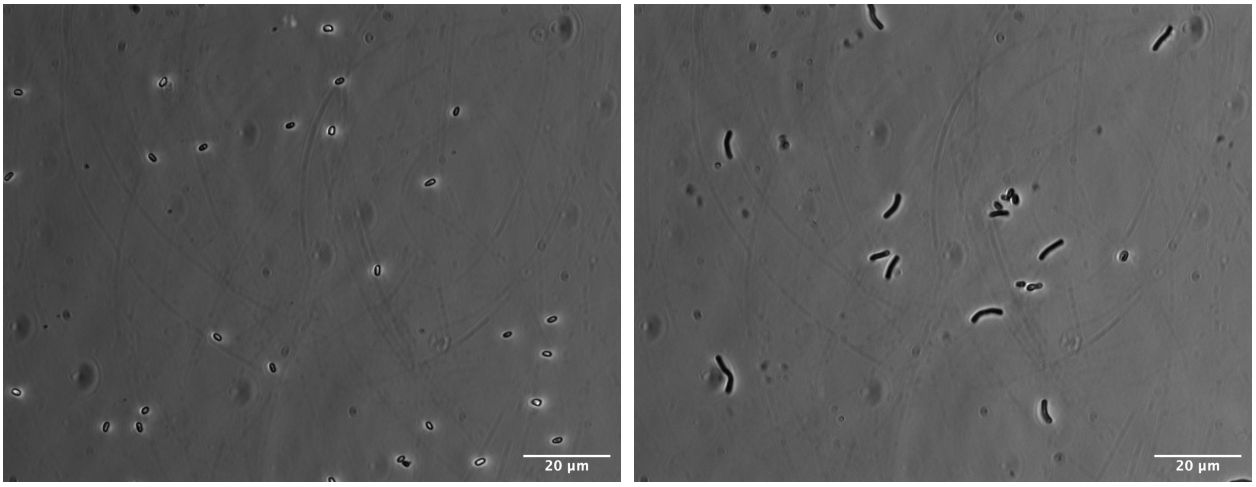


Figure 4.21 Germination of *C. difficile* R20291 spores following exposure to a mixture of germinants.

R20291 spores were incubated in the presence of different germination mixtures for 150 minutes at 37º C and the change in OD_{600} was determined. The following germination mixtures were used; 0.1% Tc and 50 mM glycine in DIW (black line), 0.1% Tc and 50 mM glycine and 12.5 mM calcium chloride in DIW (green line). 0.1% Tc and 50 mM glycine in BHI (red line), 0.1% Tc and 50 mM glycine and 12.5 mM calcium chloride in BHI (blue line). Experiment was performed in triplicate.

To confirm that the reduction of optical density seen following the addition of germinants was due to germination, phase contrast microscopy was performed (Mayer et al., 2008a). As can be seen from Figures 4.22, both groups germinated, and vegetative cells were observed in the group of spores incubated in BHI with germinants following 150 minutes.



(a) (b)

Figure 4.22 Phase contrast images of germinating *C. difficile* spores following 150 minutes of germination.

Purified *C. difficile* R20291 spores were exposed to 0.1% sodium taurocholate and 50 mM glycine and incubated for 150 minutes at 37º C in SDW (a) or BHI (b). Images were captured using Leica DMIRB Inverted Leica Modulation Contrast Microscope.

4.3.13 Sensitivity of Newly Germinated Spores to Lysis by EAD.

Next, the ability of EAD to lyse newly germinated spores of R20291 was investigated. We compared the effect of delivering EAD at the same time as the germination mixture to that of delivering them sequentially, that is germinants followed by a washing step and then EAD.

4.3.13.1 Co-delivery of Germinants and EAD.

As can be seen from Figure 4.23, spore germination was faster in the presence of additional calcium, although both germination conditions achieved a similar end point at 60 minutes, suggesting germination rather than lysin mediated disruption. This could be due to the release of calcium trapped within the spore or due to the fact that newly germinated spores are relatively resistant to EAD mediated lysis (Mehta et al., 2013).

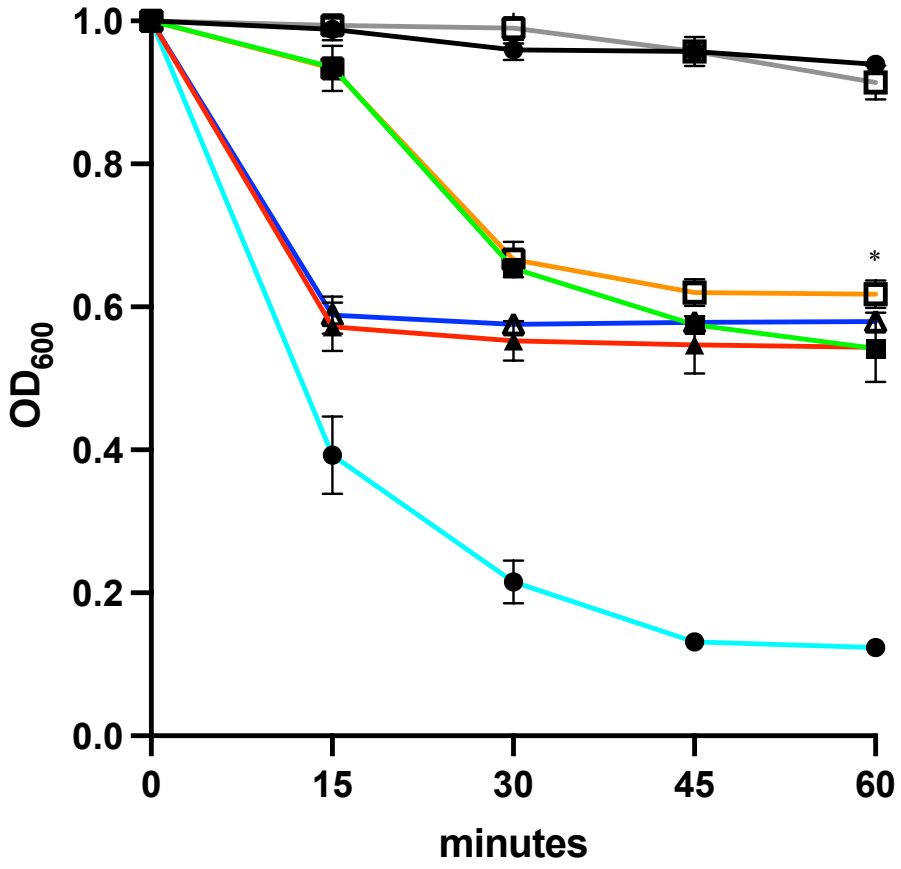


Figure 4.23 The effect of treating *C. difficile* R20291 spores with germinants, calcium and EAD at the same time.

R20291 spores was treated with a germinant mixture comprising 0.1% Tc and 50 mM glycine (green line), germinants with 60 µg/mL recombinant EAD (orange line), germinants with 12.5mM $CaCl_2$ (red line), and germinants with 12.5mM $CaCl_2$ and 60 µg/mL recombinant EAD (blue line) at 37° C. R20291 spores were suspended in DIW (black solid line) and vegetative R20291 were used with (light blue line) and without (gray line) 60 µg/mL EAD as controls. The change of the OD_{600} was monitored throughout the experiment. The experiment was performed in triplicate. Statistical analysis was performed between R20291 spores exposed to germinants with and without 60 µg/mL EAD. P value = < 0.05 (*).

4.3.13.2 Germinants followed by EAD.

Finally, we determined the impact of delaying the addition of EAD to the newly germinated spores on the subsequent lysis of the vegetative organisms. Following germination for 150 minutes, the newly germinated cells (confirmed by phase contrast microcopy) were washed with PBS and resuspended to an OD_{600} of 1 in the presence of EAD. As can be seen in Figure 4.24, the EAD treated cells showed a similar rate of reduction in OD_{600} to EAD control over the first 15 minutes, after which the EAD treated germinated spore plateaued at an OD_{600} of approximately 0.5. This could reflect the fact that not all of the newly germinated R20291 spores were sensitive to EAD induced damage. Interestingly, we saw a

reduction in OD_{600} of the germinated R20291 suspended in PBS in the absence of EAD for reasons which are unclear.

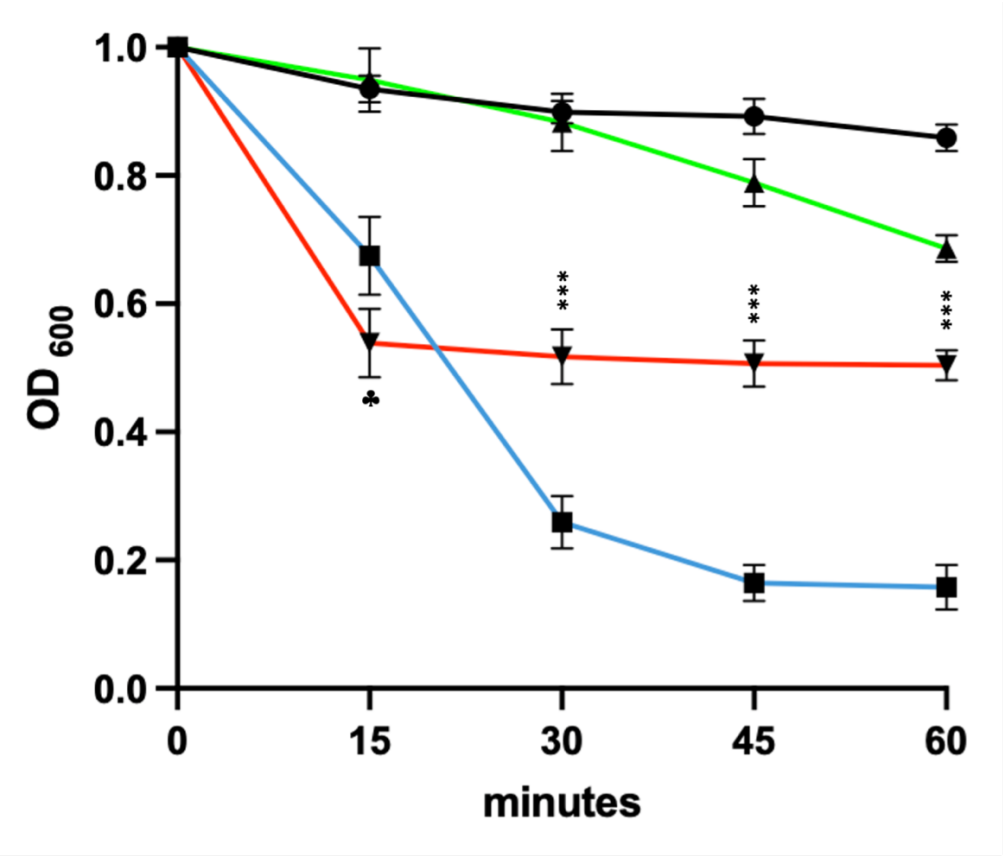


Figure 4.24 The impact of treating newly germinated spores of R20291 with recombinant EAD following a washing step.

Spores of R20291 were incubated with germinants and pre-reduced BHI for 150 minutes. Following germination, the vegetative bacteria were washed with PBS and resuspended to an OD_{600} of 0.8 to 1.2 in PBS (green line) and PBS with 60 µg/mL EAD (red line). Vegetative *C. difficile* obtained from a BHI culture and resuspended in PBS (black line) and PBS with 60 µg/mL EAD (blue line) were used as controls. The experiment was performed in triplicate and readings were normalized to initial value. ♣ refers to a significant (***) reduction of germinated spores after 15 minutes of the addition of 60 µg/mL EAD. P value = < 0.05(*), < 0.01(**), < 0.001 (***), < 0.0001 (****).

4.4 Discussion

Prior research findings demonstrated that EADs are more active than full length endolysins (Mayer et al., 2011; Wang et al., 2015). The EAD shares 77.5% and 37.4% amino acid sequence identity with PlyCD and $CD27L_{1-179}$, respectively, suggesting that it may be a zinc-dependent N-acetylmuramoyl-L-alanine amidase (Furlon et al., 2021). This hypothesis was confirmed by the ability of zinc to restore the biological activity of the EDTA-treated protein. The EAD also demonstrated a similar spectrum of enzymic activity to that of CD27L across a wide range of pH values (Mayer et al., 2008a).

In addition to determining the contribution of zinc to biological activity, we also examined the impact of calcium, as this compound occurs naturally in the gut and has been reported to inhibit the lytic activity of PlyCD (Wang et al., 2015). We found that calcium also inhibited the lytic activity of the EAD in a concentration-dependent manner. Interestingly, the enzyme retained significant lytic activity (P value= < 0.001 (***) and < 0.0001 (****)) at concentrations of calcium in excess of the levels reported to occur in the gut (15 mM), suggesting that this inhibitory effect could be overcome *in situ* by increasing the local concentration of the lysin.

Given that our ultimate aim is to eliminate the spore form of the pathogen from the gut, we next determined the ability of the EAD to lyse the vegetative form of *C. difficile* in the presence of spore germinants. Although sodium taurocholate and glycine had no impact on lytic activity, the addition of calcium chloride inhibited lysis in a concentration-dependent manner.

Interestingly, the combination of sodium taurocholate, glycine, and calcium in the absence of the EAD significantly reduced the length of the vegetative bacterium. During the course of infection, vegetative *C. difficile* encounters multiple stresses in the gastrointestinal tract, including exposure to bile salts and high osmolarity (Chowdhury et al., 1996; Begley et al., 2005a). Bile salts have detergent properties that can modify the bacterial cell wall in a manner that reduces the organism's osmotic tolerance (Oberkampf et al., 2022). It may be the case that the reduction in size was due to the combination of bile-salt-induced cell wall disruption and osmosis-driven leakage of intracellular contents due to the relativity high concentration of extracellular calcium.

Following the confirmation of the EAD's capability to lyse vegetative bacteria in the presence of germinants, the efficacy of the lysin in targeting the spore form of the pathogen was subsequently assessed.

The simultaneous delivery of germinants and lysin was not as effective as delivering them separately. It is possible that the rapid release of calcium held within the cortex of the spore upon the triggering of germination when combined with the calcium already present in the germination mixture was sufficient to inhibit lytic activity. Another possibility is that the newly germinated organisms are relatively resistant to EAD-mediated lysis. Mehta and colleagues observed that spores of the Sterne strain of *B. anthracis* required at least 2 h after the exposure to germinants before the actively growing organism developed sensitivity to lysin-mediated disruption (Mehta et al., 2013).

Delivery of intact proteins such as lysins to the lower GI tract is a particular challenge due to the harsh pH encountered in the stomach and the presence of degradative proteases. One option could be to express the lysin *in situ* from a bacterial host such as *Lactococcus lactis*, which is a food-grade probiotic organism that has been genetically engineered to express a recombinant protein. Previous studies have confirmed the feasibility of using this approach to produce therapeutic proteins in the GI tract (Van Pijkeren and Barrangou 2017). Mayer and colleagues reported that biologically active *C. difficile* ΦCD27 endolysin could also be expressed from *L. lactis,* raising the possibility that this could be a viable delivery platform (Mayer et al., 2008a).

Thus, one could envision a scenario in which patients who had experienced an episode of CDI were given a lysin-expressing strain of *L. lactis* to colonize the gut followed sometime later by a mixture of germinants and calcium. The continued expression of lysin in the gut could potentially overcome temporary germination-mediated inhibition of lysin activity.

In conclusion, we have demonstrated the ability of EAD of LysCD6356 to lyse the vegetative form of a diverse range of clinical isolates of *C. difficile,* which included members of the hypervirulent 027 ribotype. Our results suggest a strategy in which spore germination followed several hours later by exposure to lysin could potentially be developed as a strategy to reduce the risk of relapsing CDI.

Given the concentration dependent effect of the lysin, its proximity to the bacterium post germination is likely to have impact on its efficacy. In the next chapter, the ability of endolysins to attach to the surface of *C. difficile* spores will be assessed.

Chapter 5 Binding of endolysins to the *C. difficile* spore surface

5.1 Introduction

The spore form protects *C. difficile* from external stresses, such as stomach acidity, aerobic conditions, high temperatures and common disinfectants. Moreover, it plays a crucial role in shielding the pathogen from antibiotics and the action of phages and endolysins.

The ability of endolysins to bind to the surface of the spore and vegetative forms of spore forming bacteria has been reported. The phage derived endolysins PlyG and LysPBC2 bind to members of the *Bacillus spp* via a 60-62 amino acid spore binding domain located primarily in the catalytic domain (Yang et al., 2012; Kong et al., 2019). The bacteriophage endolysin of the *Clostridium tyrobutyricum* phage CTP1 (CTP1L) and its CBD also bind to the vegetative form and spores produced by *C. beijerinckii, C. tyrobutyricum, C. butyricum,* and *C. sporogenes* (Gómez-Torres et al., 2018). Additionally, Mayer and colleagues reported that phage endolysin CD27L and its catalytic domain, $CD27L_{1-179}$, bound to *C. difficile* vegetative cells (NCTC 11204), but its ability to bind to the spores was not reported (Mayer et al., 2011).

While spores are impervious to degradation by lysins, their ability to attach to the spore surface raises the intriguing possibility that, upon germination, they are released and could potentially kill the emerging vegetative bacterium. By better understanding the nature of this interaction, it may be possible to develop a new generation of lysin based therapeutic agents.

The outermost layer of the *C. difficile* spore is called the exosporium, the structure and composition of which have been investigated across a number of studies, including work undertaken in the Baillie lab (Tina Joshi 2012; Barra-Carrasco et al., 2013; Díaz-González et al., 2015; Antunes et al., 2018; Calderón-Romero et al., 2018; Malyshev 2018). The protein composition of the *C. difficile* exosporium differs from that of other bacterial spores, being comprised of a range of proteins, including cysteine rich proteins such as CdeA, CdeB, CdeC, and CdeM, in addition to collagen like glycoproteins such as BclA1, BclA2, and BclA3, which make up a hair-like projection (Paredes-Sabja et al., 2022). Recent reports indicated that CdeC and CdeM are major components of the exosporium, while CdeC was also reported to be present in the spore coat as a monomer (Barra-Carrasco et al., 2013; Antunes et al., 2018; Calderón-Romero et al., 2018).

The importance of this layer is evidenced by the contribution it makes to *C. difficile* survival and pathogenicity. The loss of CdeC from the exosporial layer results in an increase in sensitivity to lysozyme, heat, and ethanol (Barra-Carrasco et al., 2013). This protein also plays a role in mediating the attachment of spores to the cells of the host intestine, as does CdeM, and may play a role in facilitating cellular uptake (Mora-Uribe et al., 2016; Calderón-Romero et al., 2018). BclA3 is also thought to mediate *C. difficile* spore uptake into intestinal epithelial cells, allowing the pathogen to avoid antibiotic treatment and reemerge to initiate a new infection (Castro-Córdova et al., 2021).

Immunogenicity studies using recombinant CdeC and CdeM have shown that both proteins are able to protect mice and hamsters against *C. difficile* infection when employed as vaccines (Ghose et al., 2016). Studies have also shown antibodies raised against these proteins bind to intact spores, suggesting that at least part of each protein is exposed to the environment (Barra-Carrasco et al., 2013; Díaz-González et al., 2015; Antunes et al., 2018; Calderón-Romero et al., 2018).

Given that each protein plays a central role in spore structure, survival, and pathogenicity, it would suggest that they are under considerable selective pressure to remain unchanged, making them ideal molecules for targeted interventions (Luo et al., 2015). For this reason, these proteins were provisionally identified as potential lysin binding targets. To test this theory, we will attempt to generate recombinant versions of each protein.

Efforts to express non-codon optimized versions of CdeC using a pET-22b expression vector and several different *E. coli* host strains, including BL21(DE3)Ril and SHuffle T7, was reported to have an issue of being insoluble (Barra-Carrasco et al., 2013; Brito-Silva et al., 2018). This could be due to the histidine affinity tag being incorporated into the C-terminus of the expressed protein. It was previously reported that the location of the histidine tag may affect the level of solubility, stability, and activity (Woestenenk et al., 2004; Aslantas and Surmeli 2019). Additionally, researchers reported that N-terminus tags have two advantages over C-terminus tags. First, in most cases, N-terminus tags enhance recombinant protein production by providing a reliable context for efficient translation initiation (Waugh 2005). Second, it facilitates the removal of all or most

residues at the N-terminus of the protein of interest, given that the majority of the endoproteases cleave at or close to the C-terminus of their recognition sites (Malhotra 2009).

Previous attempts to express recombinant CdeC and CdeM from *E. coli* with the histidine incorporated at the N-terminus were successful. In 2016, Ghose and colleagues employed a pET-19b expression vector to express codon optimized versions of CdeC and CdeM from a BL21(DE3) *E. coli* host achieving yields of CdeC and CdeM ranging from 500 to 700 mg/L (Ghose et al., 2016).

5.2 Aims and objectives

The aims of this chapter are to determine if the CD6356 phage endolysin binds to the surface of *C. difficile* spores. Using *in silico* modeling tools, the structure of the CD6356 endolysin will be compared to that of known spore binding endolysins with a view to identifying regions of structural homology. The ability of recombinantly expressed CD6356 phage endolysin to bind to *C. difficile* spores will then be assessed. If they bind, their specificity and their ability to bind to the spore and vegetative forms of different spore forming bacterial species will be determined. Finally, we will attempt to identify the structures on the spore surface to which the endolysin binds. To do this, we will focus on two previously identified exosporial proteins, CdeC and CdeM, which are exposed to the environment.

The objectives are:

1- To identify potential spore binding regions within LysCD6356 and its EAD using a bioinformatic based approach.
2- To determine the ability of recombinant LysCD6356 and its EAD to attach to the spore and vegetative forms of *C. difficile* and other related bacterial species.
3- To express recombinant versions of the exosporial proteins CdeC and CdeM.
4- To determine the ability of recombinant LysCD6356 and its EAD to bind to CdeC and CdeM.

5.3 Results

5.3.1 Identification of potential spore binding regions using an *in silico* based approach

To identify potential spore binding regions within recombinant LysCD6356 and its EAD, a bio-informatic based approach was used. We hypothesized that LysCD6356 and/or its EAD contained a region which mediated attachment to spores.

To determine if this was the case, we first aligned the protein sequences of LysCD6356, CD27L, and CTP1L. The analysis also aligned the protein sequence of EAD of LysCD6356 to the catalytic domains of CD27L, PlyG, and LysPBC2.

A tertiary structural model of LysCD6356 EAD was created and compared to the crystal structures of PlyG and $CD27L_{1-179}$ to identify potential regions of homology.

5.3.1.1 Analyzing protein sequence alignment and percentage of identity

Using BLAST NCBI, the percent identity of LysCD6356 to CD27L was 51% which considered significant (E-value $6e^{-85}$), while the identity with CTP1L was only 19%, indicating no significant similarity. Alignment of the sequences using Clustal Omega confirmed this finding, showing higher similarity to CD27L than CTP1L (Appendex: Figure A6.8). Notably, the C- terminal domain appears to have higher level of similarity than the N- terminal domain, suggesting some form of similarity on the CBD. This finding was in alignment with previously reported analysis of *C. difficile* endolysins to have high similarity with some differences even though the CBD target the same bacteria, suggesting the binding to different targets of the cellular compartment. Investigating the structural similarity especially on the CBDs may clarify the similarity.

The low alignment similarity and percentage of identity between LysCD6356 and CTP1L suggest that if LysCD6356 endolysin binds to *C. difficile* spores, it probably uses a different region than that used by CTP1L.

In contrast, CD27L demonstrated significant sequence alignment similarity and percent identity, raising the possibility that LysCD6356 may contain a region or regions which mediate binding to the vegetative form of *C. difficile* in a similar manner to CD27L. The ability of CD27L to bind to *C. difficile* spores has not been reported.

We next determined if there were any similarities between the protein sequences of the EAD of LysCD6356 and the catalytic domains of PlyG, LysPBC2, and CD27L. The catalytic domain of LysPBC2 was reported to have a high percent identity (57%) when aligned with the catalytic domain of PlyG (Kong et al., 2019); however, the percent identity of the EAD of LysCD6356 with the catalytic domain of PlyG and LysPBC2 was only 13% and 18%, respectively. In contrast, the percent identity of EAD of LysCD6356 to $CD27L_{1\text{-}179}$ was 37.4% indicating significant similarity (E-value $5e^{-32}$) (Furlon et al., 2021). Additionally, the protein sequence alignment of EAD of LysCD6356 demonstrated no significant similarity with the catalytic domain of PlyG and LysPBC2 and the spore binding domain of PlyG and LysPBC2 (Appendex: Figure A6.9 and Figure A6.10).

The low level of protein sequence alignment and percentage of identity between the EAD of LysCD6356 and the catalytic domains of PlyG and LysPBC2 and their spore binding domains suggests that they are unlikely to share a common spore binding region.

$CD27L_{1\text{-}179}$ demonstrated better alignment and a higher percentage of identity, suggesting a higher likelihood of potentially sharing a vegetative binding sequence with the EAD of LysCD6356.

5.3.1.2 Analyzing the homology in protein tertiary structures

Since there is no crystal structure for LysCD6356 and its EAD, AlphaFold software was used to predict both structures. The confidence of AlphaFold for a predicted model is measured by the predicted local distance difference test (pLDDT) which ranges from 0 to 100. The lower pLDDT, the lower the level of confidence in the predicted structure. The predicted structures of LysCD6356 and its EAD were highly confident where the pLDDT scores were 93.8 and 97.62, respectively.

The alignment of LysCD6356 structure and CD27L demonstrated high structural alignment (RMSD = 1.202). The EADs of LysCD6356 and CD27L show better alignment than CBD, as demonstrated in Figure 5.1

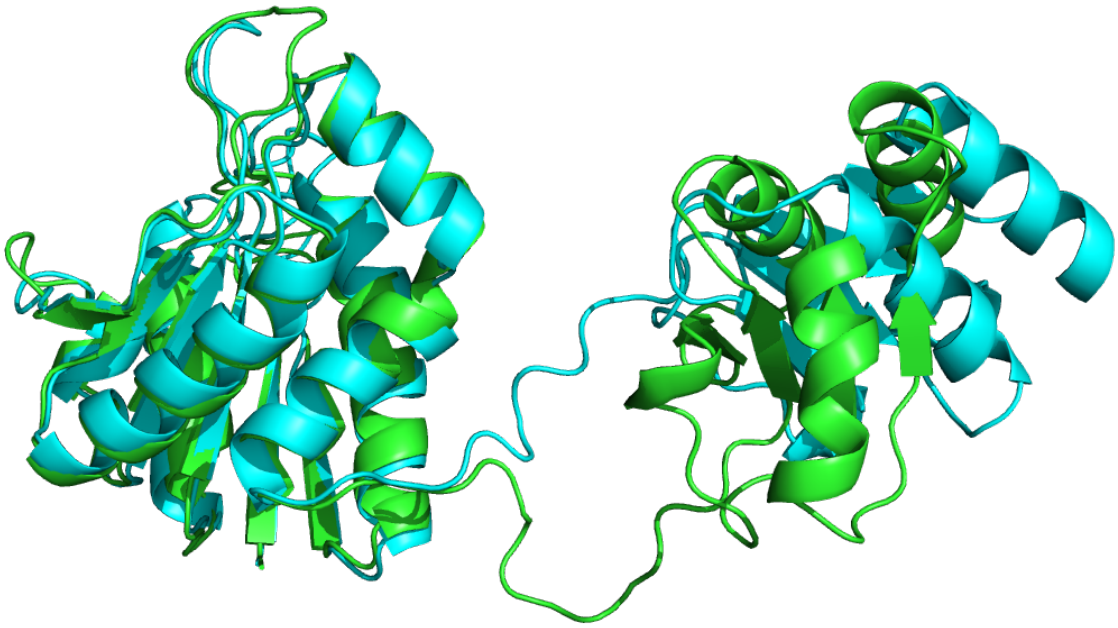


Figure 5.1 Structural alignment of LysCD6356 and CD27L

The predicted structures of CD27L (light blue) was aligned with the predicted structure of LysCD6356 (green). The left part of the the structure represent the EAD and the right part represent the CBD. Structures were predicted using AlphaFold software. Alignment and modification of colours were applied using PyMol program.

However, the alignment of CTP1L and LysCD6356 demonsted low structural alignment (RMSD = 12.643) as demonstrated in Figure 5.2.

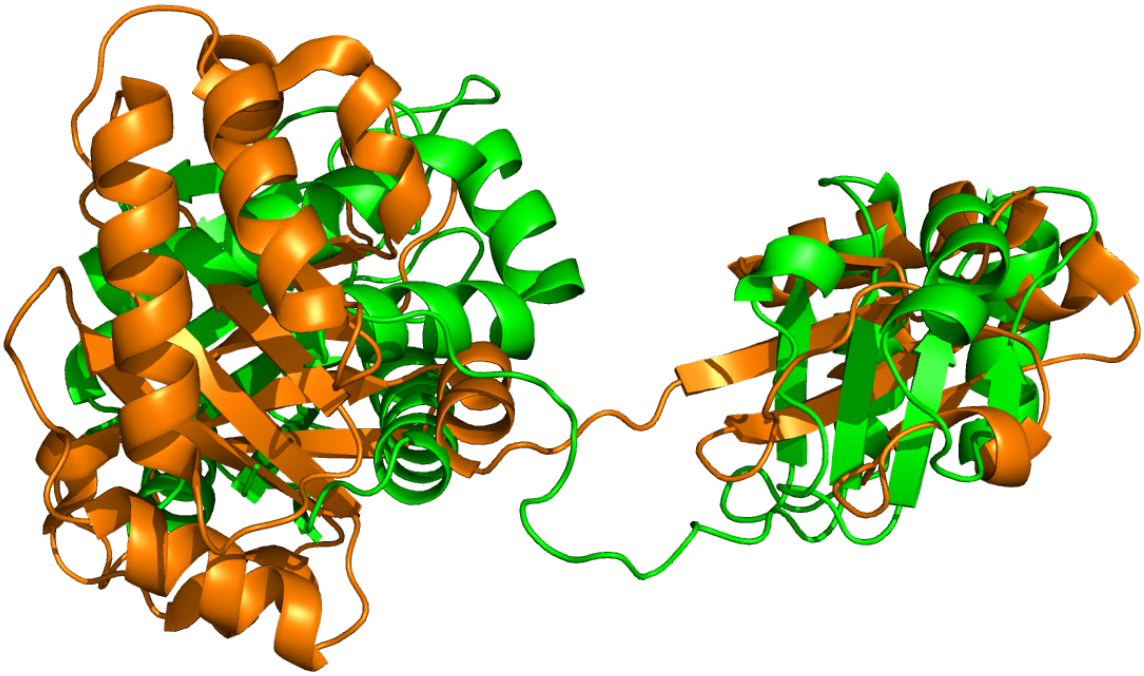


Figure 5.2 Structural alignment of LysCD6356 and CTP1L

The crystal structures of CTP1L (Orange) was aligned with the predicted structure of LysCD6356 (green). The left part of the the structure represent the EAD and the right part represent the CBD. Crystal structure of CTP1L was downloaded from PDB protein data bank (PDP code 5A6S). LysCD6356 structure was predicted using AlphaFold software. Alignment and modification of colours were applied using PyMol program.

The protein sequence alignment analysis and structural alignment analysis demonstrated that LysCD6356 has high similarity to CD27L increases the chances of sharing a region that could be responsible for protein binding to vegetative *C. difficile*.

Investigating the alignment of the EAD of LysCD6356 was then investigated. Similar to the previous observation in Figure 5.1, the EAD of LysCD6356 and $CD27L_{1\text{-}179}$ demosntrated high structural alinment (RMSD = 0.824) as represented in Figure 5.3.

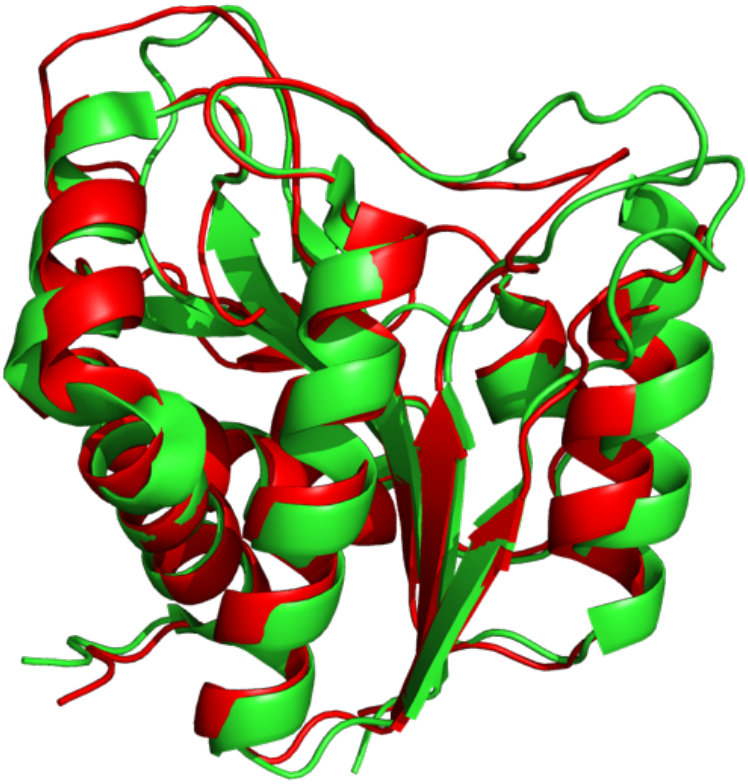


Figure 5.3 Structural alignment of the catalytic domains of LysCD6356 and CD27L$_{1-179}$
The crystal structure of the catalytic domain of CD27L$_{1-179}$ (green) was aligned with the predicted structure of EAD of LysCD6356 (red). Crystal structure of CD27L$_{1-179}$ was downloaded from PDB protein data bank (PDP code 3QAY). EAD of LysCD6356 structure was predicted using AlphaFold software. Alignment and modification of colours were applied using PyMol program.

The high similarity of the EAD of LysCD6356 to CD27L$_{1-179}$ increases the chances of sharing a region that could be responsible for protein binding to vegetative *C. difficile*. The high confidence in making EAD structure using AlphaFold program, the reported significant similarity with CD27L$_{1-179}$ (37.4% percent identity and $5e^{-32}$ E-value) (Mayer et al., 2011), requirement of ligand molecule (Zn^{+2}), demonstrating similar function, and high similarity of 3D structural alignment to

the crystal structures of $CD27L_{1-179}$ (RMSD = 0.824) confirm the protein alignment analysis between EAD of LysCD6356 and $CD27L_{1-179}$.

Alignment of AphaFold predicted EAD 3D structure with crystal structure of PlyG, a spore binding lysin, demonstrated low similarity (RMSD = 7.97) as demonstrated in figure 5.4.

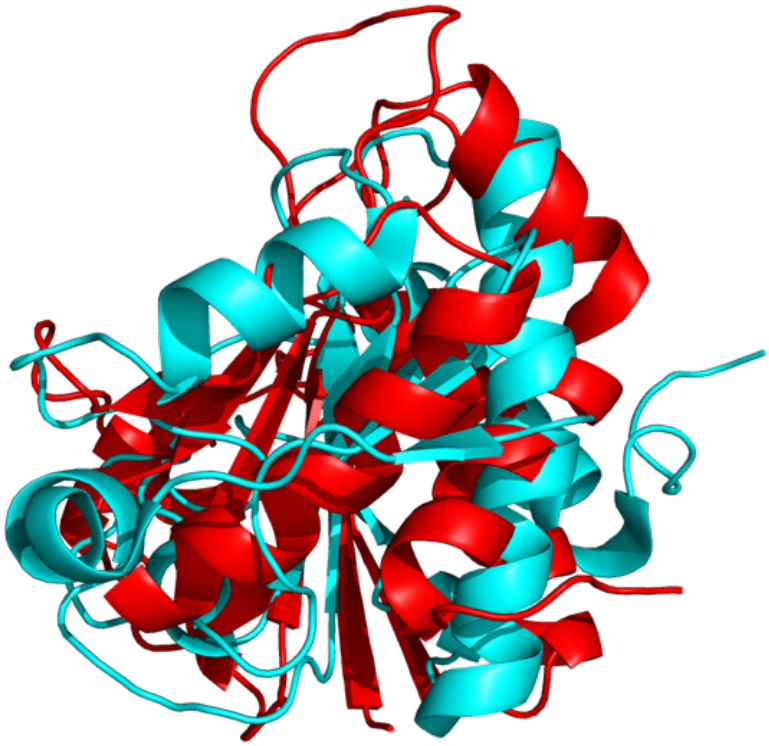


Figure 5.4 Structural alignment of EAD of LysCD6356 and the catalytic domain of PlyG
The crystal structure of the catalytic domain of PlyG (light blue) was aligned with the predicted structure of EAD of LysCD6356 (red). The crystal structure of catalytic domain of PlyG was downloaded from PDB protein data bank (PDP code 2L47) and the predicted structure of EAD of LysCD6356 was downloaded from AlphaFold software. Alignment and modification of colours were applied using PyMol program.

The structure alignment analysis demonstrated that EAD of LysCD6356 differed from PlyG suggestion that any spore binding motif differ from those found in PlyG. Further investigation would be required to determine the spore binding region of EAD.

In conclusion, bioinformatics analysis revealed that CD27L and its EAD ($CD27L_{1-179}$) shared the highest level of protein sequence similarity with LysCD6356 and its EAD (51% and 37.4%, respectively). Additionally, the high degree of structural similarity of the EAD of LysCD6356 to $CD27L_{1-179}$ (RMSD = 1.202 and 0.824, respectively) suggests that it binds to vegetative *C. difficile* in a similar manner (Mayer et al., 2011).

In terms of spore binding, the shortage of data regarding *C. difficile* spore binding sequences and the structural differences observed between the EAD and PlyG models suggest that there are spore binding motifs in EAD that differ from the ones found in PlyG.

5.3.2 Assessment of the ability of recombinant LysCD6356 and its EAD to attach to the spore and vegetative forms of *C. difficile* and other related bacterial species

5.3.2.1 Conjugation of endolysins with Alexa 488

To visualize the binding of LysCD6356 and its EAD, the recombinant proteins were first labeled with the fluorophore Alexa 488. Using the Beer-Lambert law, the mean number of Alexa 488 molecules bound to LysCD6356 (Table 5.1) and EAD (Table 5.2) was estimated to be more than 1.

Table 5.1 Characteristics of LysCD6356 conjugated with Alexa 488.

Peak high is the absorbance of LysCD6356 at 280 nm and the absorbance of fluorophore at 495 nm.

LysCD6356 conjugated to Alexa 488 (LysCD6356-Alexa 488)		
	LysCD6356 (280 nm)	Alexa 488 (495 nm)
Extinction coefficient (ε)	**37,610 M^{-1} cm^{-1}**	**71,000 M^{-1} cm^{-1}**
Peak high	**0.82**	**1.33**
Corrected peak high for fluorophore absorbance at 280 nm	**0.674**	**1.33**
Concentration of stock (μM)	**17.92**	**18.7**
Mean number of dyes per protein	**1.04**	

Readings were repeated three times.

Table 5.2 Characteristics of EAD conjugation with Alexa 488.

Peak high is the absorbance of EAD at 280 nm and the absorbance of fluorophore at 495 nm.

EAD of LysCD6356 conjugated with Alexa 488 (EAD-Alexa 488)		
	EAD of LysCD6356 (280nm)	Alexa 488 (495 nm)
Extinction coefficient (ε)	**23,045 M^{-1} cm^{-1}**	**71,000 M^{-1} cm^{-1}**
Peak high	**0.72**	**2.04**
Corrected peak high for fluorophore absorbance at 280 nm	**0.386**	**2.04**
Concentration of stock (μM)	**16.73**	**28.9**
Mean number of dyes per protein	**1.72**	

Readings were repeated three times.

5.3.2.2 Investigating attachment of Alexa labeled endolysins to *C. difficile* spores.

The ability of Alexa 488 labeled LysCD6356 and its EAD to attach to the surface of bacterial spores was investigated by incubating purified spores in the presence of each protein (Mayer et al., 2011). Two different *C. difficile* strains, R20291 (RT 027) and DS1787 (RT 106), were assessed alongside *B. subtilis* ATCC 27370 and *B. cereus,* ATCC 6464. These two later strains were included to determine the species specificity of the endolysins and to assess the potential contribution of an exosporium.

As can be seen in Figure 5.5, LysCD6356-Alexa 488 attached to the spore forms of both strains of *C. difficile* and to *B. cereus* ATCC 6464 spores but not to those of *B. subtilis* ATCC 27370 (lacks an exosporium), suggesting some form of non-specific binding to *C. difficile,* possibly to components located in the exosporium. Further studies with more isolates will need to be performed to determine if that is indeed the case.

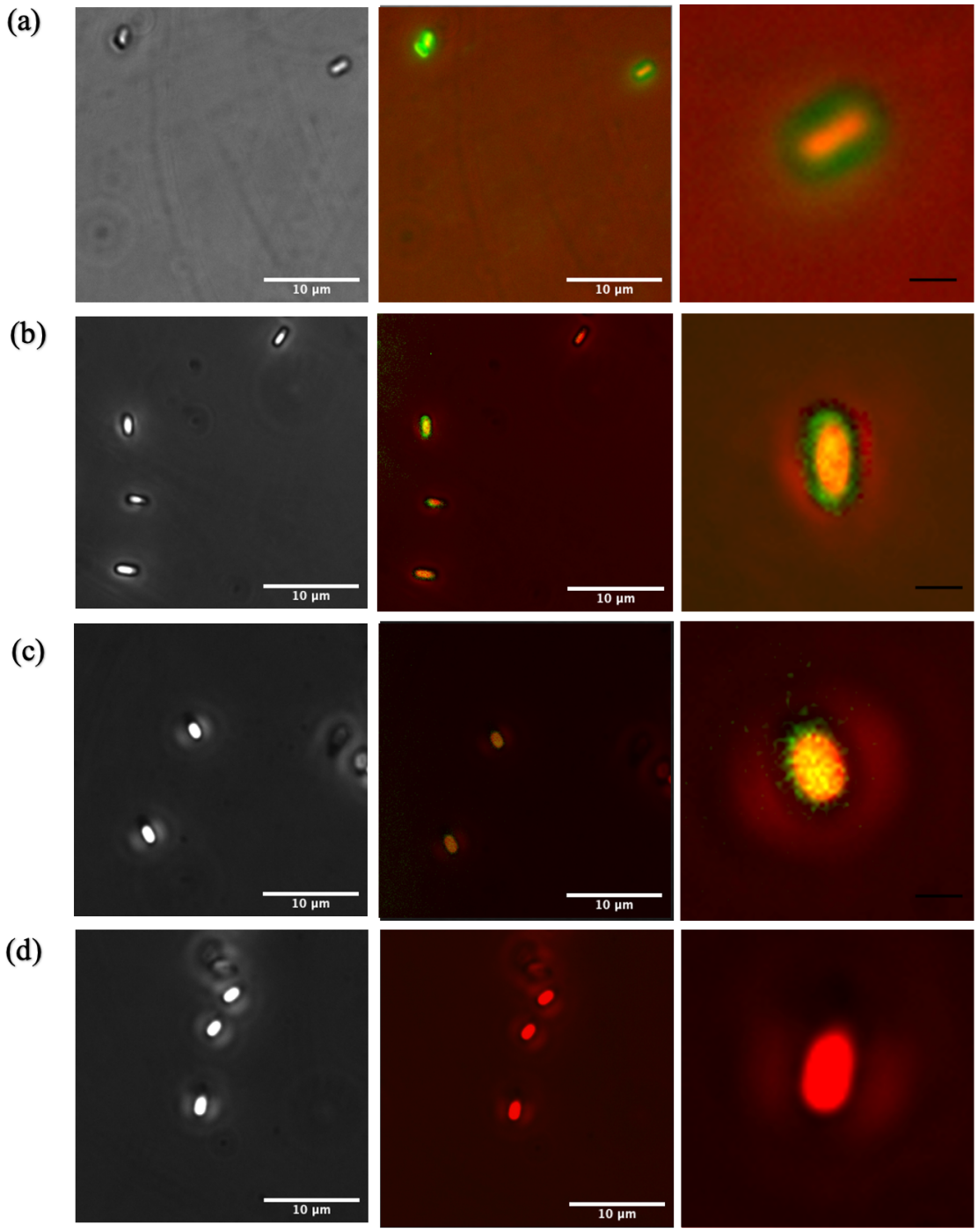


Figure 5.5 Ability of LysCD6356-Alexa 488 to attach to spore surface.

7.5 µM LysCD6356-Alexa 488 was incubated with (a) *C. difficile* R20291, (b) *C. difficile* DS1787, (c) *B. cereus* ATCC 6464 and (d) *B. subtilis* ATCC 27370 spores for 20 minutes at 37º C. Spores were then washed three times and prepared for visualization. Spores were visualized under phase contrast microscopy (left). Using ImageJ, spore image was coloured with red and merged with the image of green coloured fluorescent dye of Alexa 488 (middle). A spore of merged image was magnified (right). Experiment was performed in duplicate, and images were captured from three different fields. Black scale bars length is 1 µm.

We next examined the ability of Alexa labeled EAD to attach to the same spores. As can be seen in Figure 5.6, EAD attached to spores of the *C. difficile* isolates but failed to attach to the spore forms of the other bacterial species, suggesting some sort of specificity for *C. difficile*. Further studies with more isolations will need to be performed to confirm the validity of this statement.

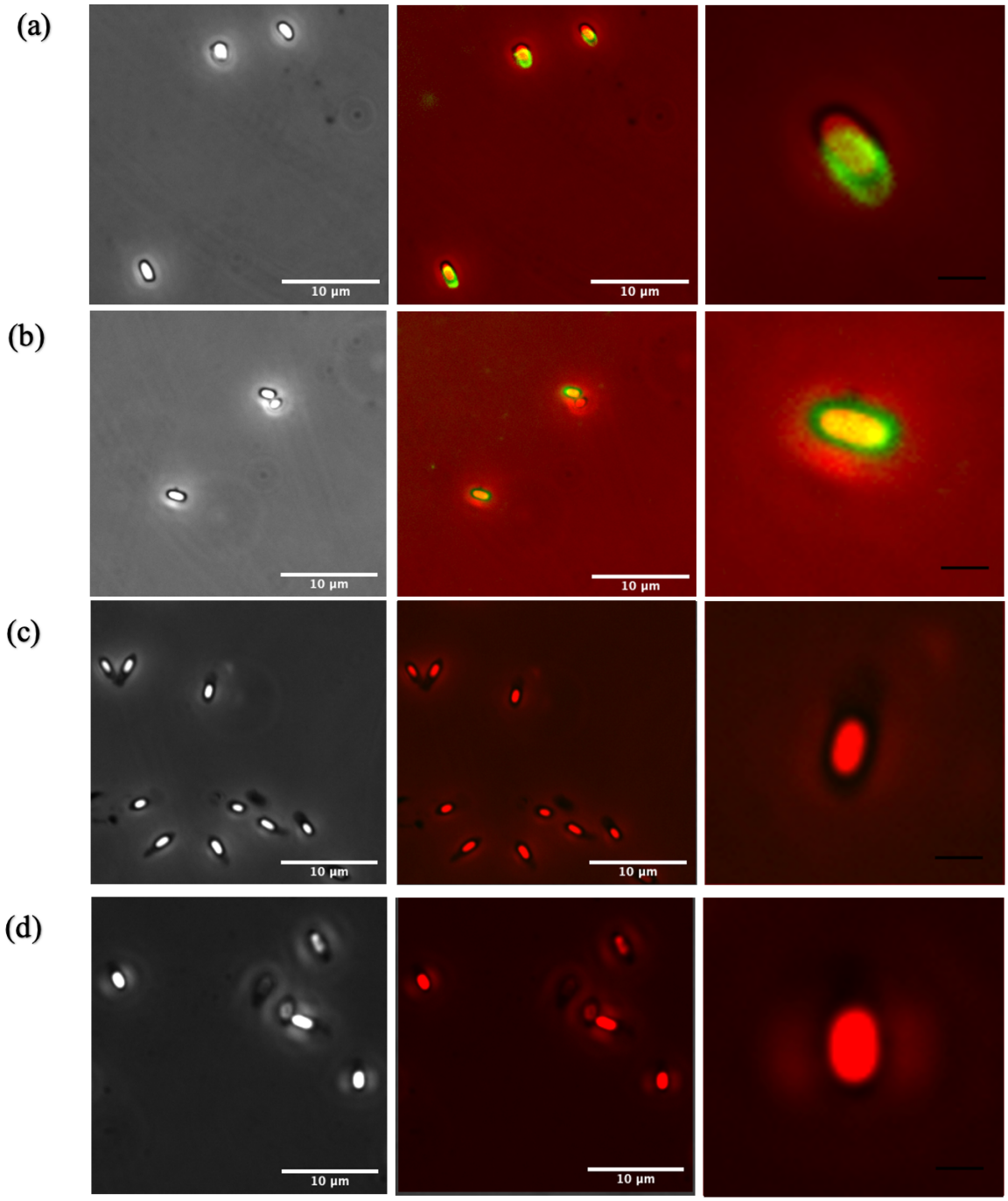


Figure 5.6 Ability of EAD-Alexa 488 to attach to spore surface.

7.5 µM EAD-Alexa 488 was incubated with a) *C. difficile* R20291, (b) *C. difficile* DS1787, (c) *B. cereus* ATCC 6464 and (d) *B. subtilis* ATCC 27370 spores for 20 minutes at 37º C. Spores were then washed three times and prepared for visualization. Spores were visualized under phase contrast microscopy (left). Using ImageJ, spore image was coloured with red and merged with the image of green coloured fluorescent dye of Alexa 488 (middle). A spore of merged image was magnified (right). Experiments was performed in duplicate, and images were captured from three different fields. Black scale bars length is 1 µm.

5.3.2.3 Investigating attachment of endolysins to the surface of vegetative bacteria

The ability of Alexa labeled proteins to bind to the vegetative form of each bacterial isolate was determined. The full length and EAD of LysCD6356 bound to the vegetative forms of the two *C. difficile* test strains but not to the vegetative forms of *B. cereus* ATCC 6464 and *B. subtilis* ATCC 27370 (Figures 5.7 and 5.8).

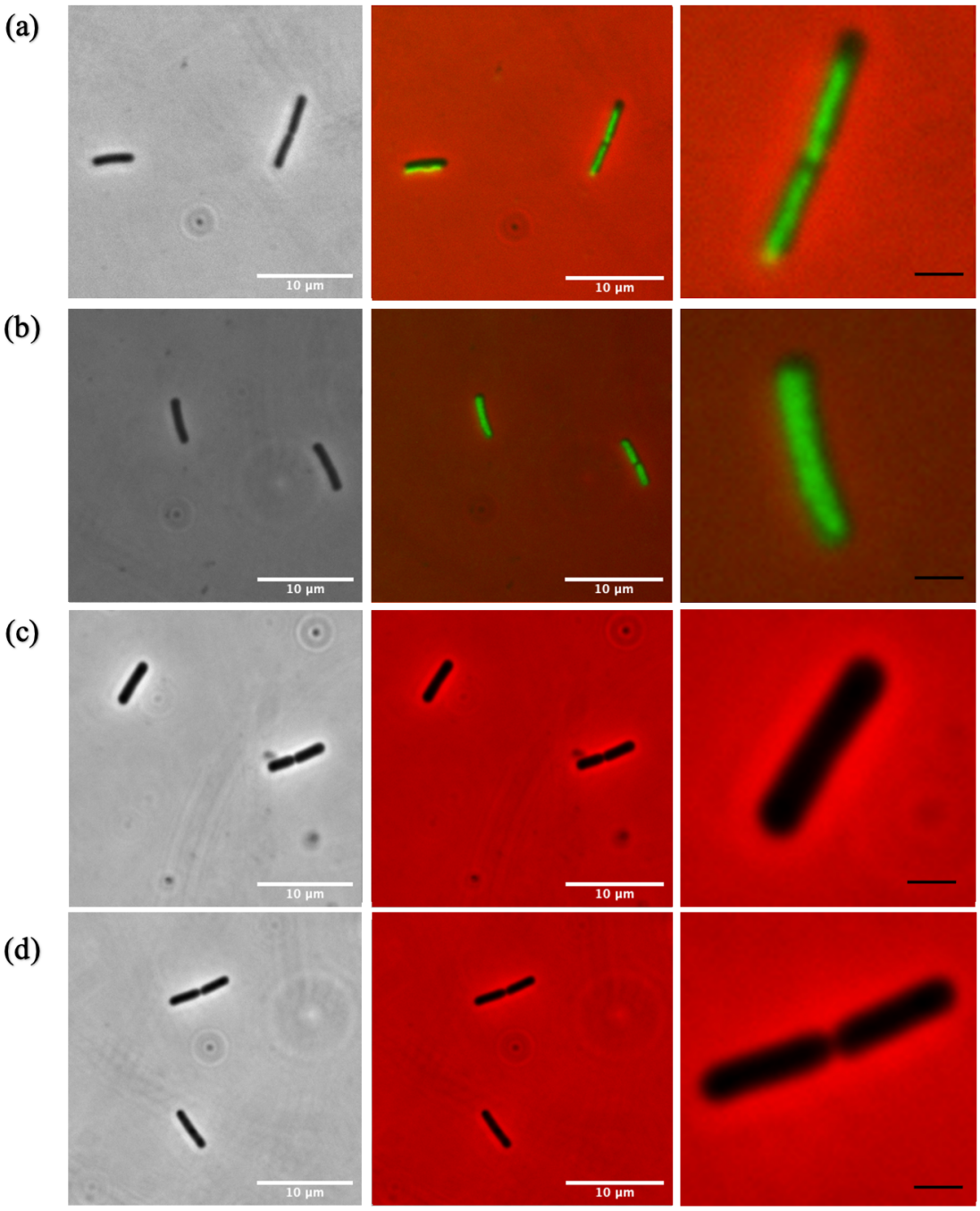


Figure 5.7 Ability LysCD6356-Alexa 488 to attach to vegetative bacterial form.

7.5 µM LysCD6356-Alexa 488 was incubated with vegetative form of a) *C. difficile* R20291, (b) *C. difficile* DS1787, (c) *B. cereus* ATCC 6464 and (d) *B. subtilis* ATCC 27370 for 20 minutes at 37º C. Cells were then washed three times and prepared for visualization. Cells were visualized under phase contrast microscopy (left). Using ImageJ, cell image was coloured with red and merged with the image of green coloured fluorescent dye of Alexa 488 (middle). A cell of merged image was magnified (right). Experiment was performed in duplicate, and images were captured from three different fields. Black scale bars length is 1 µm.

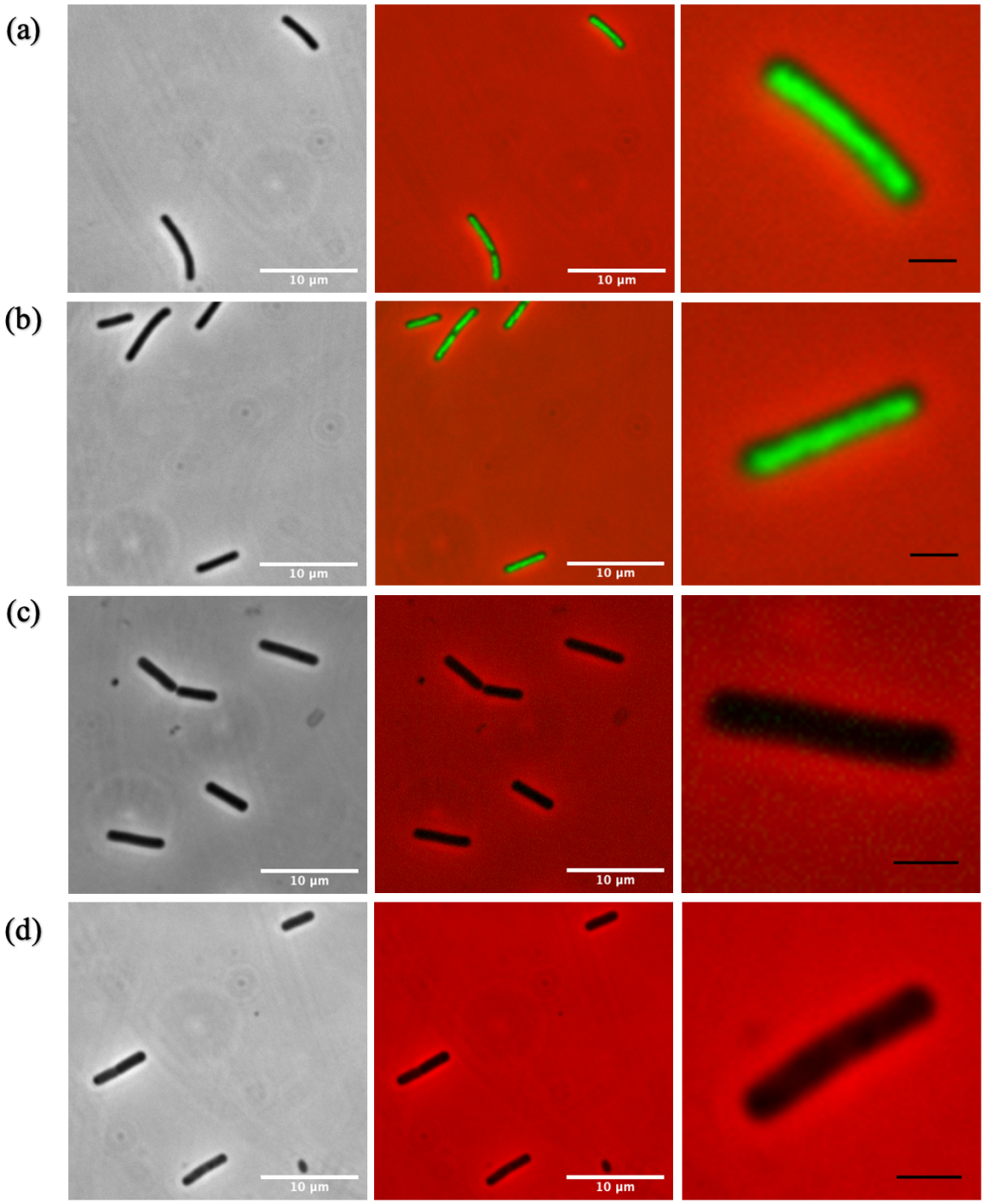


Figure 5.8 Ability EAD-Alexa 488 to attach to vegetative bacterial form.

7.5 µM EAD-Alexa 488 was incubated with vegetative form of a) *C. difficile* R20291, (b) *C. difficile* DS1787, (c) *B. cereus* ATCC 6464 and (d) *B. subtilis* ATCC 27370 for 20 minutes at 37º C. Cells were then washed three times and prepared for visualization. Cells were visualized under phase contrast microscopy (left). Using ImageJ, cell image was coloured with red and merged with the image of green coloured fluorescent dye of Alexa 488 (middle). A cell of merged image was magnified (right). Experiment was performed in duplicate, and images were captured from three different fields. Black scale bars length is 1 µm.

Based on spore and vegetative attachment results, the full length LysCD6356 binds to the spore and vegetative forms of *C. difficile* and to the spore form of one isolate of *B. cereus*, while EAD only binds to the spore and vegetative forms of *C. difficile*.

It may be the case that full length LysCD6356 being larger than EAD offers more opportunities for non-specific spore binding. The loss of these regions from EAD could account for the increased specificity. This hypothesis needs to be confirmed by assessing a wider collection of spore forming bacteria. Additionally, the ability of EAD to bind specifically to both forms of *C. difficile* provides a potential advantage if it were to be used as a diagnostic marker. Again, this needs to be confirmed using more isolates and the spore binding region would need to be identified.

5.3.3 Cloning and expression of recombinant versions of the exosporial proteins CdeC and CdeM

In this study, the binding of LysCD6356 and EAD to the surface of the spore and vegetative forms of *C. difficile* was demonstrated. The next step was to identify the structures on the spore surface to which the endolysins attached. To do this, we focused on two previously identified proteins, CdeC and CdeM, which are components of the exosporium and are known to be exposed to the environment.

While the conserved nature of these proteins across multiple isolates of *C. difficile* was reported in 2018 (Calderón-Romero et al., 2018), we performed an updated *in-silico* analysis to confirm their conservation within *C. difficile* (Appendix 7.1).

5.3.3.1 Introduction of plasmid constructs into an *E. coli* expression hosts

Each of the gene sequences was codon optimized (GenScript, Netherlands). Gene sequences were synthesized and inserted into pET-19b expression vectors, where each codon optimized gene was placed downstream of a 10 x Histidine affinity tag and enterokinase recognition and cleavage site

under the control of an inducible T7 promoter. Plasmids provided by GenScript were transformed into competent cells of *E. coli* SHuffle T7 using heat shock.

5.3.3.2 Recombinant protein expression of CdeC

In order to study the level of expression of CdeC, cell lysates (before and after induction) and fractions obtained using a Qiagen purification kit, including flow through fraction, washing fraction and elution fraction from each expression experiment, were analyzed using SDS-PAGE and Western blot. An SDS-PAGE gel showing the different fractions obtained from an induced culture of SHuffle T7 pET-19b-CdeC can be seen in Figure 5.9.

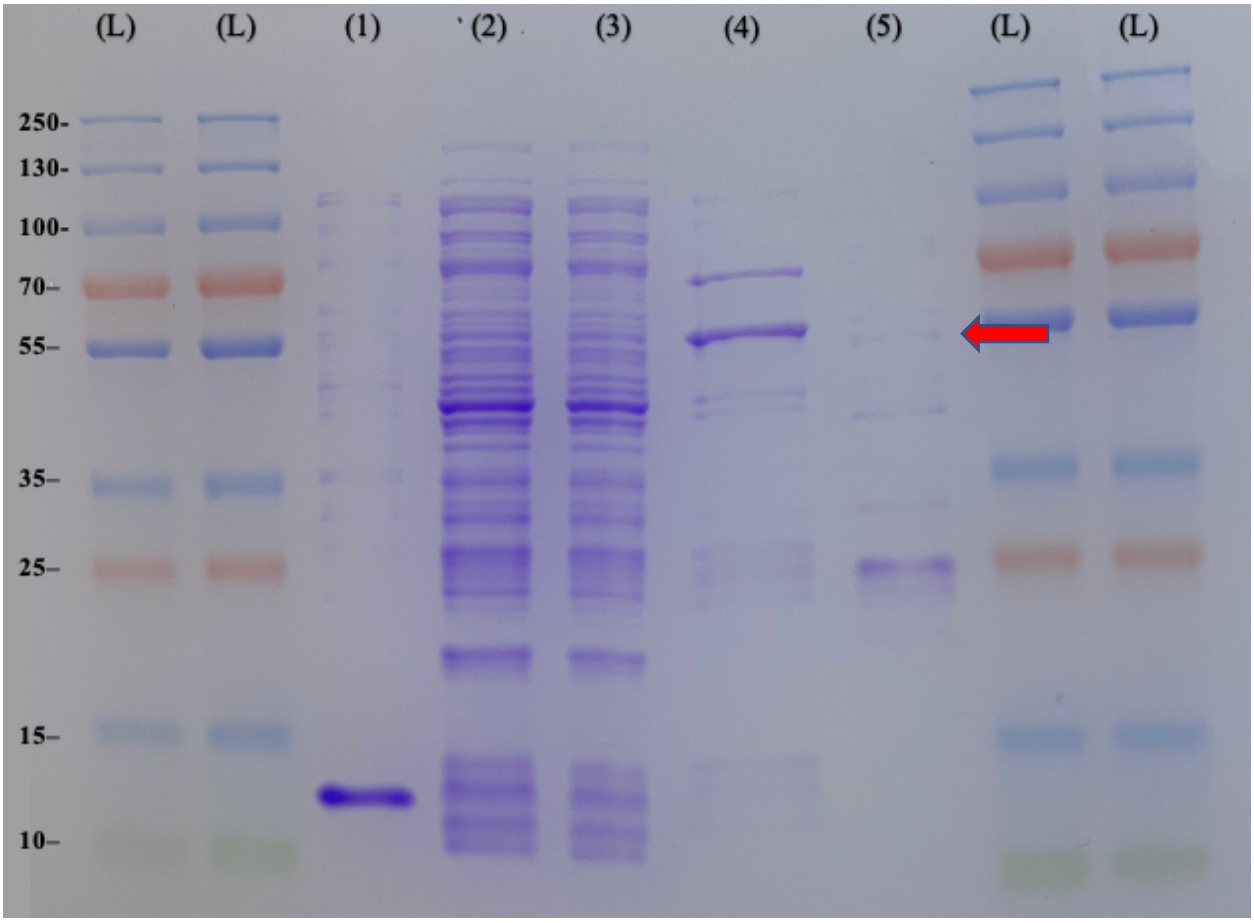


Figure 5.9 An SDS-PAGE gel showing different fractions following the induction and purification of a culture of *E. coli* SHuffle T7 CdeC expressing recombinant CdeC.

Once the culture of *E. coli* SHuffle T7 pET-19b-CdeC had reached mid log phase growth, cells were induced with 0.5 mM IPTG, and the culture was incubated at 21° C for 3 hours on a shaker incubator. Following protein purification using Qiagen purification kit, protein fractions were run on SDS-PAGE gel, which was then stained with Coomassie brilliant blue to visualize protein

bands. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) Elution fraction (250mM imidazole), and (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). 2 µg of protein were loaded from all fractions. Red arrow indicated the predicted size of the recombinant CdeC band (47.54 kDa).

To determine if the faint band in the elution fraction with an approximate size of 47 kDa was recombinant CdeC, Western blot analysis was performed. A clear band was detected, indicating the presence of the target protein, as can be seen in Figure 5.10. Other immunoreactive bands appeared on the gel (around 24 kDa), which may be the result of the degradation of CdeC.

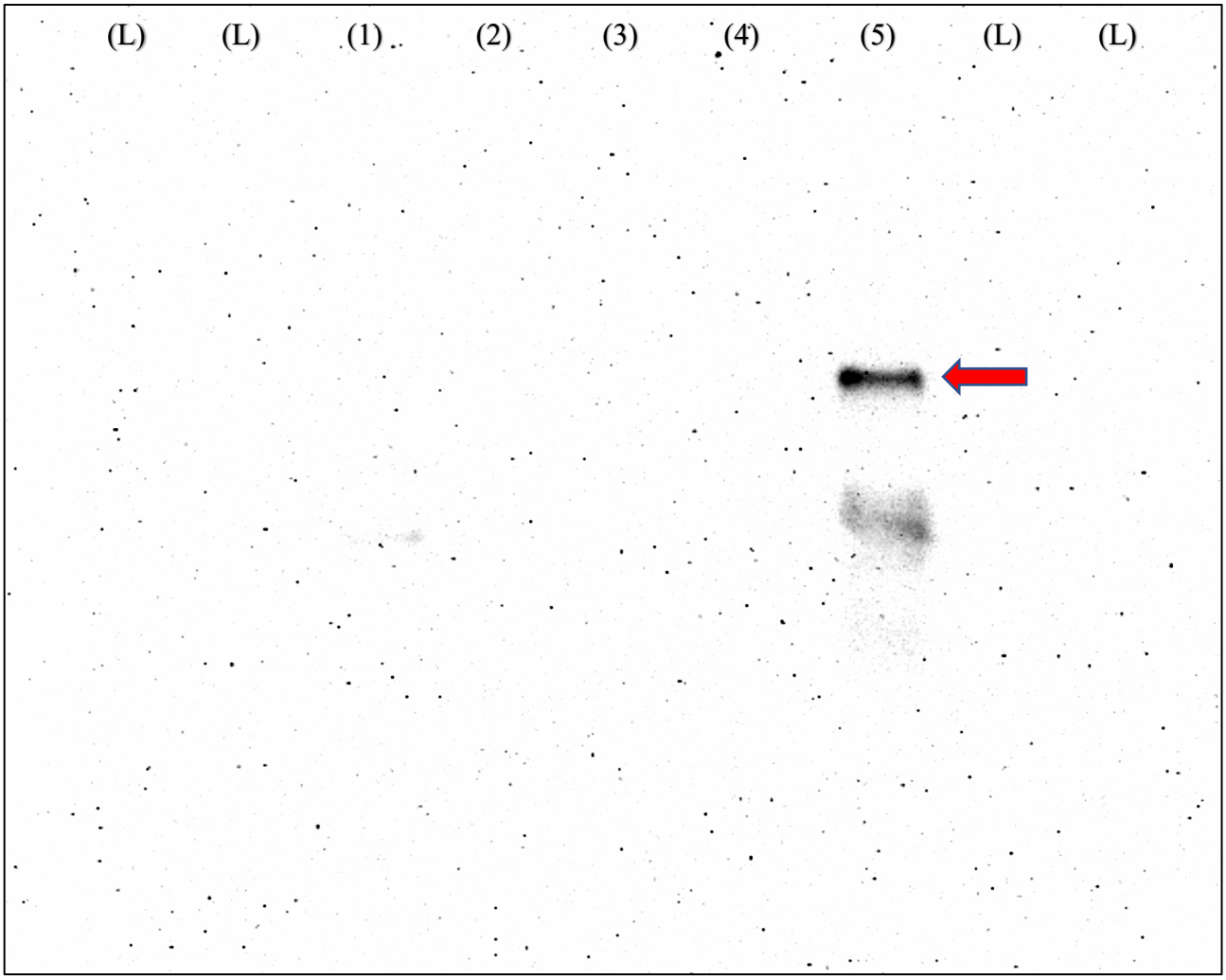


Figure 5.10 Western blot analysis of a PVDF membrane with different fractions of a *E. coli* SHuffle T7 pET-19b-CdeC culture expressing recombinant CdeC.

The membrane was stained with a histidine specific antibody and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (1) Pre-induction fraction, (2) post-induction

fraction, (3) Flow through fraction, (4) Wash fraction, (5) Elution fraction. (L) molecular wight ladder (PageRuler™ Plus Prestained Protein Ladder). The red arrow represents the predicted location of CdeC (47.54 kDa).

Using this host and purification approach, we achieved a recombinant protein yield of 243 µg/L with 6.3% purity.

To investigate the impact of using a different strain of *E. coli* as the host cell, the expression of recombinant CdeC from BL21 (DE3) was performed. A band was observed of the expected size in the elution fraction (Figure 5.11). Other bands can also be seen in the elution fraction, which may represent CdeC degradation products or possibly other contaminating proteins.

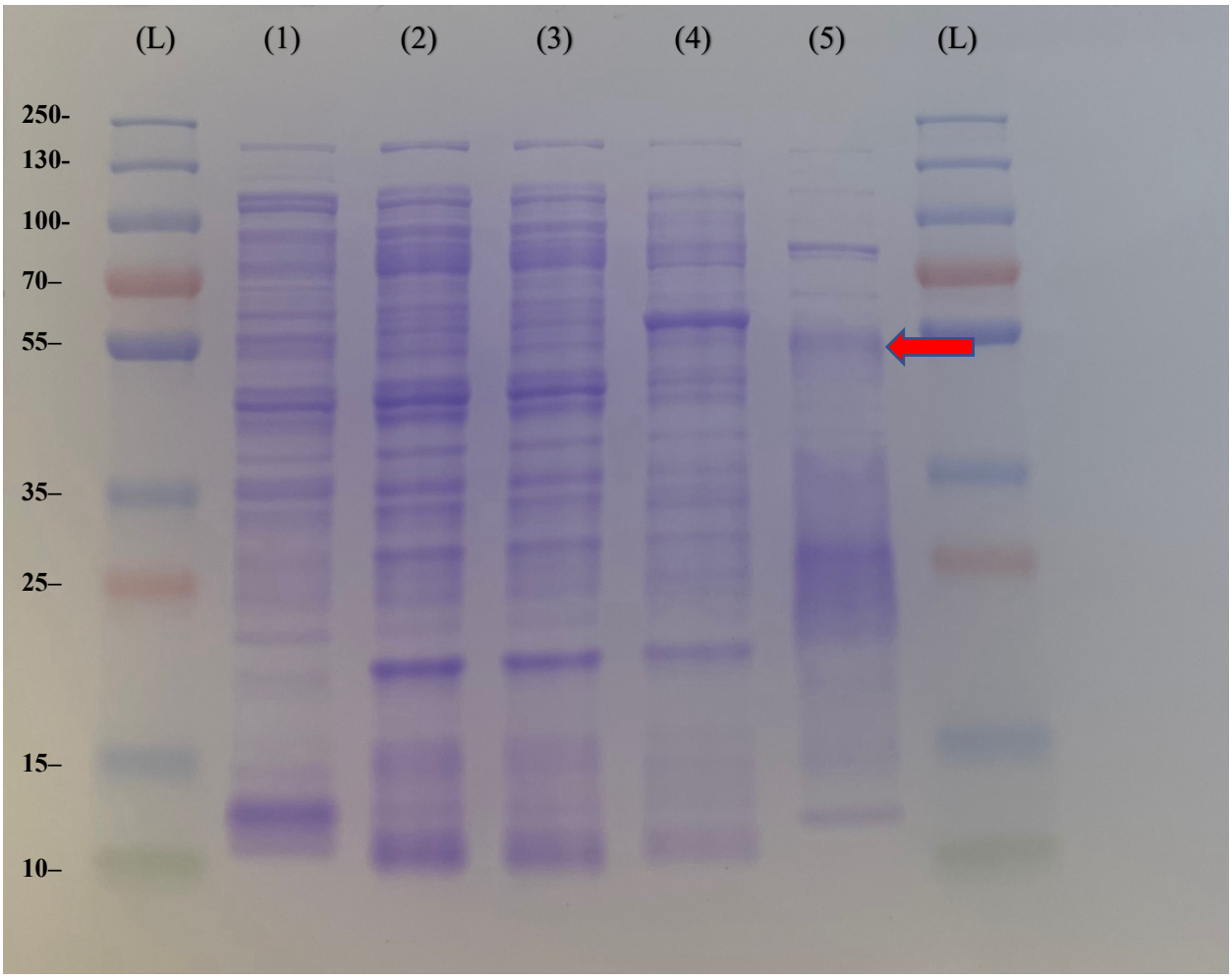


Figure 5.11 An SDS-PAGE gel showing different expression and purification fractions following the induction and purification of a culture of *E. coli* BL21(DE3) pET-19b-CdeC expressing recombinant CdeC.

The gel was stained with Coomassie brilliant blue to visualize protein bands. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) Elution fraction (250mM imidazole), and (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). 2 µg of protein were loaded from all fractions. Red arrow indicated the predicted size of the recombinant CdeC band (47.54 kDa).

To confirm the nature of this band, Western blotting was performed. As can be seen from Figure 5.12 a single band was identified, corresponding in size to the target protein. Interestingly, we only saw a single band, which may indicate less protein degradation using *E. coli* BL21 (DE3) pET-19b-CdeC compared to *E. coli* SHuffle T7 pET-19b-CdeC.

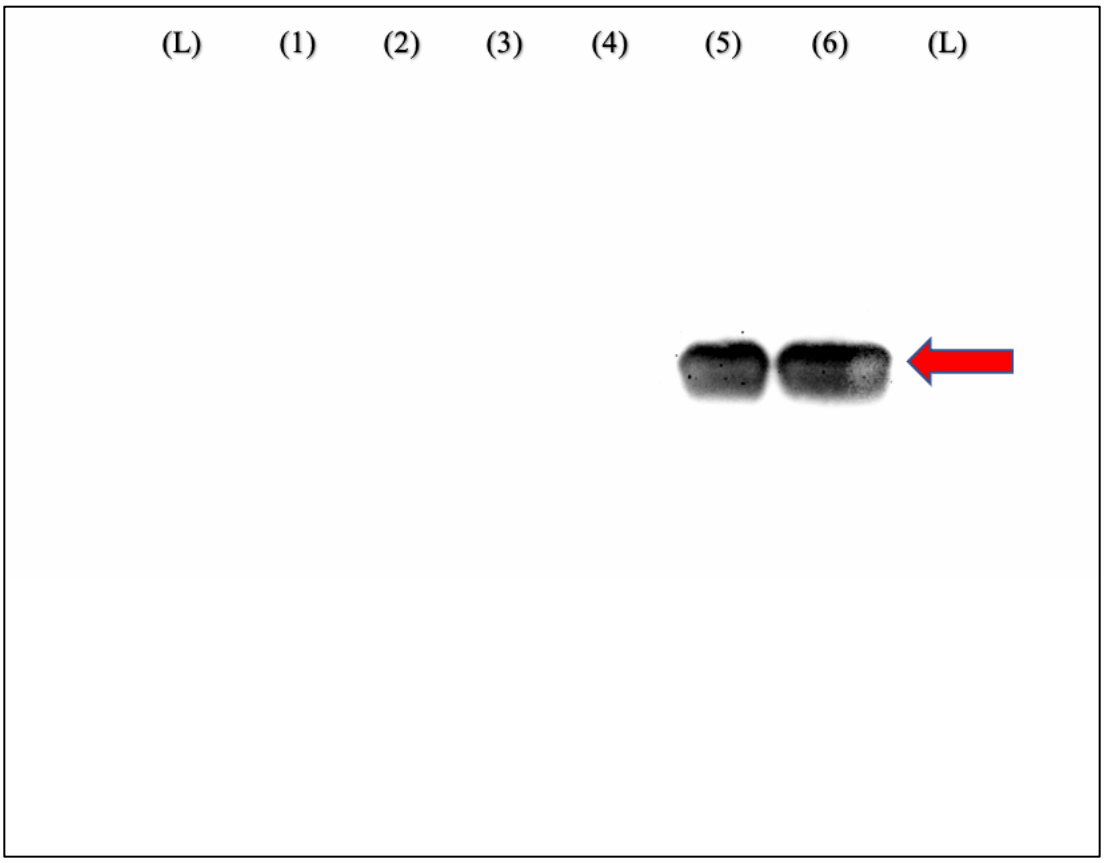


Figure 5.12 Western blot analysis of a PVDF membrane with different fractions of an *E. coli* BL21(DE3) pET-19b-CdeC culture expressing recombinant CdeC.

The membrane was stained with a histidine specific antibody and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) and (6) Elution fraction. (L) molecular wight ladder (PageRuler™ Plus Prestained Protein Ladder). The red arrow represents the predicted location of CdeC (47.54 kDa).

The production of recombinant CdeC from the different host strains in this project revealed that SHuffle T7 produced at least two bands of CdeC, which may result from the degradation of the protein of interest, while BL21(DE3) yielded a single band.

Using this host and purification approach, we achieved a recombinant protein yield of 519 µg/L with 14.8% purity. This level of expression was almost twice that achieved using the SHuffle T7, and for this reason the *E. coli* BL21 (DE3) host was used to support future studies.

5.3.3.3 Recombinant protein expression of CdeM

The expression of recombinant CdeM in different fractions was analyzed using SDS-PAGE and Western blot. Commassie blue straining failed to reveal the presence of a band of the expected size (22.03 kDa) in the elution fraction following induction of SHuffle T7 pET-19b-CdeM (Figure 5.13).

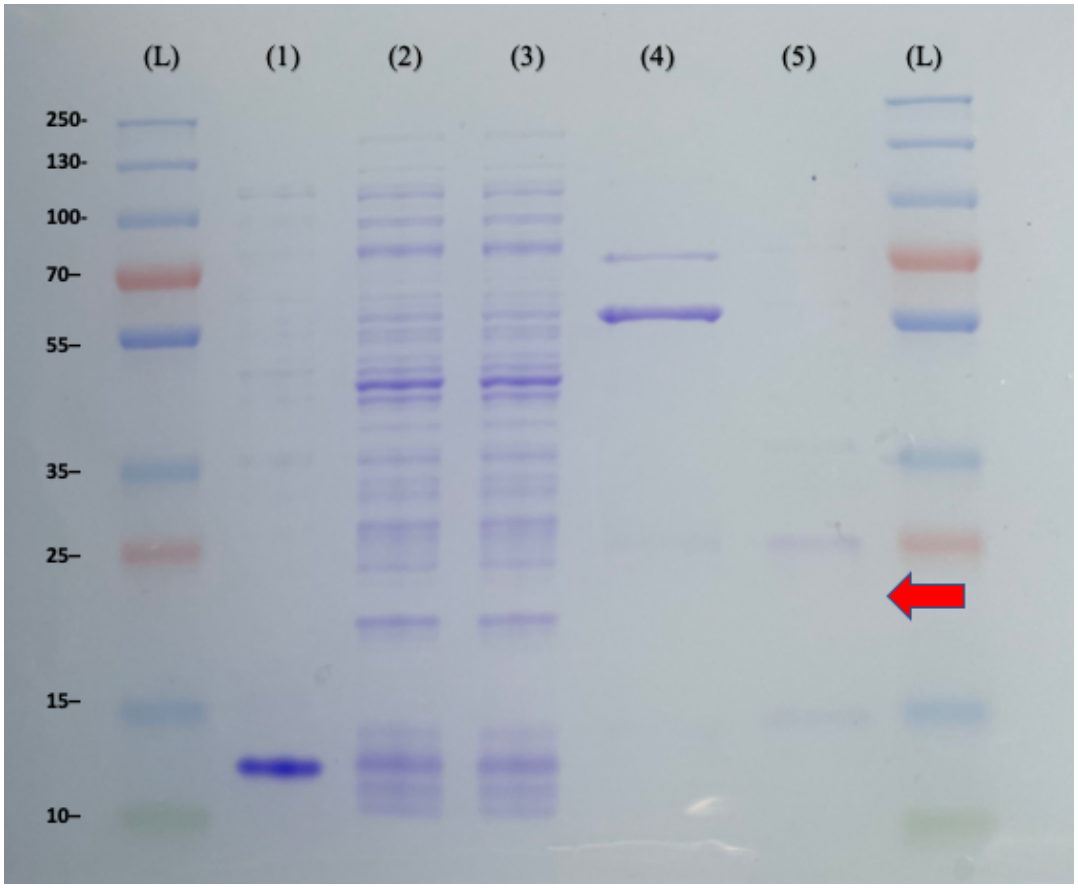


Figure 5.13 An SDS-PAGE gel showing different fraction following the induction and purification of a culture of *E. coli* SHuffle T7 pET-19b-CdeM expressing recombinant CdeM.

The gel was stained with Coomassie brilliant blue to visualize protein bands. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) Elution fraction (250mM imidazole), and (L) Molecular weight ladder (PageRuler™ Plus Prestained

Protein Ladder). 2 µg were loaded from all fractions. The red arrow indicates the predicted location of a recombinant CdeM band (22.03 kDa).

To determine if this failure to detect CdeM was due to a problem with protein expression or reflected a level of expression below the limit of detection of Coomaisse blue staining (100-500 ng) (Brunelle and Green 2014), we performed a Western blot. As can be seen in Figure 5.14, we detected a band of the expected size in the elution fraction, suggesting that the recombinant protein had been expressed but at a level lower than that of CdeC.

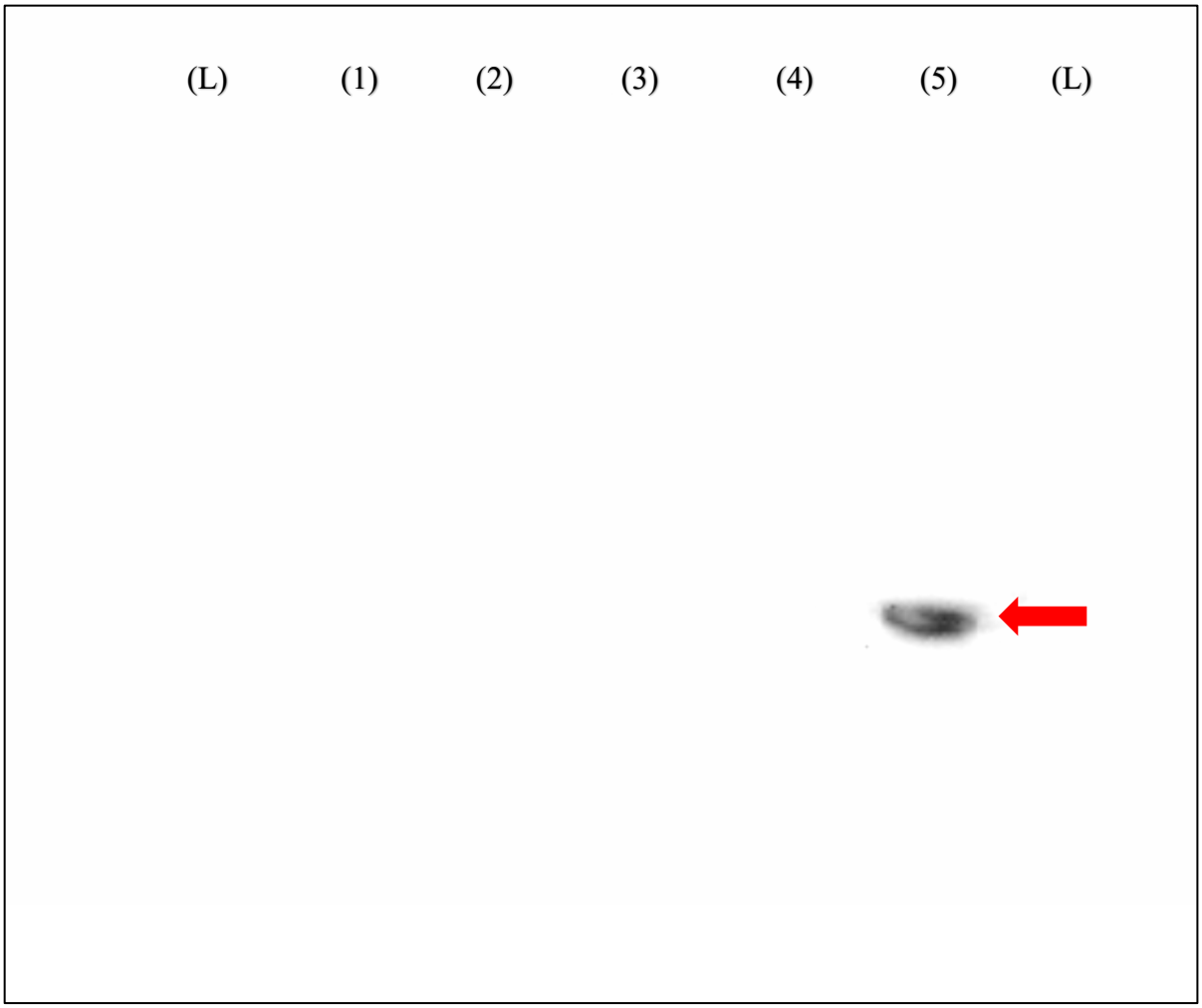


Figure 5.14 Western blot analysis of a PVDF membrane with different fraction of an *E. coli* SHuffle T7 pET-19b-CdeM culture expressing recombinant CdeM.

The membrane was stained with a histidine specific antibody and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) Elution fraction. (L) molecular wight

ladder (PageRuler™ Plus Prestained Protein Ladder). The red arrow represents the predicted location of CdeM (22.03 kDa).

To investigate the impact of using a different strain of *E. coli* as the host cell, the production of recombinant CdeM from BL21 (DE3) was performed. In contrast to the results obtained using the SHuffle T7 host, a band was observed of the expected size in the elution fraction on the PAGE-SDS gel (Figure 5.15).

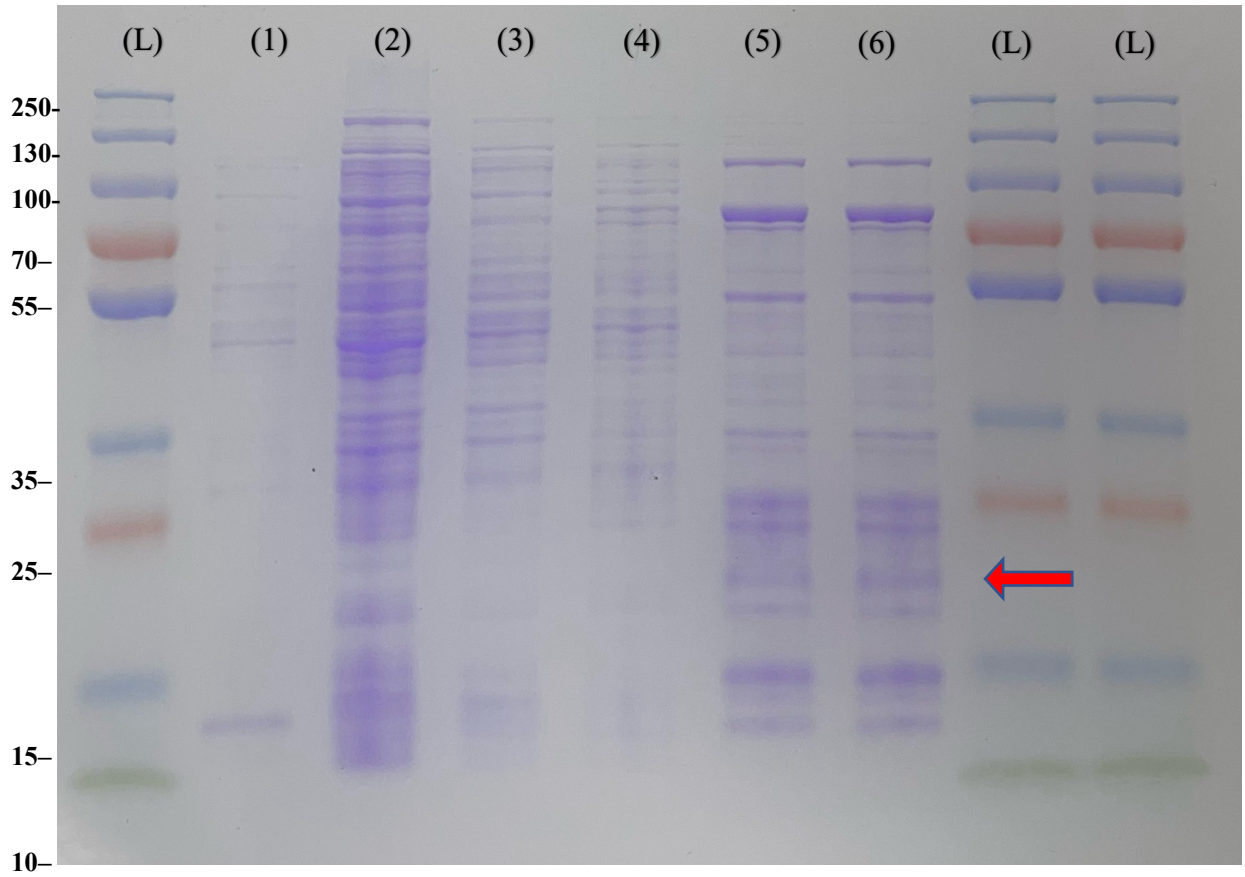


Figure 5.15 An SDS-PAGE gel showing different fraction following the induction and purification of a culture of *E. coli* BL21(DE3) pET-19b-CdeM expressing recombinant CdeM.

The gel was stained with Coomassie brilliant blue to visualize protein bands. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) and (6) Elution fraction (250mM imidazole), and (L) Molecular weight ladder (PageRuler™ Plus Prestained Protein Ladder). 2 µg were loaded from all fractions. The red arrow indicated the predicted size of the recombinant CdeM band (22.03 kDa).

Western blot analysis detected a band of the expected size in elution fractions (5 and 6), confirming that the band in the SDS-PSGE gel was the expressed recombinant protein (Figure 5.16).

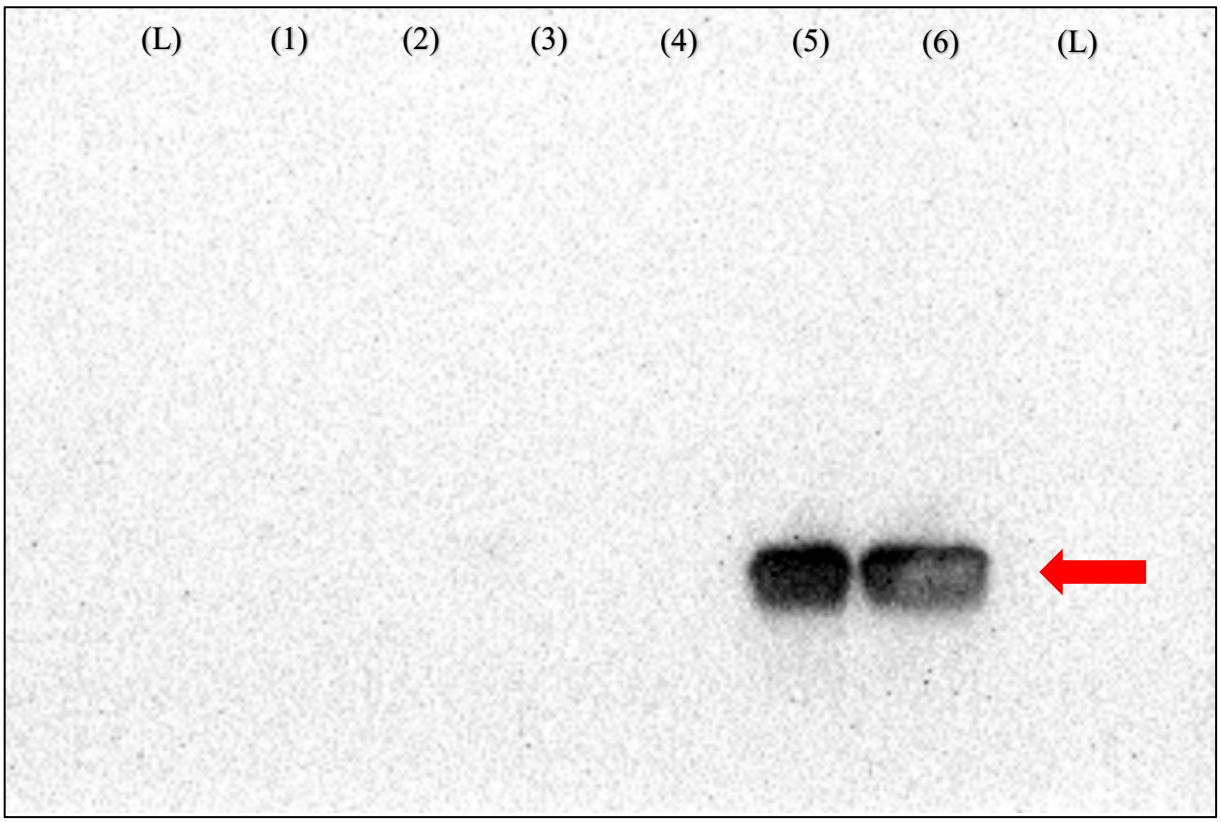


Figure 5.16 Western blot analysis of a PVDF membrane with different fractions of an *E. coli* BL21(DE3) pET-19b-CdeM culture expressing recombinant CdeM.

The membrane was stained with a histidine specific antibody and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. (1) Pre-induction fraction, (2) post-induction fraction, (3) Flow through fraction, (4) Wash fraction, (5) and (6) Elution fraction. (L) molecular wight ladder (PageRuler™ Plus Prestained Protein Ladder). The red arrow represents the predicted location of CdeM (22.03 kDa).

As was the case for CdeC, the yield of purified CdeM differed depending on the expression host. While SHuffle T7-CdeM was undetectable on SDS-PAGE, a clear band was detected using BL21 (DE3) and yielded 355 µg/L with 7.8% CdeM purity.

For this reason, BL21 (DE3) was chosen as the host for the production of recombinant CdeM in subsequent experiments.

5.3.4 Assessment of the ability of recombinant LysCD6356 and its EAD to bind to CdeC and CdeM

5.3.4.1 Western blot analysis

Aliquots of CdeC and CdeM were first run on a native PAGE, and then transferred to a PVDF membrane using a transblot turbo transfer system (Bio-Rad, UK) as described previously (Chapter 2, Section 2.2.14). To confirm protein transfer, Western blot analysis was performed using an anti-histidine antibody which recognized the histidine tag located at the N terminal of each recombinant protein (Figure 5.17).

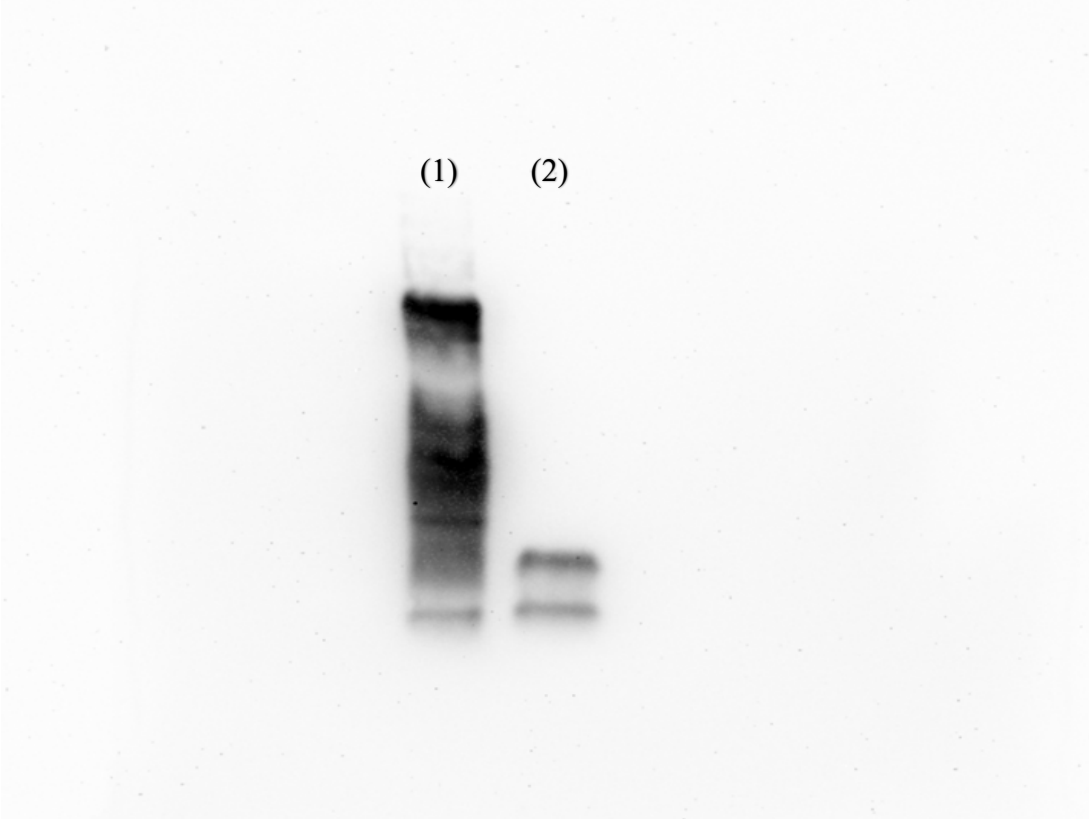


Figure 5.17 Western blot analysis of PVDF membrane following native-PAGE run for CdeC and CdeM

The membrane has been stained with a histidine specific antibody and a secondary goat anti-mouse detection antibody at a concentration of 1:200,000. The presence of the antibody complex was detected using a Chemiluminescence substrate. 7 µg were loaded of CdeC (1) and CdeM (2).

Native protein gels do not separate proteins based on their size only but rather on their size and charge in the folded state. Thus, in native SDS-PAGE, the run of non-reduced CdeC and CdeM appears to reveal the presence of different forms of what could be multimers or breakdown products.

5.3.4.2 Far western blot analysis

Following confirmation of the transfer of CdeC and CdeM to the PVDF membrane, we next probed loaded membranes with Alexa labeled LysCD6356 and EAD. While we were able to visualize the Alexa-labeled probes at concentrations of 224.6 and 145.7 μg/mL, respectively, we were unable to see any evidence of the probes binding to either protein (Figures 5.18 and 5.19).

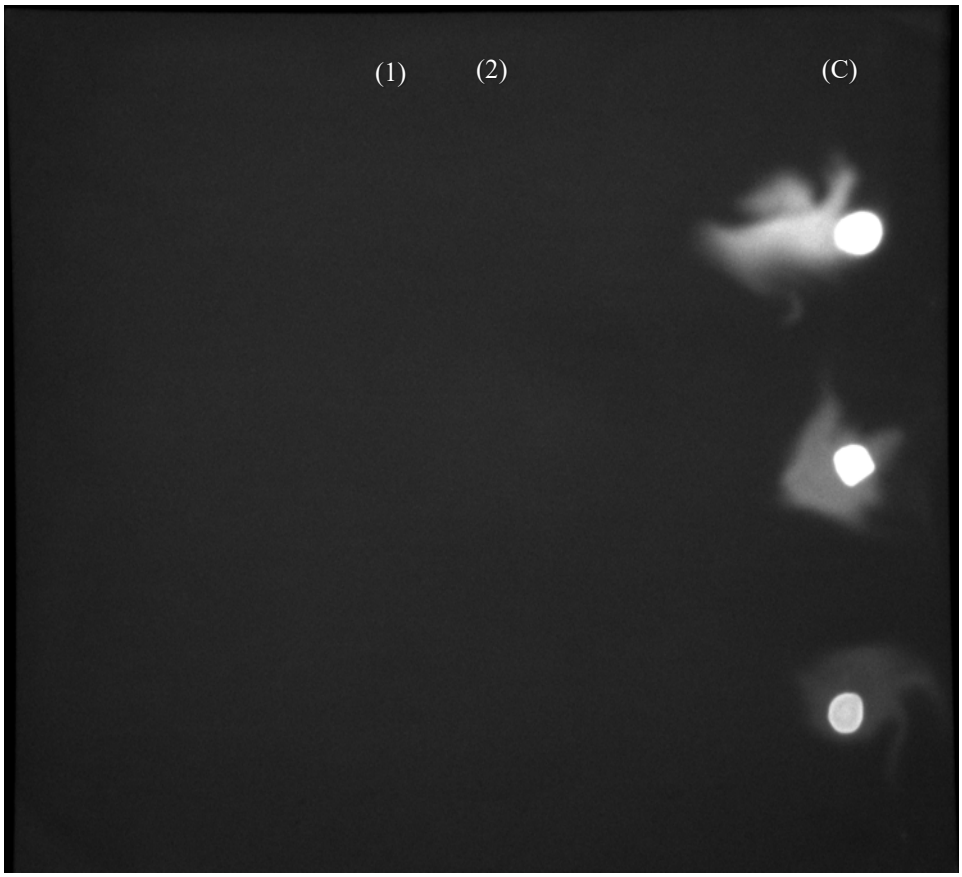


Figure 5.18 Assessing the ability of Alexa labelled LysCD6356 to attach to CdeC and CdeM using Far Western blot analysis.

CdeC (1) and CdeM (2) were run on a native–PAGE gel and transferred to a PVDF membrane. The membrane was exposed to blocking and washing steps and then was probed with Alexa488 labeled LysCD6356. The positive control (C) consisted of 5µl drops of different dilutions of the Alexa labeled protein in PBS (898.5, 449.2 and 224.6 µg/mL). Far Western blot was performed in triplicate.

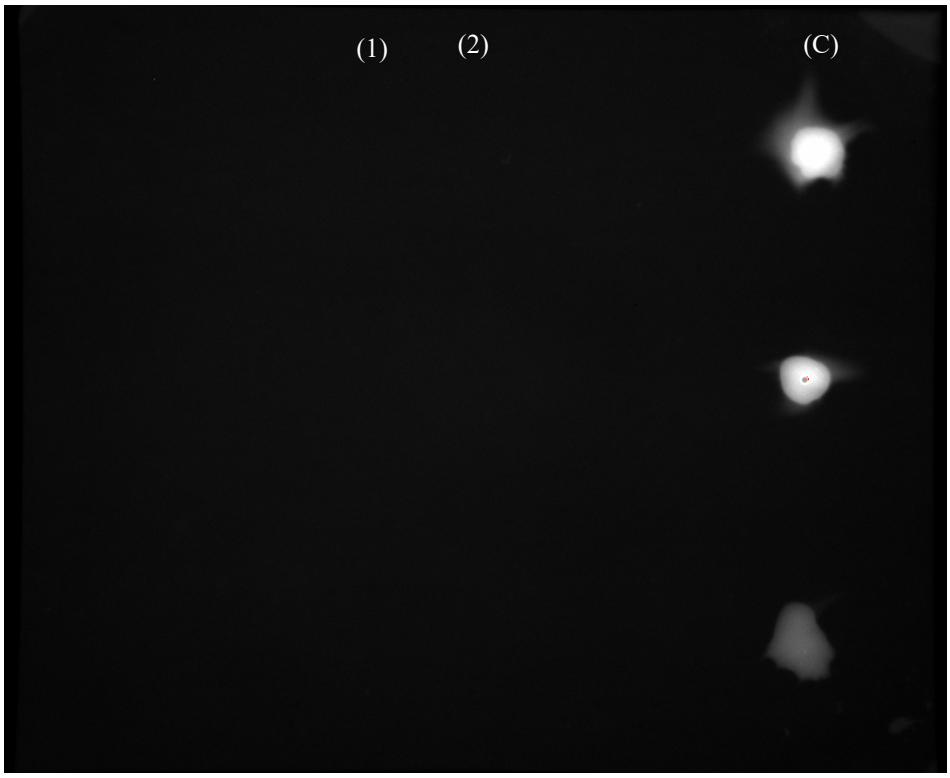


Figure 5.19 Assessing the ability of the alexa labelled EAD of LysCD6356 to attach to CdeC and CdeM using Far Western blot analysis.

CdeC (1) and CdeM (2) were run on a native–PAGE gel, transferred to a PVDF membrane. The membrane was exposed to blocking and washing steps and then was probed with Alexa488 labeled EAD. The positive control (C) consisted of 5µl drops of different dilutions of the Alexa labeled protein in PBS (582.9, 291.4 and 145.7 µg/mL). Far Western blot was performed in triplicate.

In preliminary studies, we were unable to demonstrate binding of the Alexa labeled endolysins to recombinantly expressed CdeC and CdeM. This may be due to the fact that the assay requires further optimization (the lacke an appropriate positive control), that the concentration of surface proteins was not sufficient for detection, or that the endolysin does not bind to either protein, suggesting that they bind to other targets on the spore surface. Further work will be required to address this issue.

5.4 Discussion

The aims of this chapter were to determine if the endolysin of CD6356 or its EAD were able to attach to the *C. difficile* spore surface. By using Alexa fluorophore labeled recombinant endolysin, binding to the vegetative and spore forms of *C. difficile* was investigated. The full length lysin and its EAD were able to bind to the surface of the vegetative forms of the two *C. difficile* isolates examined in this study but failed to attach to the vegetative forms of the other *Bacillus* spp. isolates, suggesting a degree of specificity. The ability of a *C. difficile* specific endolysin, CD27L, to bind to the vegetative form of *C. difficile* had previously been reported and given that this lysin shares a degree of sequence and structural homology with LysCD6356, it is tempting to speculate that they bind to similar targets.

Using AlphaFold, The alignment of LysCD6356 structure and CD27L demonstrated high structural alignment (RMSD = 1.202), which was not the case with CTP1L were the structural alignment was low (RMSD = 12.643). The protein sequrence alignment and the structural alignment investigation suggests that LysCD6356 and CD27L might sharing a region that is responsible for the binding of endolysin to the vegetative *C. difficile.* Similarly, The structure of EAD and $CD27L_{1-179}$ were compared using a bioinformatic based approach. The significant level of similarity in protein sequence alignment of EAD to $CD27L_{1-179}$ (E-value $5e^{-32}$) and the high structural alignment similarity (RMSD = 0.824) suggest that they may share common motifs.

On the other hand, spore binding analysis demonstrated that LysCD6356 was able to bind to *C. difficile* strains and *B. cereus* ATCC 6464 but not to spores of *B. subtilis* ATCC 27370, which lacks exosporium layer, suggesting that the endolysin recognizes structures present in, or attached to, the exosporium. The EAD of LysCD6356 bound only to *C. difficile* strains suggesting greater specificity than the full length lysin.

To identify potential spore attachment regions in EAD of LysCD6356, an *in silico* analysis was performed. The amino acid comparison to the known spore binding region in PlyG demonstrated no significant similarity. In addition, the percent identity demonstrated low protein sequence similarity with PlyG (13%) and structural investigation showed low similarity (RMSD = 7.97),

suggesting that EAD lacked homolog with the spore binding motif found in PlyG, suggesting that EAD may bind to spores via an unreported sequence. Further investigations would be required to characterize the phenomenon.

In an effort to identify potential spore binding targets, two major structural components of the exosporium of *C. difficile*, CdeC and CdeM, which are surface exposed, were investigated (Barra-Carrasco et al., 2013; Ghose et al., 2016; Calderón-Romero et al., 2018). By performing an *in silico* analysis, proteins of CdeC and CdeM appear to be conserved in *C. difficile*, confirming a previously published analysis (Calderón-Romero et al., 2018). However, the expression of recombinant forms of CdeC and CdeM is challenging (Barra-Carrasco et al., 2013). To date, CdeC and CdeM have been expressed in different *E. coli* strains and from different vectors in an effort to enhance expression levels and solubility. Two strains, BL21(DE3)Ril and SHuffle T7, were reported to yield enhanced levels of soluble protein from a non-codon optimized version of CdeC by lowering the incubation temperature to 21 ºC (Brito-Silva et al., 2018). The yield of expressed protein was not reported. Ghose and colleagues utilized BL21 (DE3) carrying pET19b-CdeC (codon optimized) or pET19b-CdeM (codon optimized) expression plasmids and were able to achieve yields ranging from 500-700 mg/L (Ghose et al., 2016). The expressed levels of CdeC and CdeM were higher than those achieved in present study which may reflect differences in the efficiency of commercial codon optimization algorithms (Menzella 2011).

The BL21(DE3) host is a commonly used *E. coli* host strain as it is a T7 and protease deficient strain. The BL21(DE3)Ril host offers further advantages over BL21(DE3) in that it contains transfer RNA molecules which recognize codons commonly found in Gram positive bacteria such as *C. difficile*. SHuffle T7 is used to support the formation of disulfide bonds in cysteine rich proteins and enhance the proper folding of these proteins (Lobstein et al., 2012; Brito-Silva et al., 2018). In present study, codon optimized version of CdeC and CdeM gene sequences and two *E. coli* host strains, BL21 (DE3) and SHuffle T7 were used. The variation of expression levels between the host strains can be due to the proteases deficient nature of BL21 (DE3) (*Lon*$^{-}$ and *OmpT*$^{-}$) compared to SHuffle T7. Therefore, BL21(DE3) was selected as the *E. coli* host for the production of the recombinant CdeC and CdeM, which were used in our lysin binding studies.

We attempted to determine the ability of Alexa labelled endolysin to bind to these proteins using native PAGE in an effort to retain the tertiary protein structure which would be encountered in the spore. Following separation on the native gel several bands were observed which reacted with the anti-histidine antibody, suggesting that each protein under non-denaturing conditions might form multimers. Multimeric forms of cysteine rich proteins such as CdeC and CdeM have been reported previously (Romero-Rodríguez et al., 2020). Additionally, the phenomenon was also observed with other proteins such as ExsY, an exosporial protein of *B. cereus*, and CotY, a spore coat protein of *B. subtilis* (Jiang et al., 2015; Terry et al., 2017). In this study, we were unable to demonstrate the binding of Alexa labeled LysCD6356 and its EAD to CdeC or CdeM. It may be the case that the experimental conditions require further optimization, the concentration of surface proteins was not sufficient for detection, or reflect the fact that the endolysins do not bind to CdeC or CdeM but to other spore surface exposed proteins. Further work will be required to address this issue.

In conclusion, as hypothesized, LysCD6356 and its EAD were able to attach to the spore and vegetative forms of *C. difficile*. Surprisingly, the truncated version of LysCD6356 demonstrated a higher level of specificity for *C. difficile* spores, which might be a consequence of being a smaller protein and hence reducing the potential for non-specific binding. Additionally, an *in silico* investigation of the homology of the EAD of LysCD6356 to the catalytic region of $CD27L_{1\text{-}179}$, which binds to vegetative *C. difficile*, revealed high similarity, suggesting the possibility of a common motif responsible for the binding of the protein to vegetative *C. difficile*. In contrary, the homology analysis of the EAD of LysCD6356 to PlyG which binds to spores of *B. anthracis*, revealed low similarity, suggesting that EAD may contain novel spore binding motifs. Further studies are needed to determine the regions involved. In addition, the finding that EAD is bound to all forms of *C. difficile* would suggest that it recognizes a target or targets common to both forms of the bacterium, that it pass through the exosporium and bind to the coat layer, or that it contains more than one binding region. Further studies are needed to address these questions. Two highly conserved exosporial surface exposed proteins, CdeC and CdeM, failed to bind Alexa labeled endolysins. This could reflect deficiencies with the experimental approach or the fact that the endolysin recognizes other surface targets.

Chapter 6 General discussion

6.1 Introduction

In this study, we assessed the ability of a bacteriophage-derived endolysin, LysCD6356 and its EAD, to target clinical isolates obtained primarily in the UK. This lysin, which was isolated from CD6356, a temperate phage of the Siphoviridae family, was previously demonstrated in our laboratory to be more active than CD27L against our collection of isolates (Horgan et al., 2010; Alyousef 2013).

We also undertook a preliminary assessment of the ability of this lysin in combination with germinants (taurocholate and glycine) and calcium to target the spore form of the pathogen with a view to explore the feasibility of developing a treatment regimen that minimizes the potential for recurrent infections (Budi et al., 2020).

Finally, we investigated the ability of endolysins to bind to *C. difficile* spore surface. Specific binding of the lysin to the spore surface would increase protein localization at the site of infection and might also be used as a diagnostic tool.

6.2 Selection and expression of endolysins

The aim was to use the most active endolysin against a collection of clinical isolates of *C. difficile*. As reported in the literature, the CD27L endolysin, which was isolated from *C. difficile* NCTC 12727, demonstrated activity against a wide range of clinical isolates (Mayer et al., 2008b). In 2013, a comparison of the activity of a range of endolysins, including CD27L, LysC2, LysCD119, LysCD6356, and LysCD38-2 found that LysCD6356 was the most active endolysin against a collection of clinical isolates of *C. difficile* (Alyousef 2013). Similar to CD27L, LysCD6356 retained its activity across a wide range of pH values (4 - 9), suggesting that it should retain its activity in the gastrointestinal tract (GIT) (Alyousef 2013). For these reasons LysCD6356 was chosen to be recombinantly expressed and employed in this study.

The levels of expressed recombinant LysCD6356 and EAD achieved in this study were comparable to previously reported levels (Alyousef 2013; Furlon et al., 2021). The small

differences in yields could be due to variations in codon optimization or the use of different host strains of *E. coli* (Alyousef 2013; Mauro 2018).

6.3 The level of lytic activity of recombinantly expressed endolysins against clinical isolates of *C. difficile*

Previous studies reported that EADs are more active than the full length lysin (Mayer et al., 2011; Wang et al., 2015). We were able to confirm that EAD was more active than LysCD6356 on a mass basis: however, the difference in activity cannot be confirmed without investigating the molar basis.

Visualization of the lytic activity of EAD against vegetative forms of *C. difficile* using a scanning electron microscope revealed time dependent surface disruption (Fischetti 2005; Schuch et al., 2019; Zhang et al., 2022). We also observed that exposure of the vegetative form to germinants (0.1% taurocholate and 50 mM glycine) and calcium in the absence of EAD, resulted in a significant reduction in the length of vegetative *C. difficile* from 4.46 μm (±1.05) to 1.73 μm (±0.28). During the course of infection, vegetative *C. difficile* encounters multiple stresses in the gastrointestinal tract, including exposure to bile salts and high osmolarity (Chowdhury et al., 1996; Begley et al., 2005b). Bile salts have detergent properties that can modify the bacterial cell wall in a manner which reduces the organism's osmotic tolerance (Oberkampf et al., 2022). It may be the case that the reduction in size was due to the combination of bile-salt-induced cell wall disruption and osmosis-driven leakage of intracellular contents due to the relativity high concentration of extracellular calcium.

An investigation of the role of pH on the activity of the EAD of LysCD6356 revealed that, similar to the full length lysin EAD, it retained its activity across the same range of PH values (4 to 9), suggesting it should retain its activity in the GIT. Published literature reported that several *C. difficile* endolysins remained active across a wide range of pH (Mayer et al., 2008b; Alyousef 2013; Wang et al., 2015; Peng et al., 2019; Mondal et al., 2021).

EAD, similar to $CD27_{1}$-179, was found to be zinc dependent. The high structural similarity between the three endolysins suggests that EAD also belongs to the *N*-acetylmuramoyl-L-alanine amidase class of endolysins (Ishikawa et al., 1999; Mayer et al., 2011; Vermassen et al., 2019).

Another factor that influenced EAD activity in a concentration dependent manner was the presence of calcium. A similar negative impact on enzymic activity was previously reported for PlyCD, an endolysin isolated from *C. difficile* 630 that shares 77.5% identity with EAD (Wang et al., 2015; Furlon et al., 2021). This is relevant because free calcium is naturally availabile in the GIT, although EAD retained its ability to lyse vegetative bacteria at calcium levels exceeding those naturally present in the GIT (Barrett et al., 2015; Kochan et al., 2017).

Even though the vegetative cells of *C. difficile* were reduced in size due to the action of germinants and calcium, SEM analysis demonstrated no evidence of cell disruption, suggesting that *C. difficile* resist the cell disruptive effect of germinants and calcium, as indicated by the severity of the infection in the availability of taurocholate and calcium. On the contrary, the wide range of activity of EAD of LysCD6356, the bacterial cell wall lysing effect, the activity across a wide range of pH and the minimal effect on gut microflora provide a competitive treatment tool to target *C. difficile* and became a field of interest because of the increasing antibiotic resistance.

6.4 Targeting of *C. difficile* spores

The spore form of *C. difficile* is responsible for the transmission and recurrence of CDI due to its ability to survive harsh environments combined with its resistance to antibiotics, bacteriophages, and endolysins. A strategy to eliminate spores consists of exposing them to compounds that trigger germination, followed by an antimicrobial agent which targets the emerging, sensitive vegetative bacterium. In one such example, Celebi and colleagues exposed *B. anthracis* spores to a combination of germinants (100 mM L-alanine and 5 mM inosine) and an antimicrobial agent (5000 ppm paracetic acid) separated by 1 hour and reported that they were able to reduce spore numbers (Celebi et al., 2016b).

Two approaches were employed in an effort to maximize spore germination and subsequent lysin induced killing. In the first, spores were exposed to germinants and lysin at the same time, while in the second, spores were incubated in the presence of germinants for 2.5 hours prior to the addition of lysin. The simultaneous addition of germinants and lysin failed to reduce spore numbers. Possible reasons for this failure to reduce the viability of the organism could be due to the negative impact of spore associated calcium on EAD activity, as described in chapter 4 section 4.4.7. Another possibility is that the newly germinated bacterium is relatively resistant to lysis. It has previously been reported that newly germinated spores of the Sterne strain of *B. anthracis* required at least 2 hours to develop sensitivity to lysin-mediated disruption (Mehta et al., 2013).

This might explain why we observed a reduction in *C. difficile* viability when the application of germinant followed by EAD was delayed by 2.5 hours. Following incubation, the addition of 60µg/mL EAD reduced the OD_{600} of the germinated spores significantly (P value < 0.001) in a similar pattern as the control over the first 15 minutes. The curve then plateaued, suggesting that not all of the germinated spores were sensitive to EAD, as previously reported (Mehta et al., 2013).

Thus, by adding germinants, we were able to induce spore germination. Delaying exposure of newly germinating spores to endolysins demonstrated to be a better approach for reducing bacterial numbers than simultaneous exposure to spore germinants and endolysins.

6.5 Binding endolysins to *C. difficile* spores

The ability of endolysins to bind to the spore and vegetative forms of bacteria has been reported (Mayer et al., 2011; Yang et al., 2012; Gómez-Torres et al., 2018; Kong et al., 2019). For this reason, the ability of LysCD6356 and its EAD to bind to the spore and vegetative forms of *C. difficile* and related bacteria was assessed. To the best of our knowledge, this is the first investigation of endolysin binding to both forms of *C. difficile*. While EAD failed to bind to the spore forms of *B. cereus* ATCC 6464 and *B. subtilis* ATCC 27370 spores, it was able to bind to *C. difficile* spores, suggesting greater specificity than full length lysin, which demonstrated binding capability to *B. cereus* spores in addition to *C. difficile* spores. The binding of the full length

LysCD6356 may be a consequence of its larger size compared to EAD, which provides more area for non-specific binding. Further analysis will be required to confirm this hypothesis.

Endolysins are being explored as potential diagnostic tools due to a combination of factors, including speed, specificity, and cost, when compared to approaches such as antibodies, PCR and culture. Researchers have mainly been investigating the cell binding domain of the lysins (CBD) due to their high specificity (Santos et al., 2019). *C. perfringens* was reported to be detected in 5 minutes using the CBD of LysCPAS15 linked to an enhanced fluorescent green protein reporter (Cho et al., 2021). Additionally, *B. anthracis* has been detected using a 10 amino acid motif from the CBD of PlyG (Fujinami et al., 2007; Sainathrao et al., 2009). In the present study, the EAD of LysCD6356 was able to bind to both forms of *C. difficile* (R20291 and DS1787) and may have potential as a future diagnostic tool for *C. difficile*.

In an effort to identify the lysin binding targets on the spore surface, the ability of the full length lysin and EAD to bind to recombinantly expressed major structural components of the *C. difficile* exosporium, CdeC and CdeM, was assessed (Barra-Carrasco et al., 2013; Ghose et al., 2016; Calderón-Romero et al., 2018). Surprisingly, neither protein in laboratory studies was bound to LysCD6356 or EAD. This failure could reflect issues with the experiment methods or the endolysins attached to different spore surface structures. Further work will be required to address this issue.

Finally, using a bioinformatic approach based on previously published studies, we attempted to identify specific surface binding regions within LysCD6356 and its EAD by comparing them to the structure of endolysins known to bind to bacteria, including *C. difficile* (Mayer et al., 2011; Yang et al., 2012; Gómez-Torres et al., 2018; Kong et al., 2019). The protein sequence of LysCD6356 demonstrated significant similarity with CD27L (51% identity and $6e^{-85}$ E-value), suggesting that they may share common motifs. Similarly, EAD shared significant similarity (37.4% identity and $5e^{-32}$ E-value) with $CD27_{1\text{-}179}$, suggesting the presence of a similar vegetative binding region.

6.6 Suggested future work

6.6.1 Binding specificity of endolysins

Further investigation of the specificity of EAD of LysCD6356 is required to determine its ability to bind to range of *C. difficile* isolates. This is particularly important given the genetic diversity of the species. Studies should also be undertaken into the specificity of the lysin with regards to other bacterial species as this will determine the suitability of the EAD of LysCD6356 as a diagnostic tool for the spore and vegetative forms of *C. difficile*.

6.6.2 Fully characterize lysin activity against vegetative and spore forms of *C. difficile*

Further studies should be undertaken to identify the optimum conditions under which newly germinated spores become sensitive to the lysin treatment. Pandey and colleagues studied the germination and outgrowth of *Bacillus subtilis* spores through time and found that while spore germination was rapid, subsequent outgrowth to form the biologically active vegetative form took up to five hours (Pandey et al., 2013). Similar studies are required to identify the optimum post germination period for lysin treatment of *C. difficile.*

6.6.3 Formulation and timing of delivery

There are several approaches that could be used to deliver germinants and lysin to the site of infection in the GIT. The lysin could be delivered separately from the germinants to ensure the highest possible activity using an enteric coated capsule with proteolytic enzyme inhibitors to protect the lysins from stomach acidity and proteolytic activity in the GIT, respectively. The feasibility of this approach was previously reported by using an enteric coated capsule to perform oral FMT, which reported non-inferior results of preventing rCDI compared to enema (Jiang et al., 2018).

Another possible approach would be utilizing a genetically engineered member of the bacterial flora, such as *Lactococcus lactis,* to express EAD of LysCD6356. The approach has been used to express active endolysin, such as CD27L (Mayer et al., 2008b; Guo et al., 2015; Van Pijkeren and Barrangou, 2017).

6.6.4 Combination of treatment approaches

A combination of our strategy of spore targeting with other treatment options, such as fecal microbiota transplant, could potentially further reduce the incidence of recurrent CDI. One could also explore potential synergistic interaction between the EAD of LysCD6356 and first line antibiotics, the feasibility of which, has been demonstrated for other endolysins such as Cpl-1 and $PlyCD_{1-174}$ (Djurkovic et al., 2005; Wang et al., 2015).

References

Akash, M.S.H., Rehman, K., Tariq, M. and Chen, S. 2015. Development of therapeutic proteins: Advances and challenges. *Turkish Journal of Biology* 39(3), pp. 343–358. doi: 10.3906/biy-1411-8.

Alabdali, Y.A.J., Oatley, P., Kirk, J.A. and Fagan, R.P. 2021. A cortex-specific penicillin-binding protein contributes to heat resistance in Clostridioides difficile spores. *Anaerobe* 70, p. 102379. doi: 10.1016/j.anaerobe.2021.102379.

Alves, A., Bassot, A., Bulteau, A.L., Pirola, L. and Morio, B. 2019. Glycine metabolism and its alterations in obesity and metabolic diseases. *Nutrients* 11(6), p. 1356. doi: 10.3390/nu11061356.

Alyousef, A. 2013. *Identification and characterisation of lysin enzymes as potential therapeutics for the treatment of Clostridium difficile*. Cardiff: Cardiff University. Available at: https://orca.cardiff.ac.uk/id/eprint/49926 [Accessed: 27 January 2020].

Antonova, N.P. et al. 2019. Broad bactericidal activity of the myoviridae bacteriophage Lysins LysAm24, LysECD7, and LysSi3 against Gram-Negative ESKAPE Pathogens. *Viruses* 11(3), p. 284. doi: 10.3390/v11030284.

Antunes, W. et al. 2018. Structure and assembly of a Clostridioides difficile spore polar appendage. *bioRxiv*, p. 468637. Available at: http://biorxiv.org/content/early/2018/12/03/468637.abstract [Accessed: 23 June 2022].

Aslantas, Y. and Surmeli, N.B. 2019. Effects of N-terminal and C-terminal polyhistidine tag on the stability and function of the thermophilic P450 CYP119. *Bioinorganic Chemistry and Applications* 2019, p. 8080697. doi: 10.1155/2019/8080697.

Barbut, F., Richard, A., Hamadi, K., Chomette, V.R., Burghoffer, B.A. and Petit, J.-C. 2000. *Epidemiology of Recurrences or Reinfections of Clostridium difficile-Associated Diarrhea*.

Barra-Carrasco, J. et al. 2013. The Clostridium difficile exosporium cysteine (CdeC)-rich protein is required for exosporium morphogenesis and coat assembly. *Journal of Bacteriology* 195(17), pp. 3863–3875. doi: 10.1128/JB.00369-13.

Barrett, K.E., Barman, S.M., Boitano, S. and Brooks, H. 2015. Hormonal Control of Calcium & Phosphate Metabolism the Physiology of Bone. In: *Ganong's Review of Medical Physiology*. Available at: https://accessmedicine.mhmedical.com/content.aspx?bookid=1587&sectionid=97164382.

Bartlett, J.G., Chang, T.W., Gurwith, M., Gorbach, S.L. and Onderdonk, A.B. 1978. Antibiotic-Associated Pseudomembranous Colitis Due to Toxin-Producing Clostridia. *New England Journal of Medicine* 298(10), pp. 531–534. doi: 10.1056/NEJM197803092981003.

Bauer, P. et al. 2011. Clostridium difficile infection in Europe: a hospital-based survey. *Lancet* 377, pp. 63–73. Available at: www.thelancet.com.

Begley, M., Gahan, C.G.M. and Hill, C. 2005a. The interaction between bacteria and bile. *FEMS Microbiology Reviews* 29(4), pp. 625–651. doi: 10.1016/j.femsre.2004.09.003.

Begley, M., Gahan, C.G.M. and Hill, C. 2005b. The interaction between bacteria and bile. *FEMS Microbiology Reviews* 29(4), pp. 625–651. doi: 10.1016/j.femsre.2004.09.003.

Brito-Silva, C., Pizarro-Cerda, J., Gil, F. and Paredes-Sabja, D. 2018. Identification of Escherichia coli strains for the heterologous overexpression of soluble Clostridium difficile exosporium proteins. *Journal of Microbiological Methods* 154(April), pp. 46–51. Available at: https://doi.org/10.1016/j.mimet.2018.10.002.

Brunelle, J.L. and Green, R. 2014. Coomassie blue staining. In: *Methods in Enzymology*. Academic Press Inc., pp. 161–167. doi: 10.1016/B978-0-12-420119-4.00013-6.

Budi, N., Safdar, N. and Rose, W.E. 2020. Treatment issues in recurrent Clostridioides difficile infections and the possible role of germinants. *FEMS Microbes* 1(1), p. xtaa001. doi: 10.1093/femsmc/xtaa001.

Calderón-Romero, P. et al. 2018. Clostridium difficile exosporium cysteine-rich proteins are essential for the morphogenesis of the exosporium layer, spore resistance, and affect C. difficile pathogenesis. *PLoS Pathogens* 14(8), p. 1007199. doi: 10.1371/journal.ppat.1007199.

Castro-Córdova, P. et al. 2021. Entry of spores into intestinal epithelial cells contributes to recurrence of Clostridioides difficile infection. *Nature Communications* 12(1), p. 1140. doi: 10.1038/s41467-021-21355-5.

Celebi, O. et al. 2016. The use of germinants to potentiate the sensitivity of Bacillus anthracis spores to peracetic acid. *Frontiers in Microbiology* 7, p. 18. doi: 10.3389/fmicb.2016.00018.

Chen, G., Kang, W., Li, W., Chen, S. and Gao, Y. 2022. Oral delivery of protein and peptide drugs: From non-specific formulation approaches to intestinal cell targeting strategies. *Theranostics* 12(3), pp. 1419–1439. doi: 10.7150/thno.61747.

Chilton, C.H., Pickering, D.S. and Freeman, J. 2018. Microbiologic factors affecting Clostridium difficile recurrence. *Clinical Microbiology and Infection* 24(5), pp. 476–482. doi: 10.1016/j.cmi.2017.11.017.

Cho, J.H., Kwon, J.G., O'Sullivan, D.J., Ryu, S. and Lee, J.H. 2021. Development of an endolysin enzyme and its cell wall–binding domain protein and their applications for biocontrol and rapid detection of Clostridium perfringens in food. *Food Chemistry* 345, p. 128562. doi: 10.1016/j.foodchem.2020.128562.

Chowdhury, R., Sahu, G.K. and Das, J. 1996. Stress response in pathogenic bacteria. *J. Biosci* 21, pp. 149–160. doi: 10.1007/BF02703105.

Clifford McDonald, L. et al. 2005. An Epidemic, Toxin Gene-Variant Strain of Clostridium difficile. *The New England journal of medicine* 353(23), pp. 2433–2441. doi: 10.1056/NEJMoa051590.

Craik, D.J., Fairlie, D.P., Liras, S. and Price, D. 2013. The Future of Peptide-based Drugs. *Chemical Biology and Drug Design* 81(1), pp. 136–147. doi: 10.1111/cbdd.12055.

Crivaro, A.N. et al. 2023. Clostridioides difficile: Characterization of the circulating toxinotypes in an Argentinean public hospital. *Revista Argentina de Microbiologia* 55(1), pp. 73–82. doi: 10.1016/j.ram.2022.05.010.

Curry, S. 2010. Clostridium difficile. *Clinics in Laboratory Medicine* 30(1), pp. 329–342. doi: 10.1016/j.cll.2010.04.001.

Darkoh, C. et al. 2022. Emergence of Clinical Clostridioides difficile Isolates With Decreased Susceptibility to Vancomycin. *Clinical Infectious Diseases* 74(1), pp. 120–126. doi: 10.1093/cid/ciaa912.

Dawson, L.F., Valiente, E., Donahue, E.H., Birchenough, G. and Wren, B.W. 2011. Hypervirulent clostridium difficile pcr-ribotypes exhibit resistance to widely used disinfectants. *PLoS ONE* 6(10), p. e25754. doi: 10.1371/journal.pone.0025754.

Díaz-González, F. et al. 2015. Protein composition of the outermost exosporium-like layer of Clostridium difficile 630 spores. *Journal of Proteomics* 123, pp. 1–13. doi: 10.1016/j.jprot.2015.03.035.

Dimitrov, D.S. 2012. Therapeutic Proteins. *Methods in Molecular Biology* 899, pp. 489–497. doi: 10.1007/978-1-61779-921-1.

Dunne, M. et al. 2014. The CD27L and CTP1L Endolysins Targeting Clostridia Contain a Built-in Trigger and Release Factor. *PLoS Pathogens* 10(7), p. e1004228. doi: 10.1371/journal.ppat.1004228.

Edmondson, D.G. and Roth, S.Y. 2001. Identification of protein interactions by far Western analysis. In: *Current protocols in molecular biology*. pp. 17–2. doi: 10.1002/0471142727.mb2006s55.

Elmi, A. et al. 2018. The bile salt sodium taurocholate induces Campylobacter jejuni outer membrane vesicle production and increases OMV-associated proteolytic activity. *Cellular Microbiology* 20(3), p. e12814. doi: 10.1111/cmi.12814.

Fatima, R. and Aziz, M. 2019. The Hypervirulent Strain of Clostridium Difficile: NAP1/B1/027 - A Brief Overview. *Cureus* 11(1), pp. 1–8. doi: 10.7759/cureus.3977.

Feuerstadt, P., Theriault, N. and Tillotson, G. 2023. The burden of CDI in the United States: a multifactorial challenge. *BMC Infectious Diseases* 23(1), p. 132. doi: 10.1186/s12879-023-08096-0.

Fischetti, V.A. 2005. Bacteriophage lytic enzymes: Novel anti-infectives. *Trends in Microbiology* 13(10), pp. 491–496. doi: 10.1016/j.tim.2005.08.007.

Fortier, L.C. 2017. The Contribution of Bacteriophages to the Biology and Virulence of Pathogenic Clostridia. *Advances in Applied Microbiology* 101, pp. 169–200. doi: 10.1016/bs.aambs.2017.05.002.

Frieden, T. 2013. *Antibiotic Resistance Threats in the United States*. CDC. Available at: https://stacks.cdc.gov/view/cdc/20705 [Accessed: 27 June 2024].

Fu, H. et al. 2020. Codon optimization with deep learning to enhance protein expression. *Scientific Reports* 10(1), p. 17617. doi: 10.1038/s41598-020-74091-z.

Fujinami, Y., Hirai, Y., Sakai, I., Yoshino, M. and Yasuda, J. 2007. Sensitive detection of Bacillus anthracis using a binding protein originating from γ-phage. *Microbiology and Immunology* 51(2), pp. 163–169. doi: 10.1111/j.1348-0421.2007.tb03894.x.

Furlon, J.M., Mitchell, S.J., Bailey-Kellogg, C. and Griswold, K.E. 2021. Bioinformatics-driven discovery of novel Clostridioides difficile lysins and experimental comparison with highly active benchmarks. *Biotechnology and Bioengineering* 118(7), pp. 2482–2492. doi: 10.1002/bit.27759.

Gerding, D.N., Johnson, S., Rupnik, M. and Aktories, K. 2013. Clostridium difficile binary toxin CDT: Mechanism, epidemiology, and potential clinical importance. *Gut Microbes* 5(1), pp. 15–27. doi: 10.4161/gmic.26854.

Gergely Szabo, B. et al. 2016. Use of intravenous tigecycline in patients with severe Clostridium difficile infection: a retrospective observational cohort study. *Clinical Microbiology and Infection* 22(12), pp. 990–995. doi: 10.1016/j.cmi.2016.08.017.

Ghose, C., Eugenis, I., Edwards, A.N., Sun, X., McBride, S.M. and Ho, D.D. 2016. Immunogenicity and protective efficacy of Clostridium difficile spore proteins. *Anaerobe* 37, pp. 85–95. doi: 10.1016/j.anaerobe.2015.12.001.

Gómez-Torres, N., Dunne, M., Garde, S., Meijers, R., Narbad, A., Ávila, M. and Mayer, M.J. 2018. Development of a specific fluorescent phage endolysin for in situ detection of Clostridium species associated with cheese spoilage. *Microbial Biotechnology* 11(2), pp. 332–345. doi: 10.1111/1751-7915.12883.

Gonzales-Luna, A.J. et al. 2020. PCR ribotypes of Clostridioides difficile across Texas from 2011 to 2018 including emergence of ribotype 255. *Emerging Microbes and Infections* 9(1), pp. 341–347. doi: 10.1080/22221751.2020.1721335.

Gonzales-Luna, A.J. et al. 2021. Reduced Susceptibility to Metronidazole Is Associated with Initial Clinical Failure in Clostridioides difficile Infection. *Open Forum Infectious Diseases* 8(8), p. ofab365. doi: 10.1093/ofid/ofab365.

Grisshammer, R. and Tucker, J. 1997. Quantitative Evaluation of Neurotensin Receptor Purification by Immobilized Metal Affinity Chromatography. *PROTEIN EXPRESSION AND PURIFICATION* 11, pp. 53–60.

Guo, S. et al. 2015. The recombinant Lactococcus lactis oral vaccine induces protection against C. difficile spore challenge in a mouse model. *Vaccine* 33(13), pp. 1586–1595. doi: 10.1016/j.vaccine.2015.02.006.

Gustafsson, C., Govindarajan, S. and Minshull, J. 2004. Codon bias and heterologous protein expression. *Trends in Biotechnology* 22(7), pp. 346–353. doi: 10.1016/j.tibtech.2004.04.006.

Herbert, R. et al. 2019. Two-year analysis of Clostridium difficile ribotypes associated with increased severity. *Journal of Hospital Infection* 103(4), pp. 388–394. doi: 10.1016/j.jhin.2019.06.003.

Ho, M.K.Y., Zhang, P., Chen, X., Xia, J. and Leung, S.S.Y. 2022. Bacteriophage endolysins against gram-positive bacteria, an overview on the clinical development and recent advances on the delivery and formulation strategies. *Critical Reviews in Microbiology* 48(3), pp. 303–326. doi: 10.1080/1040841X.2021.1962803.

Horgan, M., O'Sullivan, O., Coffey, A., Fitzgerald, G.F., van Sinderen, D., McAuliffe, O. and Ross, R.P. 2010. Genome analysis of the Clostridium difficile phage φCD6356, a temperate phage of the Siphoviridae family. *Gene* 462(1–2), pp. 34–43. doi: 10.1016/j.gene.2010.04.010.

Icho, S. and Melnyk, R.A. 2018. Dismantling a Toxin to Disarm a Superbug. *Cell Press* 40(3), pp. 155–156. doi: 10.1016/j.tips.2019.01.007.

Ishikawa, S., Kawahara, S. and Sekiguchi, J. 1999. Cloning and expression of two autolysin genes, cwlU and cwlV, which are tandemly arranged on the chromosome of Bacillus polymyxa var. colistinus. *Molecular and General Genetics* 262(4–5), pp. 738–748. doi: 10.1007/s004380051136.

Jiang, S., Wan, Q., Krajcikova, D., Tang, J., Tzokov, S.B., Barak, I. and Bullough, P.A. 2015. Diverse supramolecular structures formed by self-assembling proteins of the Bacillus subtilis spore coat. *Molecular Microbiology* 97(2), pp. 347–359. doi: 10.1111/mmi.13030.

Jiang, Z.D. et al. 2018. Safety and preliminary efficacy of orally administered lyophilized fecal microbiota product compared with frozen product given by enema for recurrent Clostridium difficile infection: A randomized clinical trial. *PLoS ONE* 13(11), p. e0205064. doi: 10.1371/journal.pone.0205064.

Johnson, S., Lavergne, V., Skinner, A.M., Gonzales-Luna, A.J., Garey, K.W., Kelly, C.P. and Wilcox, M.H. 2021. Clinical Practice Guideline by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA): 2021 Focused Update Guidelines on Management of Clostridioides difficile Infection in Adults. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 73(5), pp. 1029–1044. doi: 10.1093/cid/ciab549.

Kailas, L. et al. 2011. Surface architecture of endospores of the Bacillus cereus/anthracis/ thuringiensis family at the subnanometer scale. *Proceedings of the National Academy of Sciences of the United States of America* 108(38), pp. 16014–16019. doi: 10.1073/pnas.1109419108.

Kochan, T.J. et al. 2017. Intestinal calcium and bile salts facilitate germination of Clostridium difficile spores. *PLoS Pathogens* 13(7), p. e1006443. doi: 10.1371/journal.ppat.1006443.

Kochan, T.J. et al. 2018. Germinant Synergy Facilitates Clostridium difficile Spore Germination under Physiological Conditions. *mSphere* 3(5), pp. e00335-18. doi: 10.1128/msphere.00335-18.
Kociolek, L.K., Gerding, D.N., Hecht, D.W. and Ozer, E.A. 2018. Comparative genomics analysis of Clostridium difficile epidemic strain DH/NAP11/106. *Microbes and Infection* 20(4), pp. 245–253. Available at: https://doi.org/10.1016/j.micinf.2018.01.004.
Kong, M., Na, H., Ha, N.-C. and Ryu, S. 2019. LysPBC2, a Novel Endolysin Harboring a Bacillus cereus Spore Binding Domain. *ASM Journals* 85(5), pp. e02462-18. doi: 10.1128/AEM.02462-18.
Kuijper, E.J. et al. 2008. Update of Clostridium difficile Infection due to PCR Ribotype 027 in Europe. *EUROROUNDUPS* 13(31), p. 18942.
Lawson, P.A., Citron, D.M., Tyrrell, K.L. and Finegold, S.M. 2016. Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prévot 1938. *Anaerobe* 40, pp. 95–99. doi: 10.1016/j.anaerobe.2016.06.008.
Leslie, J.L. et al. 2021. Protection from lethal clostridioides difficile infection via intraspecies competition for cogerminant. *mBio* 12(2), pp. e00522-21. doi: 10.1128/mBio.00522-21.
Lessa, F.C. et al. 2015. Burden of Clostridium difficile Infection in the United States . *New England Journal of Medicine* 372(9), pp. 825–834. doi: 10.1056/nejmoa1408913.
Leverrier, P., Dimova, D., Pichereau, V., Auffray, Y., Boyaval, P. and Jan, G. 2003. Susceptibility and adaptive response to bile salts in Propionibacterium freudenreichii: Physiological and proteomic anlysis. *Applied and Environmental Microbiology* 69(7), pp. 3809–3818. doi: 10.1128/AEM.69.7.3809-3818.2003.
Liu, H., Hu, Z., Li, M., Yang, Y., Lu, S. and Rao, X. 2023. Therapeutic potential of bacteriophage endolysins for infections caused by Gram-positive bacteria. *Journal of Biomedical Science* 30(1), p. 29. doi: 10.1186/s12929-023-00919-1.
Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P. and Berkmen, M. 2012. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories* 11(1), p. 1. doi: 10.1186/1475-2859-11-56.
Low, L.Y., Yang, C., Perego, M., Osterman, A. and Liddington, R.C. 2005. Structure and lytic activity of a Bacillus anthracis prophage endolysin. *Journal of Biological Chemistry* 280(42), pp. 35433–35439. doi: 10.1074/jbc.M502723200.

Magnusson, C., Mernelius, S., Bengnér, M., Norén, T., Serrander, L., Forshell, S. and Matussek, A. 2022. Characterization of a Clostridioides difficile outbreak caused by PCR ribotype 046, associated with increased mortality. *Emerging Microbes and Infections* 11(1), pp. 850–859. doi: 10.1080/22221751.2022.2049981.

Malhotra, A. 2009. *Chapter 16 Tagging for Protein Expression*. Academic Press Inc. doi: 10.1016/S0076-6879(09)63016-0.

Malyshev, D. 2018. *Disruption of bacterial spores using microwaves and nanoparticles*.

Mann, S. and Chen, Y.P.P. 2010. Bacterial genomic G + C composition-eliciting environmental adaptation. *Genomics* 95(1), pp. 7–15. doi: 10.1016/j.ygeno.2009.09.002.

Marković, V., Kostić, M., Iličković, I. and Janković, S.M. 2014. Cost-Effectiveness Comparison of Fidaxomicin and Vancomycin for Treatment of Clostridium difficile Infection: A Markov Model Based on Data from a South West Balkan Country in Socioeconomic Transition. *Value in Health Regional Issues* 4(1), pp. 87–94. doi: 10.1016/j.vhri.2014.08.001.

Marsh, J.W., Arora, R., Schlackman, J.L., Shutt, K.A., Curry, S.R. and Harrison, L.H. 2012. Association of relapse of Clostridium difficile disease with BI/NAP1/027. *Journal of Clinical Microbiology* 50(12), pp. 4078–4082. doi: 10.1128/JCM.02291-12.

Mauro, V.P. 2018. Codon Optimization in the Production of Recombinant Biotherapeutics: Potential Risks and Considerations. *BioDrugs* 32(1), pp. 69–81. doi: 10.1007/s40259-018-0261-x.

Mayer, M.J., Garefalaki, V., Spoerl, R., Narbad, A. and Meijers, R. 2011. Structure-based modification of a Clostridium difficile-targeting endolysin affects activity and host range. *Journal of Bacteriology* 193(19), pp. 5477–5486. doi: 10.1128/JB.00439-11.

Mayer, M.J., Narbad, A. and Gasson, M.J. 2008. Molecular characterization of a Clostridium difficile bacteriophage and its cloned biologically active endolysin. *Journal of Bacteriology* 190(20), pp. 6734–6740. doi: 10.1128/JB.00686-08.

Mckenney, P.T., Driks, A. and Eichenberger, P. 2013. The Bacillus subtilis endospore: Assembly and functions of the multilayered coat. *Nature Reviews Microbiology* 11(1), pp. 33–44. doi: 10.1038/nrmicro2921.

Meessen-Pinard, M., Sekulovic, O. and Fortier, L.C. 2012. Evidence of in vivo prophage induction during clostridium difficile infection. *Applied and Environmental Microbiology* 78(21), pp. 7662–7670. doi: 10.1128/AEM.02275-12.

Mehta, K.K., Paskaleva, E.E., Azizi-Ghannad, S., Ley, D.J., Page, M.A., Dordick, J.S. and Kanea, R.S. 2013. Characterization of AmiBA2446, a novel bacteriolytic enzyme active against bacillus species. *Applied and Environmental Microbiology* 79(19), pp. 5899–5906. doi: 10.1128/AEM.02235-13.

Menzella, H.G. 2011. Comparison of two codon optimization strategies to enhance recombinant protein production in Escherichia coli. *Microbial Cell Factories* 10, pp. 11–15. doi: 10.1186/1475-2859-10-15.

Miller, B.A., Chen, L.F., Sexton, D.J. and Anderson, D.J. 2011. Comparison of the Burdens of Hospital-Onset, Healthcare Facility-Associated Clostridium difficile Infection and of Healthcare-Associated Infection due to Methicillin-Resistant Staphylococcus aureus in Community Hospitals. *Infection Control & Hospital Epidemiology* 32(4), pp. 387–390. doi: 10.1086/659156.

Mondal, S.I., Akter, A., Draper, L.A., Ross, R.P. and Hill, C. 2021. Characterization of an endolysin targeting clostridioides difficile that affects spore outgrowth. *International Journal of Molecular Sciences* 22(11), p. 5690. doi: 10.3390/ijms22115690.

Mora-Uribe, P. et al. 2016. Characterization of the adherence of Clostridium difficile spores: The integrity of the outermost layer affects adherence properties of spores of the epidemic strain R20291 to components of the intestinal mucosa. *Frontiers in Cellular and Infection Microbiology* 6, p. 99. doi: 10.3389/fcimb.2016.00099.

Murray, E., Draper, L.A., Ross, R.P. and Hill, C. 2021. The advantages and challenges of using endolysins in a clinical setting. *Viruses* 13(4), p. 680. doi: 10.3390/v13040680.

Nazzaro, F., Fratianni, F., De Martino, L., Coppola, R. and De Feo, V. 2013. Effect of essential oils on pathogenic bacteria. *Pharmaceuticals* 6(12), pp. 1451–1474. doi: 10.3390/ph6121451.

Nelson, D.C., Schmelcher, M., Rodriguez-Rubio, L., Klumpp, J., Pritchard, D.G., Dong, S. and Donovan, D.M. 2012. Endolysins as Antimicrobials. In: *Advances in Virus Research*. Academic Press Inc., pp. 299–365. doi: 10.1016/B978-0-12-394438-2.00007-4.

Nerandzic, M.M. and Donskey, C.J. 2016. A quaternary ammonium disinfectant containing germinants reduces clostridium difficile spores on surfaces by inducing susceptibility to environmental stressors. *Open Forum Infectious Diseases* 3(4), p. ofw196. doi: 10.1093/ofid/ofw196.

Nerandzic, M.M. and Donskey, C.J. 2017. Sensitizing clostridium difficile spores with germinants on skin and environmental surfaces represents a new strategy for reducing spores via ambient mechanisms. *Pathogens and Immunity* 2(3), pp. 404–421. doi: 10.20411/pai.v2i3.221.
NICE. 2017. Preventing recurrence of Clostridium difficile infection: bezlotoxumab. p. 38. Available at: https://www.nice.org.uk/advice/es13/resources/preventing-recurrence-of-clostridium-difficile-infection-bezlotoxumab-pdf-1158113662405.
NICE. 2021. *Clostridioides difficile infection: antimicrobial prescribing*. Available at: https://www.nice.org.uk/guidance/ng199/resources/clostridioides-difficile-infection-antimicrobial-prescribing-pdf-66142090546117 [Accessed: 21 February 2023].
Oberkampf, M. et al. 2022. c-di-AMP signaling is required for bile salt resistance, osmotolerance, and long-term host colonization by Clostridioides difficile. *Science Signaling* 15(750), p. 8171. doi: 10.1126/scisignal.abn8171.
Office for National Statistics. 2014. *Deaths Involving Clostridium difficile, England and Wales 2012*. Available at: http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsinvolvingclostridiumdifficileenglandandwales/2013-08-22.
Office for National Statistics. 2017. *Deaths involving Clostridium difficile*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsinvolvingclostridiumdifficilereferencetables [Accessed: 27 February 2024].
Orrell, K.E., Zhang, Z., Sugiman-Marangos, S.N. and Melnyk, R.A. 2017. Clostridium difficile toxins A and B: Receptors, pores, and translocation into cells. *Critical Reviews in Biochemistry and Molecular Biology* 52(4), pp. 461–473. doi: 10.1080/10409238.2017.1325831.
Pandey, R., Ter Beek, A., Vischer, N.O.E., Smelt, J.P.P.M., Brul, S. and Manders, E.M.M. 2013. Live Cell Imaging of Germination and Outgrowth of Individual Bacillus subtilis Spores; the Effect of Heat Stress Quantitatively Analyzed with SporeTracker. *PLoS ONE* 8(3), p. e58972. doi: 10.1371/journal.pone.0058972.
Paredes-Sabja, D., Cid-Rojas, F. and Pizarro-Guajardo, M. 2022. Assembly of the exosporium layer in Clostridioides difficile spores. *Current Opinion in Microbiology* 67, p. 102137. doi: 10.1016/j.mib.2022.01.008.

Paredes-Sabja, D., Shen, A. and Sorg, J.A. 2014. Clostridium difficile spore biology: Sporulation, germination, and spore structural proteins. *Trends in Microbiology* 22(7), pp. 406–416. Available at: http://dx.doi.org/10.1016/j.tim.2014.04.003.

Peng, Z. et al. 2019. A novel bacteriophage lysin-human defensin fusion protein is effective in treatment of clostridioides difficileinfection in mice. *Frontiers in Microbiology* 9, p. 3234. doi: 10.3389/fmicb.2018.03234.

Pham, P. V. 2018. Medical biotechnology: Techniques and applications. In: *Omics Technologies and Bio-engineering: Towards Improving Quality of Life*. Elsevier Inc., pp. 449–469. doi: 10.1016/B978-0-12-804659-3.00019-1.

Phothichaisri, W., Chankhamhaengdecha, S., Janvilisri, T., Nuadthaisong, J., Phetruen, T., Fagan, R.P. and Chanarat, S. 2022. 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. *Microbiology Spectrum* 10(2), pp. e02361-21. doi: 10.1128/spectrum.02361-21.

Van Pijkeren, J.-P. and Barrangou, R. 2017. Genome Editing of Food-Grade Lactobacilli To Develop Therapeutic Probiotics. *Microbiology spectrum* 5(5), pp. 10–1128. doi: 10.1128/microbiolspec.BAD-0013-2016.

Pizarro-Guajardo, M., Ravanal, M.C., Paez, M.D., Callegari, E. and Paredes-Sabja, D. 2018. Identification of Clostridium difficile Immunoreactive Spore Proteins of the Epidemic Strain R20291. *Proteomics - Clinical Applications* 12(5), pp. 10–14. doi: 10.1002/prca.201700182.

van Prehn, J. et al. 2021. European Society of Clinical Microbiology and Infectious Diseases: 2021 update on the treatment guidance document for Clostridioides difficile infection in adults. *Clinical Microbiology and Infection* 27, pp. S1–S21. doi: 10.1016/j.cmi.2021.09.038.

Public Health England. 2016a. Clostridium difficile Ribotyping Network (CDRN) for England and Northern Ireland. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/491253/CDRN_2013-15_Report.pdf [Accessed: 10 January 2023].

Public Health England. 2016b. *UK Standards for Microbiology Investigations, Identification of Clostridium species*. Available at: http://www.apsi.it/public/ufiles/smi/id8_3_2_rev_en_140310.pdf [Accessed: 27 June 2024].

Public Health Wales. 2013. *Clostridium difficile PCR Ribotype Surveillance in Wales 2012*.

Puetz, J. and Wurm, F.M. 2019. Recombinant Proteins for Industrial versus Pharmaceutical Purposes: A Review of Process and Pricing. *Processes* 7(8), p. 476. doi: 10.3390/pr7080476.
Qiagen. 2011. *QIAexpress ® Ni-NTA Fast Start Handbook*. Available at: https://www.qiagen.com/fi/resources/resourcedetail?id=e58890b4-a93f-4a4d-961d-8bf82e2d74d6&lang=en [Accessed: 20 March 2023].
Qiagen. 2023. *Should I use Ni-NTA Agarose in column or batch format for purification of 6xHis-tagged proteins?* Available at: https://www.qiagen.com/us/resources/faq?id=6edf1bf1-b412-48ce-82fb-f87c6ae0d972&lang=en [Accessed: 4 July 2023].
Romero-Calle, D., Benevides, R.G., Góes-Neto, A. and Billington, C. 2019. Bacteriophages as alternatives to antibiotics in clinical care. *Antibiotics* 8(3), p. 138. doi: 10.3390/antibiotics8030138.
Romero-Rodríguez, A., Troncoso-Cotal, S., Guerrero-Araya, E. and Paredes-Sabjaa, D. 2020. The Clostridioides difficile Cysteine-Rich Exosporium Morphogenetic Protein, CdeC, Exhibits Self-Assembly Properties That Lead to Organized Inclusion Bodies in Escherichia coli. *American Society for Microbiology* 5(6), pp. 1–19. doi: 10.1128/mSphere.01065-20.
Sainathrao, S., Mohan, K. and Atreya, C. 2009. Gamma-phage lysin PlyG sequence-based synthetic peptides coupled with Qdot-nanocrystals are useful for developing detection methods for Bacillus anthracis by using its surrogates, B. anthracis-Sterne and B. cereus-4342. *BMC Biotechnology* 9, p. 67. doi: 10.1186/1472-6750-9-67.
Salazar, C.L. et al. 2017. Molecular, microbiological and clinical characterization of Clostridium difficile isolates from tertiary care hospitals in Colombia. *PLoS ONE* 12(9), p. e0184689. doi: 10.1371/journal.pone.0184689.
Santos, S.B., Oliveira, A., Melo, L.D.R. and Azeredo, J. 2019. Identification of the first endolysin Cell Binding Domain (CBD) targeting Paenibacillus larvae. *Scientific Reports* 9(1), p. 2568. doi: 10.1038/s41598-019-39097-2.
Schmelcher, M., Donovan, D.M. and Loessner, M.J. 2012. Bacteriophage endolysins as novel antimicrobials. *Future Microbiology* 7(10), pp. 1147–1171. doi: 10.2217/fmb.12.97.
Schmidt, J.A. et al. 2016. Plasma concentrations and intakes of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans: A cross-sectional analysis in the EPIC-Oxford cohort. *European Journal of Clinical Nutrition* 70(3), pp. 306–312. doi: 10.1038/ejcn.2015.144.

Schuch, R., Pelzek, A.J., Nelson, D.C. and Fischetti, V.A. 2019. The PlyB Endolysin of Bacteriophage vB_BanS_Bcp1 Exhibits Broad-Spectrum Bactericidal Activity against Bacillus cereus Sensu Lato Isolates. *Applied and environmental microbiology* 85(9), pp. 00003–00019. doi: 10.1128/AEM.00003-19.

Sebaihia, M. et al. 2006. The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature Genetics* 38(7), pp. 779–786. doi: 10.1038/ng1830.

Setlow, P., Wang, S. and Li, Y.-Q. 2017. Germination of Spores of the Orders Bacillales and Clostridiales. *Annual Review of Microbiology* 71(1), pp. 459–477. doi: 10.1146/annurev-micro-090816-093558.

Shen, A., Edwards, A.N., Sarker, M.R. and Paredes-Sabja, D. 2019. Sporulation and Germination in Clostridial Pathogens. *Physiology & behavior* 176(5), pp. 139–148. doi: 10.1016/j.physbeh.2017.03.040.

Shen, Z. et al. 2016. A Catalytic DNA Activated by a Specific Strain of Bacterial Pathogen. *Angewandte Chemie - International Edition* 55(7), pp. 2431–2434. doi: 10.1002/anie.201510125.

Shrestha, R. and Sorg, J.A. 2018. Hierarchical recognition of amino acid co-germinants during Clostridioides difficile spore germination. *Anaerobe* 49, pp. 41–47. doi: 10.1016/j.anaerobe.2017.12.001.

Sievers, S. et al. 2019. Differential view on the bile acid stress response of clostridioides difficile. *Frontiers in Microbiology* 10, p. 258. doi: 10.3389/fmicb.2019.00258.

Sorg, J.A. and Sonenshein, A.L. 2008. Bile salts and glycine as cogerminants for Clostridium difficile spores. *Journal of Bacteriology* 190(7), pp. 2505–2512. doi: 10.1128/JB.01765-07.

Sorg, J.A. and Sonenshein, A.L. 2010. Inhibiting the initiation of Clostridium difficile spore germination using analogs of chenodeoxycholic acid, a bile acid. *Journal of Bacteriology* 192(19), pp. 4983–4990. doi: 10.1128/JB.00610-10.

Stare, B.G., Delmée, M. and Rupnik, M. 2007. Variant forms of the binary toxin CDT locus and tcdC gene in Clostridium difficile strains. *Journal of Medical Microbiology* 56(3), pp. 329–335. doi: 10.1099/jmm.0.46931-0.

Terry, C., Jiang, S., Radford, D.S., Wan, Q., Tzokov, S., Moir, A. and Bullough, P.A. 2017. Molecular tiling on the surface of a bacterial spore – the exosporium of the Bacillus anthracis/cereus/thuringiensis group. *Molecular Microbiology* 104(4), pp. 539–552. doi: 10.1111/mmi.13650.

Thabit, A.K., Alsolami, M.H., Baghlaf, N.A., Alsharekh, R.M., Almazmumi, H.A., Alselami, A.S. and Alsubhi, F.A. 2019. Comparison of three current Clostridioides difficile infection guidelines: IDSA/SHEA, ESCMID, and ACG guidelines. *Infection* 47(6), pp. 899–909. doi: 10.1007/s15010-019-01348-9.

Tina Joshi, L. 2012. *Pathogenicity & A Bedside Real-Time Detection Assay For Clostridium difficile In The Faeces of Hospitalized Patients*. Available at: https://orca.cardiff.ac.uk/id/eprint/37079 [Accessed: 27 June 2024].

Tonna, I. and Welsby, P.D. 2005. Pathogenesis and treatment of Clostridium difficile infection. *Postgraduate Medical Journal* 81(956), pp. 367–369. doi: 10.1136/pgmj.2004.028480.

Vasina, D. v. et al. 2021. Discovering the Potentials of Four Phage Endolysins to Combat Gram-Negative Infections. *Frontiers in Microbiology* 12. doi: 10.3389/fmicb.2021.748718.

Vermassen, A., Leroy, S., Talon, R., Provot, C., Popowska, M. and Desvaux, M. 2019. Cell wall hydrolases in bacteria: Insight on the diversity of cell wall amidases, glycosidases and peptidases toward peptidoglycan. *Frontiers in Microbiology* 10, p. 331. doi: 10.3389/fmicb.2019.00331.

Voth, D.E. and Ballard, J.D. 2005. Clostridium difficile toxins: Mechanism of action and role in disease. *Clinical Microbiology Reviews* 18(2), pp. 247–263. doi: 10.1128/CMR.18.2.247-263.2005.

Wang, G., Yi, X., Li, Y.Q. and Setlow, P. 2011. Germination of individual Bacillus subtilis spores with alterations in the GerD and SpoVA proteins, which are important in spore germination. *Journal of Bacteriology* 193(9), pp. 2301–2311. doi: 10.1128/JB.00122-11.

Wang, Q., Euler, C.W., Delaune, A. and Fischetti, V.A. 2015. Using a novel lysin to help control Clostridium difficile infections. *Antimicrobial Agents and Chemotherapy* 59(12), pp. 7447–7457. doi: 10.1128/AAC.01357-15.

Waugh, D.S. 2005. Making the most of affinity tags. *Trends in Biotechnology* 23(6), pp. 316–320. doi: 10.1016/j.tibtech.2005.03.012.

Woestenenk, E.A., Hammarström, M., Van Den Berg, S., Härd, T. and Berglund, H. 2004. His tag effect on solubility of human proteins produced in Escherichia coli: A comparison between four expression vectors. *Journal of Structural and Functional Genomics* 5(3), pp. 217–229. doi: 10.1023/B:jsfg.0000031965.37625.0e.

Yang, H. et al. 2012. Existence of separate domains in lysin PlyG for recognizing Bacillus anthracis spores and vegetative cells. *Antimicrobial Agents and Chemotherapy* 56(10), pp. 5031–5039. doi: 10.1128/AAC.00891-12.

Żebrowska, J., Żołnierkiewicz, O., Ponikowska, M., Puchalski, M., Krawczun, N., Makowska, J. and Skowron, P. 2022. Cloning and Characterization of a Thermostable Endolysin of Bacteriophage TP-84 as a Potential Disinfectant and Biofilm-Removing Biological Agent. *International Journal of Molecular Sciences* 23(14), p. 7612. doi: 10.3390/ijms23147612.

Zhang, S., Chang, Y., Zhang, Q., Yuan, Y., Qi, Q. and Lu, X. 2022. Characterization of Salmonella endolysin XFII produced by recombinant Escherichia coli and its application combined with chitosan in lysing Gram-negative bacteria. *Microbial Cell Factories* 21(1), p. 171. doi: 10.1186/s12934-022-01894-2.

Zhu, D., Sorg, J.A. and Sun, X. 2018. Clostridioides difficile biology: Sporulation, germination, and corresponding therapies for C. difficile infection. *Frontiers in Cellular and Infection Microbiology* 8, pp. 1–10. doi: 10.3389/fcimb.2018.00029.

Appendices

Appendix 1 Bio-informatic analysis

Bioinformatic analysis was utilized to confirm the specificity of each protein for *C. difficile* and the degree of structure conservation within each target protein. The two targets were CdeC and CdeM, which were identified from the literature.

1.1 CdeC

Homologs to the CdeC protein sequence of CD630 were found primarily in isolates of *C. difficile* but were also observed in member of other anaerobic bacterial species. To demonstrate the degree of similarity, a phylogenetic tree was constructed using the neighbor joining model in the NCBI database (Figure A1). The percent identity of *C. difficile* isolates ranged from 99.63% to 100% while the nearest non-*C. difficile* homolog which was found in *Clostridioides mangenotii* showing 59.95% identity.

Figure A6.1 A phylogenetic tree showing the relationship between CdeC homologs listed in the NCBI database.

The phylogenetic tree was constructed using the neighbor joining model in the NCBI database. The search sequence from *C. difficile* CD630, is highlighted in yellow.

To better understand the variations in CdeC amino acid sequences which were observed, we made a detailed comparison of the individual amino acid sequences of each homolog with that of the reference sequence derived from *C. difficile* CD630. A multiple sequence alignment was performed using Clustal Omega software (Figure A2).

Figure A6.2 Alignment of the amino acid sequences of homolog to the *C. difficile* CD630 CdeC protein sequence.

NCBI reference codes on the left refers to different *C. difficile* strains that have homology to the CdeC reference sequence of CD630. Symbol (*) indicates perfect alignment regions, (:) indicates difference in amino acids but within a group of strong similarity amino acids. Sequences were collected from BLAST-NCBI and uploaded into Clustal Omega software to show the multiple sequence alignment. The Red box highlighted area of possible deletion.

Multiple sequence alignment of the first 16 sequences demonstrated that all sequences belonging to *C. difficile* strains showed high similarity to the query protein sequence. It revealed a highly conserved sequence with one deletion site between positions 60 and 70 on the N terminus (highlighted in red box in Figure A2). The deleted region contains a repeated amino acid sequence comprising CNP, which codes for cysteine, asparagine and proline, respectively.

Studying of each sequence using Identical Protein Groups (IPG), which arranges strains based on shared uploaded sequence to the databases, revealed that each accession number is shared among a group of *C. difficile* strains. IPG-CdeC groups have different number of strains which range from 1 to 1116 strains. Reference strain, CD630, is a member of the biggest group (WP_003437985) in term of number of strains indicating that the chosen sequence of CdeC (CD630) has a certain length that is shared among most *C. difficile* strains.

1.1.1 Structural analysis of CdeC

Due to a lack of a crystal structure with which to predict the impact of these mutations on the structure of the homologs, we analyzed individual gene sequences, in FASTA format, using the I-TASSER structural prediction model. We aligned structures of the reference strain, CD630, with the Ish01 strain (NCBI reference sequence NZ_CAAJWN000000000.1), a *Clostridioides difficile* strain as a representative of an undeleted CdeC protein sequence, to investigate the role of the deletion on protein structure (Figure A3).

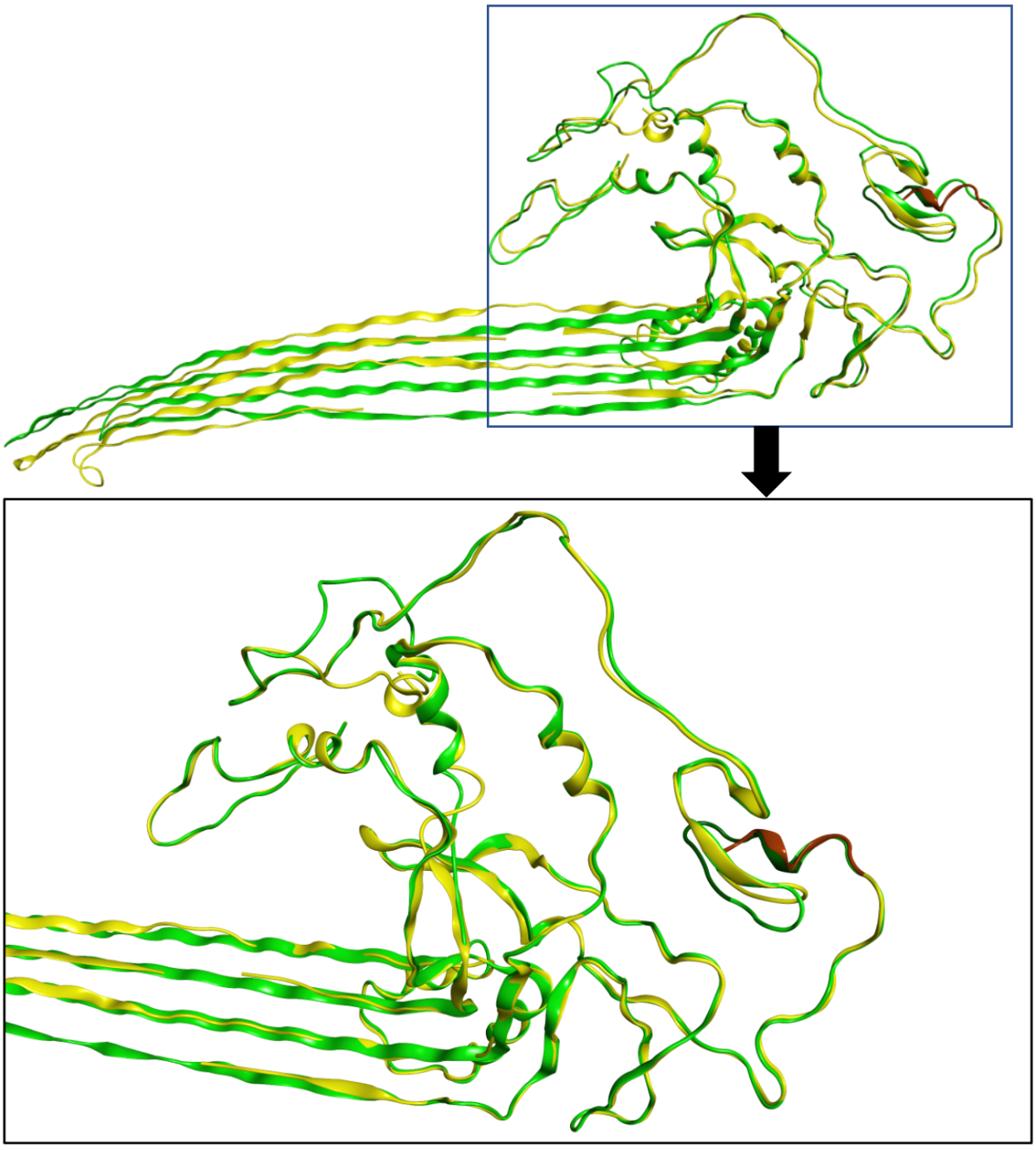


Figure A6.3 CdeC prediction models showing CdeC sequence with deletion aligned and compared to complete CdeC sequence.

In yellow is the predicted structure of the CdeC protein (405 amino acids) from CD630, and in green is the predicted structure of the CdeC homolog (408 amino acids) from Ish01. The enlarged structure is to highlight the deleted area of Ish01 (dark green) when compared to CD630 (brown). Models were predicted by I-TASSER software, and the modification of the colours and the alignment of the structures were demonstrated by MOE (Molecular Operating Environment) program.

The comparison revealed that the amino acid deletion in the CD630 CdeC protein sequence occurs in a region predicted to form an unstructured loop and thus is unlikely to contribute to overall protein structural stability and thus is unlikely to under pressure to be conserved (Gilles-Gonzalez et al., 1991; Tastan et al., 2009).

1.2 CdeM

In contrast to CdeC, homologs to the CD630 protein sequence of CdeM were mainly found in *C. difficile* (Figure 7.6). Other than *C. difficile*, two sequences belong to a bacterium (*Clostridioides mangenotii*) and a mammal (*Ailuropoda melanoleuca*) were identified but with very low percentage identity compared to the query sequence of 26 % and 32.93 %, respectively. The APC membrane recruitment protein 1 of *Ailuropoda melanoleuca* is produced inside Panda bear cells and thus is unlikely to be a problem when developing a therapy for *C. difficile.*

Similar to *C. difficile, Clostridioides mangenotii* is a strictly anaerobic Gram-positive bacteria with the ability to form spores (McLaughlin et al., 2014; Lawson et al., 2016). Both species belong to *Peptostreptococcaceae* family (Lawson et al., 2016). Additionally, it has been isolated from soil and animal feces, and could potentially cross react with an approach which targets these proteins (McLaughlin et al., 2014).

To demonstrate the degree of similarity, a phylogenetic tree was constructed using the neighbor joining model in the NCBI database (Figure A4)

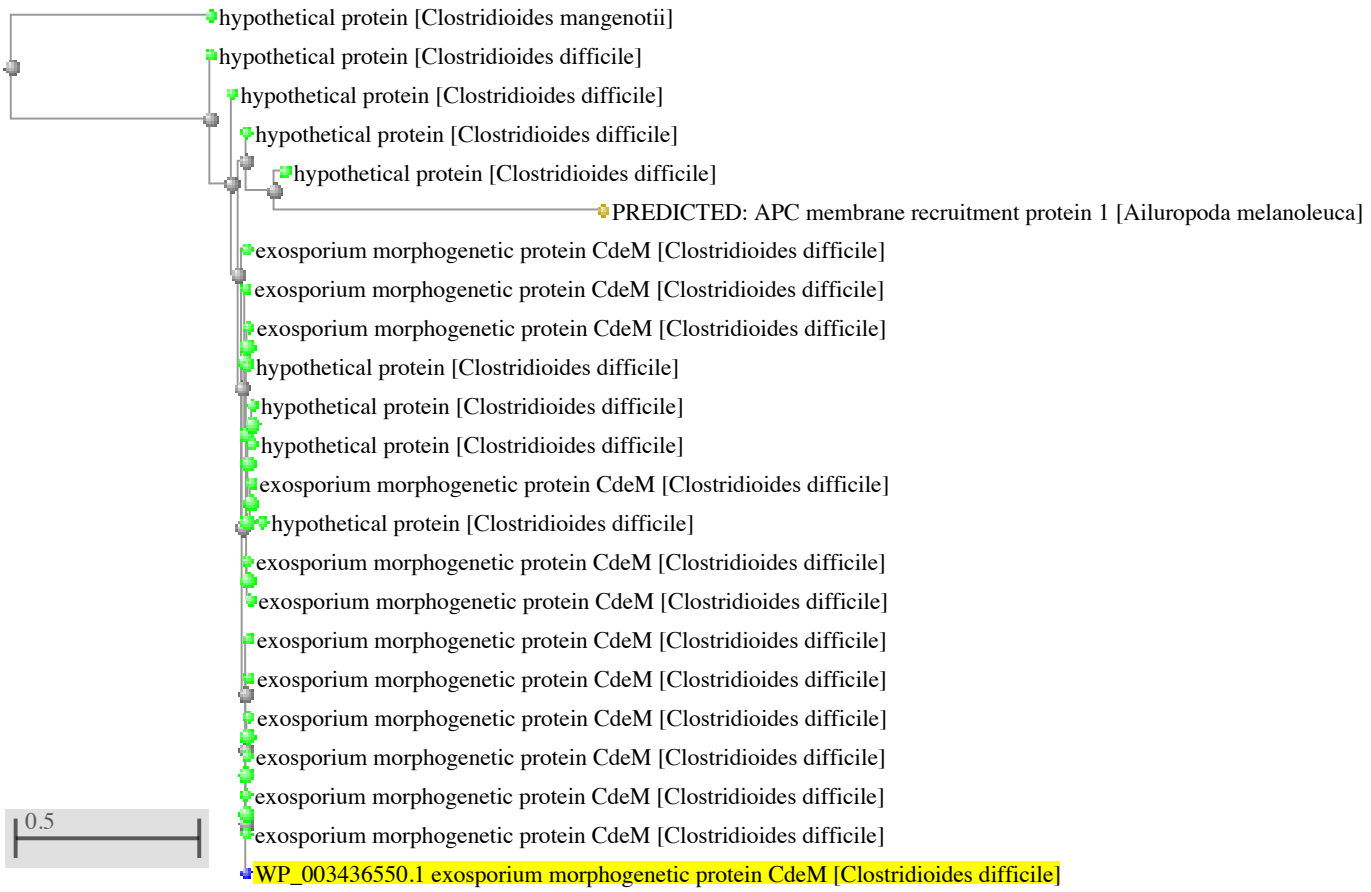


Figure A6.4 A phylogenetic tree showing the relationship between CdeM homologs

The phylogenetic tree was constructed using the neighbor joining model in the NCBI database. The query sequence, CdeM of *C. difficile* CD630, is highlighted in yellow. *C. mangenotii* is highlighted in blue.

Alignment of the *Clostridioides mangenotii* sequence against the query sequence revealed that the protein sequence of *C. mangenotii* is 29 amino acids shorter than the CD630 sequence and the similarity between the two sequences was only 26% (Figure A5). Additionally, the sequence of *Clostridioides mangenotii* is hypothetical protein which may or may not participate in spore structure.

Figure A6.5 Alignment of *C. mangenotii* hypothetical protein against *C. difficile* CdeM protein sequence.

The protein sequence alignment revealed that both sequences are cysteine rich proteins, which may involve in the structure folding due to the formation of disulfide bonds. Moreover, *C. mangenotii* hypothetical protein has a shortage in its sequence in both N and C termini. NCBI reference codes on the left refers to CdeM sequence of the reference strain *C. difficile* CD630 (WP_003436550.1) and *C. mangenotii* (WP_024620077.1). Symbol (*) indicates regions of a single conserved residue, (-) indicates regions of sequence shortage, (:) indicates difference in amino acids but within a group of strong similar properties, (.) indicates difference in amino acids but within a group of weak similar properties. The multiple sequence alignment was performed using Clustal Omega.

Following our analysis of the suitability of CdeM as a marker for *C. difficile,* the degree of amino acid variation amongst CdeM homologs of different *C. difficile* isolates compared to that of *C. difficile* CD630 was investigated. A multiple sequence alignment was performed using Clustal Omega software (Figure A6)

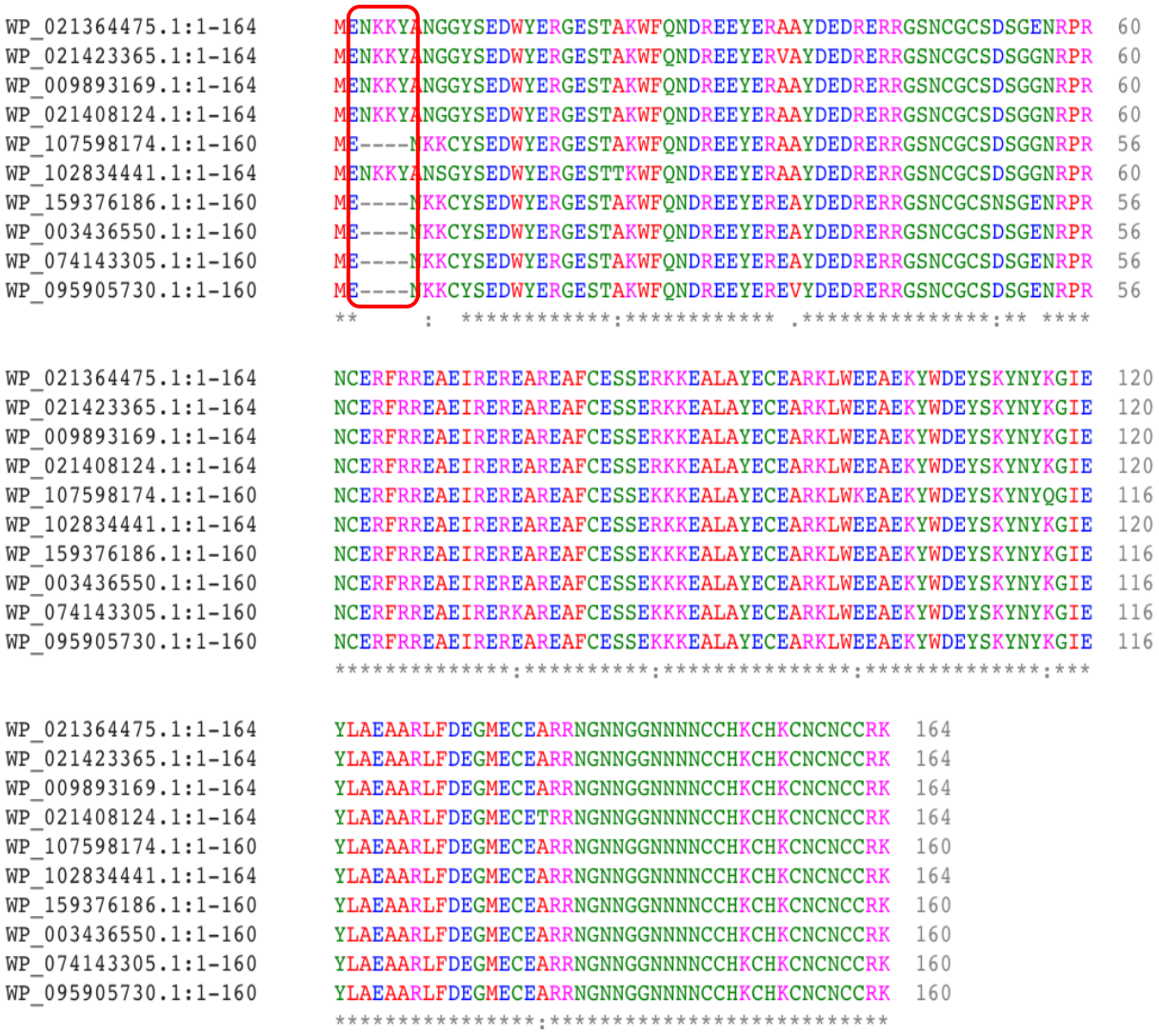


Figure A6.6 Alignment of multiple *C. difficile* CdeM protein sequences

Some *C. difficile* strains demonstrate a deletion of four amino acids at the N-terminal site. Symbol (*) indicates regions of a single conserved residue, (-) indicates regions of sequence loss, (:) indicates difference in amino acids but within a group of strong similar properties, (.) indicates difference in amino acids but within a group of weak similar properties. Sequences were collected from BLAST-NCBI and uploaded into Clustal Omega software to show the multiple sequence alignment. The Red box highlighted area of possible deletion.

Multiple sequence alignment of the amino acid sequences from different *C. difficile* isolates showed a high degree of similarity to the CdeM sequence of the reference strain CD630. Alignment revealed a highly conserved sequence with the exception of amino acids deletion occurring at the N terminus between positions 2 and 7.

Studying of each sequence using IPG revealed that each accession number represents CdeM protein sequence in different number of *C. difficile* strains ranging from 1 to 945 strains. Reference strain, CD630, is a member of the biggest CdeM protein group (WP_003436550) in term of number of strains indicating that the chosen sequence of CdeM (CD630) has a certain length that is shared among most *C. difficile* strains.

1.2.1 Structural analysis of CdeM

Due to a lack of a crystal structure to predict the effect of these mutations on the tertiary structure of the homologs, we analyzed individual gene sequences, in FASTA format, using the I-TASSER structural prediction server. We aligned structures of the reference strain, CD630, with CD196 strain (WP_009893169), as a complete sequence of CdeM, to investigate the role of the deletion on protein structure (Figure A7).

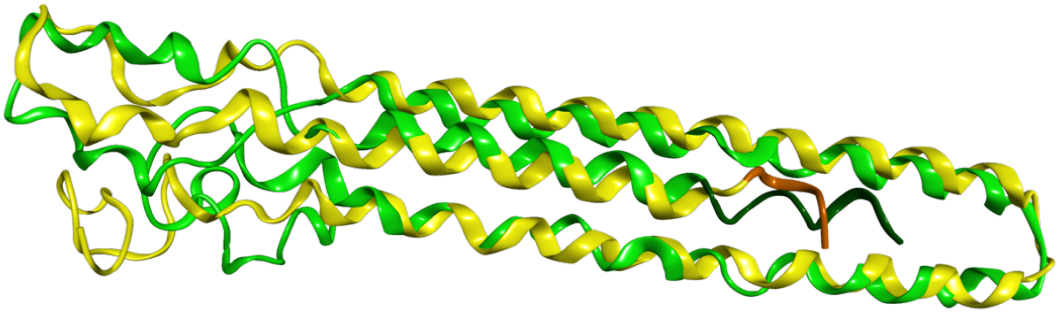


Figure A6.7 CdeM prediction models showing CdeM sequence with deletion aligned and compared to complete CdeM sequence.
The green coloured structure represents CD196 strain, which is a strain with 164 amino acids. The yellow structure represents the CD630, which is the reference strain with 160 amino acids. The N terminus deleted region is highlighted in brown and dark green for CD630 and CD196, respectively. Models were predicted by I-TASSER software, and the modification of the colours and the alignment of the structures were demonstrated by MOE (Molecular Operating Environment) program.

The structure revealed that when the deletions occurred, they were found at the beginning of the N terminus region of the protein which is not predicted to be involved in the stability of the α helix structure. Additionally, it seems not to be involved in the interaction between the helix strands suggesting that this deletion may not impact on protein folding.

On the basis of the analysis described above, the gene sequences for CdeC and CdeM from reference strain *C. difficile* 630 were selected for cloning and expression as recombinant proteins.

Appendix 2 Codon optimised sequences utilized in current study.

The following codon optimised sequences have NdeI restriction site (highlighted in green), BamHI (highlighted in blue), and stop codon (highlighted in red). The codon optimisation was performed by GenScript, UK.

2.1 Endolysins

2.1.1 full length LysCD6356

CATATGAAGGTGGTTCTGACCGCGGGTCACACCCTGACCGGTAAAGGTACCGGTGC
GACCGGTTACATCAACGAGGGCAAGGAAAACCGTATTCTGATGGACCTGGTGGTTA
AATGGCTGAAGAAAGGTGGCGCGACCGTGTACAGCGGCAAGGTTGATAAAAGCAAC
AACTATCTGAGCGAGCAGTGCCAAATCGCGAACAAGCGTAACGTGGACCTGGCGGT
TCAGATTCACTTCAACGCGAACAAAACCACCATCAACCCGATGGGTACCGAGACCA
TTTACAAGACCAACAACGGCAAAGTGTATGCGGAGCGTGTTAACGAAAAGCTGGCG
ACCGTGTTCGAAAACCGTGGTGCGAAAAGCGATGCGCGTGGCCTGTATTGGCTGCGT
CACACCAAGGCGCCGGCGATCCTGATTGAGGTGTGCTTTGTTGACAGCAAAGCGGA
CACCGATTACTATATCCGTCACAAGGATATTGTGGCGAAACTGATCGCGGAAGGTAT
TCTGAACAAGAACATCGACAACAAAGAGAACGGCGAAGATAAGAAAATGTACAAG
CACACCATCGTGTATGACGGTGAAGTTGATAAAATTCTGGCGACCGTTCTGGGTTGG
GGCTACAGCAGCAGCAAGGTGCTGGTTTGCGACATCAAAGATTACATTCCGGGCCA
GACCCAAAACCTGTATGTGGTTGGTGGCGGTGCGTGCGAGAAGATCGGTAGCATTA
CCAAAGAACACTATACCACCATCAAGGGCAACGACCGTTTCGATACCCTGCACAAA
GTTCTGGACTTTATTAAGAAATAA**TAA**GGATCC

2.1.2 EAD of LysCD6356

CATATGAAGGTTGTTCTGACCGCGGGTCACACCCTGACCGGTAAAGGCACCGGTGC
GACCGGCTACATCAACGAGGGCAAGGAAAACCGTATTCTGATGGACCTGGTGGTTA
AATGGCTGAAGAAAGGTGGCGCGACCGTGTACAGCGGCAAGGTTGATAAAAGCAAC
AACTATCTGAGCGAGCAGTGCCAAATCGCGAACAAGCGTAACGTGGACCTGGCGGT
TCAGATTCACTTCAACGCGAACAAAACCACCATCAACCCGATGGGCACCGAGACCA

TTTACAAGACCAACAACGGTAAAGTGTATGCGGAGCGTGTTAACGAAAAGCTGGCG
ACCGTGTTCGAAAACCGTGGCGCGAAAAGCGATGCGCGTGGTCTGTATTGGCTGCGT
CACACCAAGGCGCCGGCGATCCTGATTGAGGTGTGCTTTGTTGACAGCAAAGCGGA
CACCGATTACTATATCCGTCACAAGGATATTGTTGCGAAACTGATCGCGGAAGGCAT
TCTGAACAAGAACATCGAC**TAA**GGATCC

2.2 Spore surface proteins

2.2.1 CdeC

CATATGCAGGACTACAAGAAGAATAAGCGCCGTATGATGAACCAACCGATGAGCAC
CATGAACGAAGAGGAAGTTTACACCGATGAGATTAACAGCGAAGATATGCGTGGCT
TTAAGAAAAGCCACCACCACAACGGTTGCAACACCGACAACAAGTGCGAGTGCCAC
GACGATTGCAACCCGTGCAATCCGTGCAATCCGTGCAAGCCGAACCCGTGCAACCC
GTGCAAACCGAACCCGTGCGACGATAACTGCGGCTGCCACGATAACTGCAAGTGCG
ACTGCGAGCCGTGCGAAATGGACAGCGATGAGTGCTTCGAAAACAAATGCGGTCCG
GAATGCTGCAACCCGATCAGCCCGCGTAACTTCAGCGTGAGCAACGCGGTTCCGTTT
GCGATCGAGGCGAACCGTATTTTCGATACCATGCAGTTCCAAACCTTTACCGATGCG
ACCGGTCCGAACGGTGAACCGCTGACCTTTGAAACCGAGGTGGTTGAGGTGTTTGGC
AGCGTTCCGAGCGCGGGTCAGGCGAGCGTGACCATCGAAAAAATTTGCCTGAGCAA
CGATGGCATCGTTATTGACACCGGTATGACCACCCTGGAGGACTTTGATCTGGACCC
GCTGGGCGATATCGTGGGTCGTAACTGCGAAACCACCTTCGAGTTTGCGGTTTGCGG
CGAACGTAACAGCGAGTGCTGCCGTCAGGGCAAGGGTAAAAGCGTGGCGTACAAGC
AACGTGGTCTGACCGTGGCGGTTCGTAACCTGGTTCTGGAACTGCGTGGCCGTTGCG
GTTGCACCGAGTTTGTTGCGCTGGCGTTTCCGGCGGTTCGTGCGGGTGGCGGTTGCA
AACGTCGTGTGGACTATGTTGAATTCACCTTTAACACCCTGAGCGCGCCGATTTGCC
TGCCGGCGGATGGTCGTGCGGTGACCCTGCGTCAGGAGTACCAAACCAACCTGACC
GTTGACTGCATCGGCAAGAGCATTCTGAAACTGGAGTGCAACGAATGCTGCGAGCC
GTTTTATGAACTGATCATTCCGAACGATATCGACCTGGTGCTGTGCCTGCAGGAAAC
CGTTAGCACCCTGATCAGCGAGCAAATTGTGGTTCTGGCGAGCCCGAACCCGATTCA
ACCGCGTCTGGTGGATACCTTCAGCAAGGTGTGCGACTTTAGCCAATGCGGTCCGAA
TCACGGTAGCGGCAAGCCGAGCTGCCATCGT**TAA**GGATCC

2.2.2 CdeM

CATATGGAGAATAAAAAATGCTACAGCGAAGACTGGTACGAGCGCGGCGAGAGCA
CCGCGAAATGGTTCCAAAACGACCGTGAAGAATACGAGCGTGAGGCGTACGACGAG
GATCGTGAACGTCGTGGTAGCAACTGCGGCTGCAGCGACAGCGGCGAGAACCGTCC
GCGTAACTGCGAACGTTTCCGTCGTGAGGCGGAAATCCGTGAGCGTGAAGCGCGTG
AGGCGTTTTGCGAGAGCAGCGAAAAGAAAAAGGAAGCGCTGGCGTACGAGTGCGA
AGCGCGTAAACTGTGGGAGGAAGCGGAGAAGTACTGGGACGAATACAGCAAGTAC
AACTACAAGGGTATCGAGTATCTGGCGGAAGCGGCGCGTCTGTTCGATGAAGGCAT
GGAGTGCGAAGCGCGTCGTAACGGTAACAACGGCGGTAATAACAATAACTGCTGCC
ACAAATGCCACAAATGCAACTGCAACTGCTGCCGCAAG**TAA**GGATCC

Appendix 3 Primers sequences utilized in current study.

Table A3.1 Summary of characteristics of the forwards and reverse primers used in this study.

	Forward primer	Reverse primer
Sequence	5'- AGCGGATAACAATTCCCCTC -3'	5'- AGTTATTGCTCAGCGGTGG -3'
Length	20 bp	19 bp
GC clamp at the 3' end	Yes	Yes
GC content	50 %	52.6 %
Melting temperature	54.7 ºC and 51.8 ºC	55.3 ºC and 51.1 ºC
Hairpin formation	23.6 ºC	15.6 ºC

Appendix 3 Investigating the availability of a spore binding domain

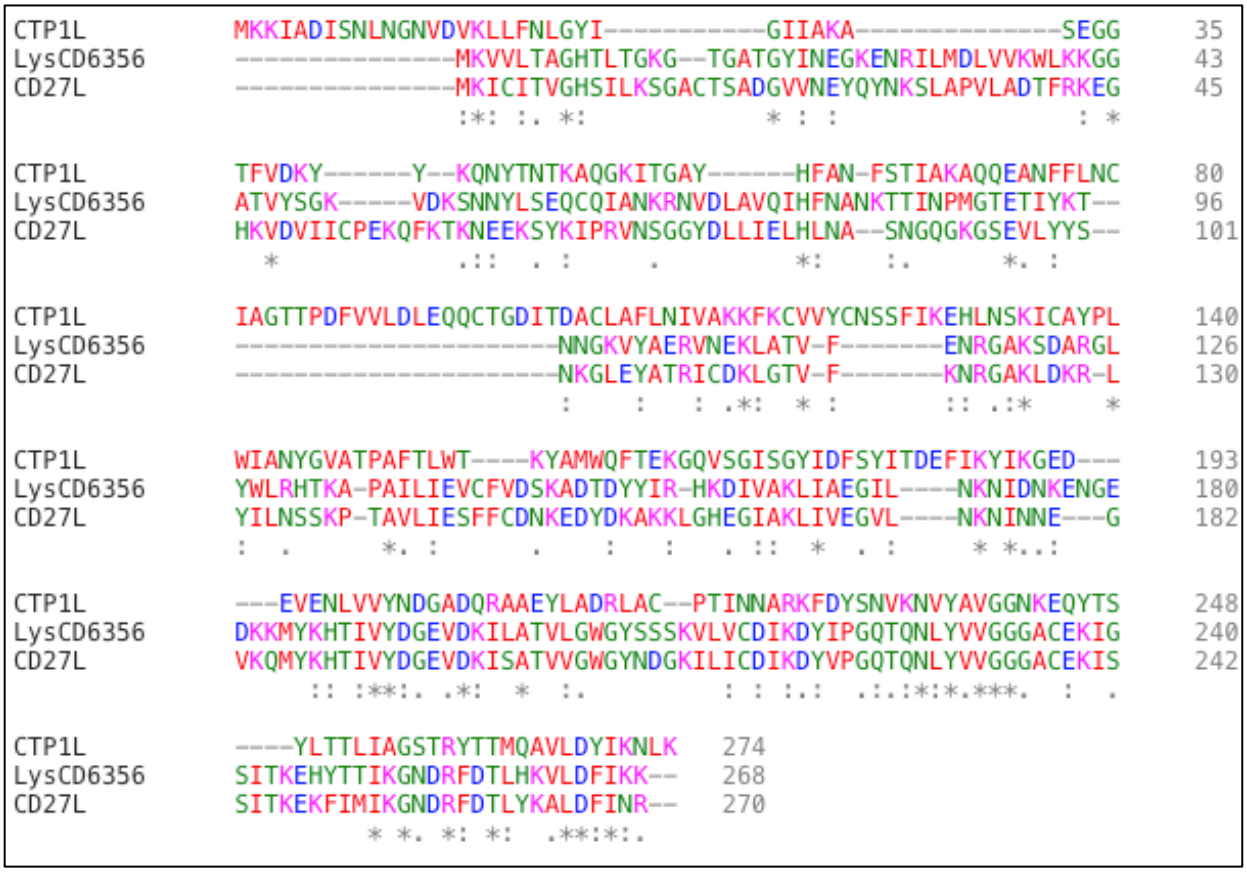


Figure A6.8 Protein sequence alignment of LysCD6356, CD27L and CTP1L using Clustal Omega.

Amino acid sequences of LysCD6356, CD27L and CTP1L were downloaded from NCBI database, and a multiple sequence alignment was performed using Clustal Omega software. Symbol (*) indicates regions of a single conserved residue, (-) indicates regions of sequence shortage, (:) indicates difference in amino acids but within a group of strong similar properties, (.) indicates difference in amino acids but within a group of weak similar properties.

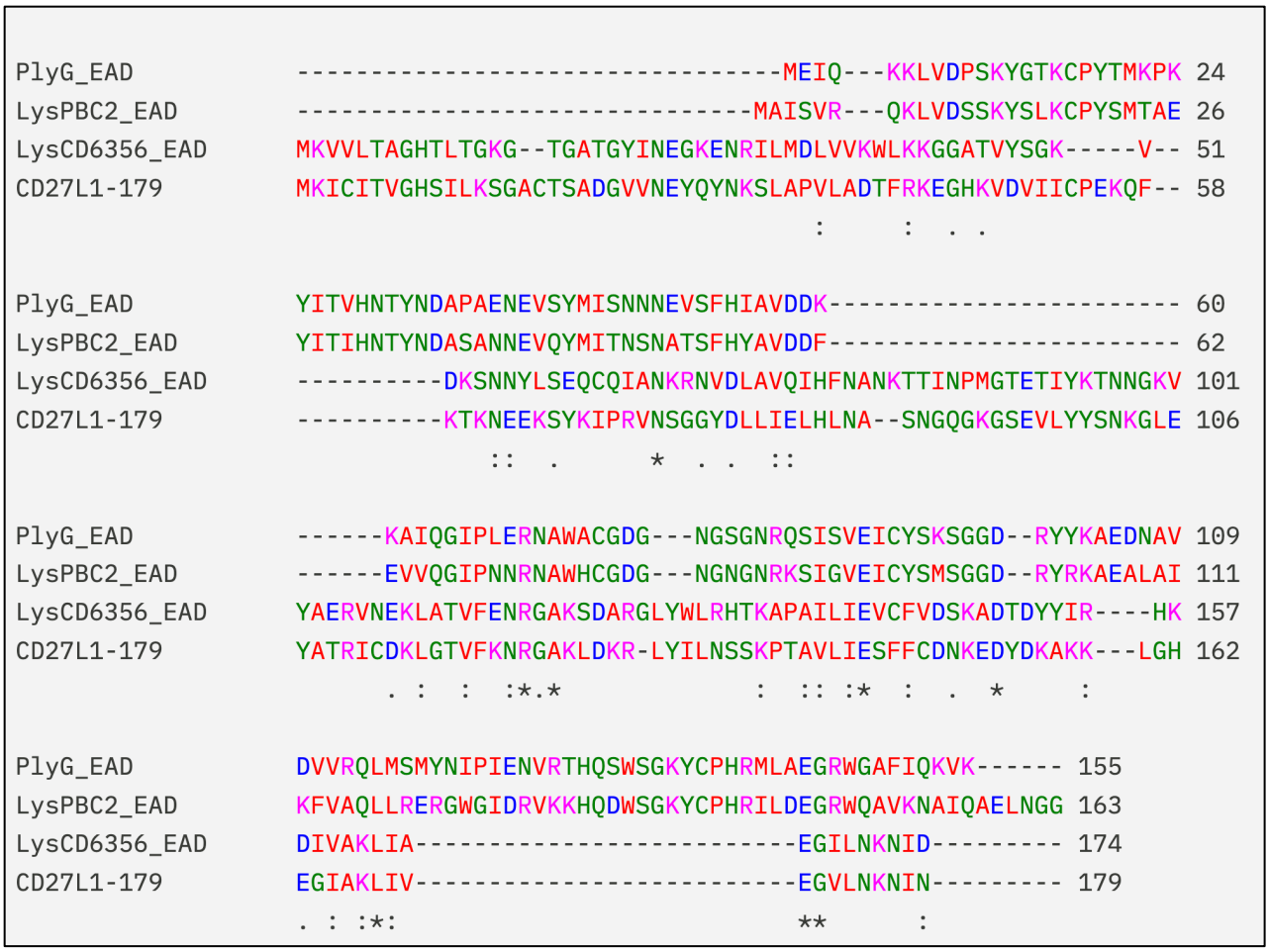


Figure A6.9 Protein sequence alignment of EAD of LysCD6356 and the catalytic domain of LysPBC2 and PlyG using Clustal Omega

Catalytic domains of LysCD6356, LysPBC2, PlyG and $CD27L_{1-179}$ were downloaded from NCBI database and the multiple sequence alignment was performed using Clustal Omega software. Symbol (*) indicates regions of a single conserved residue, (-) indicates regions of sequence shortage, (:) indicates difference in amino acids but within a group of strong similar properties, (.) indicates difference in amino acids but within a group of weak similar properties.

Figure A6.10 Protein sequence alignment of EAD of LysCD6356 and the spore binding domain of LysPBC2 and PlyG using Clustal Omega

Catalytic domain of LysCD6356 was downloaded from NCBI database and the spore binding domain of LysPBC2 and PlyG were taken from literature (Yang et al., 2012; Kong et al., 2019). The multiple sequence alignment was then performed using Clustal Omega software. Symbol (*) indicates regions of a single conserved residue, (-) indicates regions of sequence shortage, (:) indicates difference in amino acids but within a group of strong similar properties, (.) indicates difference in amino acids but within a group of weak similar properties.