The identification and characterization of novel haemolysin genes from *Clostridium difficile*

Thesis submitted by

Zeina Subhi B. Alkudmani

For the degree of

DOCTOR OF PHILOSOPHY

In the

Faculty of Medical Sciences

University College London

Department of Microbial Diseases

UCL Eastman Dental Institute

256 Grey's Inn Road

London WC1X 8LD

UK

2018

Declaration

I hereby certify that the work embodied in this thesis is the result of my own investigation, except where otherwise stated.

Abstract

Clostridium difficile is a Gram positive spore forming bacteria that is the leading cause of antibiotic associated diarrhoea and pseudomembranous colitis. *C. difficile* infection is a significant health care burden in hospital settings and in the community. One of the main factors that lead to *C. difficile* infection is the prolonged use of antimicrobial agents that disturb the microflora of the gut. This allows it to colonize and produce toxins which disrupt the integrity of the intestinal epithelium. Other than toxin production, *C. difficile* has a number of virulence factors including surface proteins and enzymes that aid in bacterial attachment and spread of infection. Another possible virulence factor that was described in clostridia and other bacteria is the production of haemolysins. These are proteins (sometimes lipids) produced by the bacterium and cause cytolysis of red blood and eukaryotic cells for example epithelial cells, leukocytes, lymphocytes, and macrophages.

This work is the first report of *C. difficile* showing haemolytic activity and the aim of this thesis is to identify genes responsible for haemolysis in *C. difficile*.

A CD630 genomic library was constructed in *E. coli* and six genes were found to confer haemolysis. One clone contained a gene (designated *atlf*) encoding a peptide homologous to an anthrax toxin lethal factor which exhibited large zones of haemolysis. Clostron and CRISPR-Cas9 based CD630 mutants were constructed with insertion and deletion in *atlf*

respectively. The haemolysis of the mutants was tested on agar and by a quantitative assay, which revealed no change in the haemolytic phenotype compared to the wild type.

This work suggests that haemolysis is a multifactorial phenomenon in *C. difficile* and may involve several genes. The CRISPR-Cas9 mutagenesis system constructed in this work will aid in many mutagenic studies to understand haemolysis and other virulence mechanisms in *C. difficile.*

Impact statement

The genome of *Clostridium difficile* is very adaptive and many genes have been associated with sporulation, colonisation, virulence, and antibiotic resistance. Our understanding of *C. difficile* and the infection it causes is better today largely due to the development of many molecular techniques including whole genome sequencing, DNA annotation, and genetic manipulation. Nevertheless *C. difficile* infection it still a major concern to health-care systems around the world. It imposes a huge burden due to the high cost of diagnosis, patient care, treatment, and infection control. However, a lot of genes still have unknown functions, many of which might contribute to pathogenesis and spread of infection.

Haemolysins are virulence factors produced by a number of Gram positive and negative bacteria. They can aid in the pathogenesis of bacteria and evasion of host immune responses. This work confirmed that *C. difficile* can show haemolysis on blood agar. Six genes conferring haemolysis when cloned in *E. coli* were identified and one was analysed in depth. Future work studying the genes could provide potential insights into their role in virulence of *C. difficile* and dissemination of infection.

A CRISPR-Cas9 plasmid was designed and used successfully as a selection tool for the deletion of a putative haemolysin encoding gene in CD630. This plasmid can be adapted as a general mutagenesis tool for *C. difficile* to generate deletions, insertions, and substitutions which can contribute to a fuller understanding of the bacterium.

Acknowledgements

Firstly, I would like to express my immense gratitude to my supervisors Prof Peter Mullany and Dr Adam P. Roberts for their continuous guidance and support of my PhD and related research. Their patience, motivation, and vast knowledge helped me complete this thesis.

Beside my supervisors, I would like to thank all the staff and students in the Department of Microbial Diseases at UCL Eastman Dental Institute. I really appreciate everyone answering my questions, teaching me how to do things in the lab, and for making me feel like a member of a loving family. I am also indebted to Dr Haitham Hussain for his help in the technical part of my research. Also I would like to thank Dr John T. Heap and his group at the Imperial College London for kindly providing plasmid pMTL83151.

I would like to acknowledge my sponsor King Saud University, Riyadh, Saudi Arabia and the Saudi Cultural Bureau in London for the funding and the opportunity they gave me to achieve my dream.

Finally and most importantly, no words would be enough to thank my family, my dad, mom, and my brothers and sister for their encouragement and continuous love and support through every step of my life and career.

Table of contents

List of figures

List of tables

Abbreviations

%	Percent
°C	Degrees Celsius
µ	Micro
3D	Three dimensional
ADP	Adenosine DiPhosphate
Amp	Ampicillin
ATc	Anhydrous Tetracycline
ATLF	Anthrax toxin lethal factor
BHI	Brain Heart Infusion
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BSA	bovine serum albumin
Cas	CRISPR-associated sequences
Cbp	collagen-binding protein
CD630	*Clostridium difficile* strain 630
CDC	Centres for Disease Control and Prevention
CDC	cholesterol dependent cytolysins
CDI	*C. difficile* infection
CDRN	*C. difficile* Ribotyping Network
CDSs	coding sequences
CDT	*Clostridium difficile* transferase or Binary toxin
cfu	Colony-forming unit
CIAP	Calf Intestinal Alkaline Phosphatase
Cm	Chloramphenicol
CRISPR	Clustered regularly interspaced short palindromic repeats
CROP	Combined Repeats Oligopeptides
crRNA	CRISPR RNA
CTns	conjugative transposons
CWPs	cell wall proteins
DNA	Deoxyribonucleic Acid
DSB	double strand break
ECDC	European Centres for Disease Control and Prevention

EDTA	Ethylene Diamine Tetra-acetic Acid
EF	Edema factor
EMBL-EBI	European Molecular Biology Laboratory and European Bioinformatics Institute
Erm	Erythromycin
FBP	fibronectin-binding protein
FDA	food and drug administration
FMT	Faecal microbiota transplantation
GO	Gene Ontology
gRNA	Guide RNA
HDR	homology-directed repair
HHH-grade	High yield, High purity, and High activity protein
His	Histidine
HMMs	hidden markov models
IDA	Iminodiacetic Acid
IDSA	Infectious Diseases Society of America
IMAC	Immobilized Metal Affinity Chromatography
IPTG	isopropyl β-D-thiogalactopyranoside
IS	Insertion sequence
Kb	Kilo base pair
kDa	Kilo dalton
Km	Kanamycin
LacZ	β-galactosidase which cleaves lactose
LB	Luria-Bertani
LF	Lethal factor
m	Milli
M	Molar
MEP	mid-exponential phase
mgna	mannosyl-glycoprotein endo-beta-N-acetylglucosamidase
MTn	mobilizable transposon
n	Nano
NAP1	North American PFGE type 1
NHEJ	non-homologous end joining
NHS	National Health Service

Ni-NTA	Nickel-Nitrilotriacetic Acid
NTCD	Non-Toxigenic *C. difficile*
OD	Optical Density
ORF	open reading frame
PA	Protective antigen
PaLoc	Pathogenicity Locus
PAM	proto-spacer adjacent motif
PBS	phosphate buffered saline
PC	phosphatidylcholine
PCR	Polymerase Chain Reaction
Pfam	Protein family
PFGE	Pulsed Field Gel Electrophoresis
PLases	phospholipases
qRT-PCR	Quantitative Reverse Transcription PCR
RAM	retrotransposable activated marker
rATLF	Recombinant ATLF protein
RBCs	Red blood cells
RBS	Ribosome binding site
RE	Restriction endonucleases
REA	Restriction Endonuclease Analysis
RNA	Ribonucleic Acid
rpm	Revolutions per minute
SCOP	Structural Classification of Proteins
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis
sgRNA	Single Guide RNA
SHEA	Society for Health Care Epidemiology of America
SLPs	Surface layer proteins (S-layer)
SM	sphingomyelin
SMART	Simple Modular Architecture Research Tool
SMases	sphingomyelinases
SNPs	single nucleotide polymorphisms
SP	stationary phase
TCD	Toxigenic *C. difficile*

TcdA	*C. difficile* Toxin A
TcdB	*C. difficile* Toxin B
Tet	Tetracycline
Tm	Thiamphenicol
TMHMM	Tied Mixture Hidden Markov Model
Trp	Tryptophan
v/v	Volume for volume
WT	Wild type
x g	Gravitational force
X-Gal	5-bromo-4-chloro-3-indolyl β-D-galactopyranoside
Zn^{2+}	Zinc
α-	alpha
β-	beta
γ-	gamma

Chapter 1

Introduction

Chapter 1 Introduction

1.1 *Clostridium difficile*: The bacterium

Clostridium difficile is a Gram positive, spore-forming, motile, anaerobic bacillus that was first described by Hall and O'Toole in infant faeces in 1935 (Figure 1-1A). It was originally called *Bacillus difficilis* based on its shape and difficulty of isolation in the laboratory (Hall & O'Toole, 1935). On Brain Heart Infusion (BHI) blood agar *C. difficile* has non-haemolytic, flat with irregular-edges, ground-glass like colonies (Figure 1-1B) (Berger, 2018; Edwards *et al.*, 2013). A few decades after its discovery, *C. difficile* was recognized as the leading cause of nosocomial antibiotic associated diarrhoea and pseudomembranous colitis (Bartlett *et al.*, 1978).

In 2009, *C. difficile* was reclassified in Bergey's Manual of Systematic Bacteriology into the family *Peptostreptococcaceae* because it was genetically distinct from its close relatives in the *Clostridiaceae* family (*Clostridium botulinum*, *Clostridium perfringens*, and *Clostridium tetani*) (Figure 1-2) (Ludwig W. *et al.*, 2009). Therefore, it was suggested to change the nomenclature of *Clostridium difficile* to *Peptoclostridium difficile* to avoid confusing it with other *Clostridium spp.* (Yutin & Galperin, 2013). Changing the name to *P. difficile* would be cumbersome, time consuming, commercially expensive, and confusing due to the fact that *C. difficile*'s name is internationally ubiquitous in commercial and clinical settings. Therefore, it was proposed to change the name again from *P. difficile* to *Clostridioides difficile*, which means similar to *Clostridium* (Lawson *et al.*, 2016).

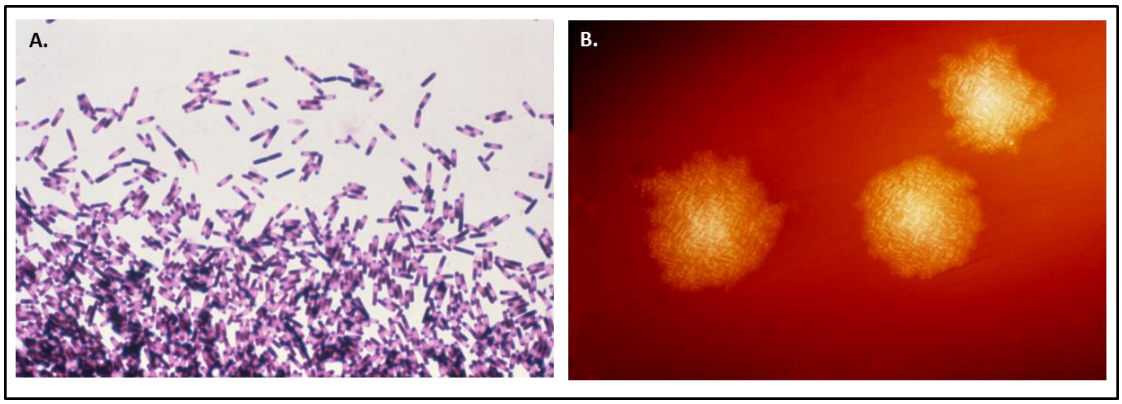


Figure 1-1 Morphology of *C. difficile*.

A- Microscopic image of *C. difficile* showing Gram positive rods with sub-terminal endospores retrieved from (Jones, 1980). **B-** Photograph depicts *C. difficile* colonies after 48 hours growth on blood agar demonstrating a ground-glass appearance, flat and irregular non-haemolytic colonies (Magnified 4.8X). Figure is taken from (Holdeman, 1965).

Figure 1-2 Phylogenetic tree of members of the family Peptostreptococcaceae.

The tree demonstrates the close relationship between *C. difficile* with *C. bifermentans* and *C. sordelii* (indicated by red arrows). Figure is taken with permission from (Lawson *et al.*, 2016).

1.2 *C. difficile* carriage and infection

C. difficile is part of the intestinal microflora in 60-70% of healthy neonates and up to 15% of adults (Jangi & Lamont, 2010; Kato *et al.*, 2001). In a hospital environment, it can be transmitted by ingestion of spores from patients, carriers, or inanimate surfaces (Claro *et al.*, 2014). Once spores are in the intestine, they germinate into the vegetative form in the presence of primary bile acids (Sorg & Sonenshein, 2008). In healthy individuals, the immune system and normal flora provide colonization resistance to *C. difficile* (Sassone-Corsi & Raffatellu, 2015). But when the microbiota of the colon is disrupted or lost, due to mainly prolonged treatment with broad-spectrum antibiotics, this provides an opportunity for *C. difficile* to colonize and produce its toxins that lead to *C. difficile* infection (CDI) (Smits *et al.*, 2016; Theriot *et al.*, 2016). Most known antibiotics have been associated with CDI especially cephalosporins, fluoroquinolones, and clindamycin (Higa & Kelly, 2014; Slimings & Riley, 2014). Symptoms of CDI vary between mild self-limited diarrhoea to severe pseudomembranous colitis (Abt *et al.*, 2016). The risk of complications is 11% in infected individuals and include toxic megacolon, perforation, systemic inflammatory response syndrome, and multisystem organ failure while the risk of all-cause 30-day mortality was estimated as 13% (Abou Chakra *et al.*, 2015; Carter *et al.*, 2015; Dobson *et al.*, 2003). Recurrent CDI occurs in 20-30% of treated patients, and out of these 40-60% will develop second recurrence. The recurrence of CDI is mainly a relapse of the same *C. difficile* strain but sometimes it can be re-infection by a different strain (Kamboj *et al.*, 2011).

1.3 Epidemiology of *C. difficile*

The incidence of CDI was reasonably stable until the late 1990s where it started to rise and become a bigger health and financial burden (Archibald *et al.*, 2004; Elixhauser & Jhung, 2006; Pépin *et al.*, 2004). According to the Centres for Disease Control and Prevention (CDC), the number of CDI cases is 250,000 per year resulting in 14,000 deaths and excess medical costs of $1 billion annually (CDC, 2013).The change in *C. difficile* epidemiology was attributed to the emergence of a hypervirulent strain (Wilcox *et al.*, 2012). In 2002, an epidemic of CDI was reported in more than 30 hospitals in Montreal and southern Quebec, Canada. The hypervirulent strain 027 (80% of isolates) had a common profile: group BI by restriction endonuclease analysis (REA), toxinotype III, produced toxins A and B, produced binary toxin, and had an 18-bp deletion (position 330 to 347) in the *tcdC* gene (Labbe *et al.*, 2008). The same strain was responsible for outbreaks in six states in the US between 2000 and 2003 (McDonald *et al.*, 2005). In the United Kingdom, two outbreaks with predominance of the same profile occurred in the Stoke Mandeville Hospital (Buckinghamshire Hospitals NHS Trust) between 2003 and 2005 (Freeman *et al.*, 2010; Smith, 2005). This strain was characterized in 2005 as toxinotype III, North American pulsed field gel electrophoresis (PFGE) type 1 (NAP1), PCR-ribotype 027, and was associated with *in vitro* increased production of toxins A and B as compared to the reference toxinotype 0 (16 and 23 times higher respectively) (Warny *et al.*, 2005). The overproduction of toxins might be related to the second mutation, single nucleotide deletion at position 117 in the *tcdC* gene. This

deletion resulted in a frameshift and a premature stop in the early part of this gene that would imply disruption of its function (MacCannell *et al.*, 2006). Also the presence of five genetic regions not found in CD630 and non-epidemic 027 (CD196) might explain its epidemic nature. This includes a Tn*6103* (previously described as phage island), two component regulatory system and transcriptional regulators (Brouwer *et al.*, 2011; Stabler *et al.*, 2009).

Due to these outbreaks, the CDC and the European CDC (ECDC) started reporting the incidence of ribotype 027 and its spread in a number of North American and European countries. Reports showed outbreaks in 40 US states and 16 European countries including the UK (Kuijper E J *et al.*, 2007; Kuijper *et al.*, 2008; O'Connor *et al.*, 2009). Between 2005 and 2007, ribotype 027 was one of the three predominant ribotypes in the UK (25.9%), whereas between 2007 and 2008 it was the most predominant with 41.3% of ribotypes isolated (Cartman *et al.*, 2010). In 2007, England started an enhanced surveillance and a *C. difficile* Ribotyping Network (CDRN) to access ribotyping centrally. Since then reported CDI cases and deaths has decreased significantly probably due to the increased awareness and enhanced diagnosis and prevention of CDI (Figure 1-3) (Freeman *et al.*, 2010; Health Protection Agency, 2016).

Figure 1-3 Global epidemiology of *C. difficile*.

a) The global distribution of common *C. difficile* ribotypes are indicated in the pie charts. The prevalence of ribotype 027 is high in Europe and USA while it is almost absent in Asia and Australia. **b)** The distribution of common *C. difficile* ribotypes in England after the mandatory surveillance shows decrease in the prevalence of ribotype 027 and increase in the diversity of other ribotypes. Figure is taken with permission from (Martin *et al.*, 2016).

1.4 CDI in the community

Although *C. difficile* was mainly known to cause hospital-acquired infection, incidence of CDI in the community has emerged. In 2015, the CDC issued a report indicating that 39% of cases of CDI (2,787 out of 7,092) were community acquired and among these 8% were ribotype 027 (CDC, 2015). Between 2010 and 2012, Western Australia had an outbreak of *C. difficile* ribotype 244 and 56% of the cases were community-associated. The ribotype 244 isolates were very similar to ribotype 027 in toxin production and hypervirulence but analysis of the whole genome showed that they are substantially different (Eyre *et al.*, 2015). It has been noted that a lot of the community CDI patients were not previously hospitalised, elderly, or took antibiotics (risk factors known for hospital-acquired infection) which raised questions about the source of CDI in the community (Leffler & Lamont, 2012). Few reports linked it to food and animal origin as spores of toxigenic *C. difficile* strains were isolated from meat products in Canada and the US and from retail salads in Scotland and France (Bakri *et al.*, 2009; Eckert *et al.*, 2013; Gould & Limbago, 2010; Rodriguez-Palacios *et al.*, 2007; Rupnik, 2007; Songer *et al.*, 2009). It is possible that contaminated environments (soil or irrigation water) could be the source of spores in food products as seen in public lawns in Australia (Moono *et al.*, 2017). *C. difficile* is known to cause disease in animals like mature horses, foals, hamsters, pigs, elephants, calves, rabbits, guinea-pigs, dogs, and laboratory animals (Baverud, 2002; Keessen *et al.*, 2011). Also some animals such as sheep, lambs, poultry, camels, donkeys, and domestic animals were found to be carriers for *C.*

difficile (Moono *et al.*, 2016). But to date there is no correlation between isolates from animals and isolates from human.

1.5 Spores of *C. difficile*

Spores of *C. difficile* are crucial for the transmission of CDI. It was even suggested that spores play a role in attachment of *C. difficile* to intestinal cells (Paredes-Sabja & Sarker, 2012). The sporulation triggers are not yet known but most probably are related to environmental stress such as nutrient limitation (Paredes-Sabja *et al.*, 2014). The sporulation rate of *C. difficile* strains varies and hypervirulent types do not have a higher capacity to produce spores (Burns *et al.*, 2011). Spores are heat stable and can survive unfavourable environmental conditions such as gastric acidity, antibiotics, and disinfectants (Lawley *et al.*, 2010; McFee & Abdelsayed, 2009; Rodriguez-Palacios & Lejeune, 2011). As a result, they can persist on contaminated surfaces and medical equipment and are readily transmitted to patients in hospital settings (Dumford *et al.*, 2009; Gerding *et al.*, 2008). The ultrastructure of the *C. difficile* spore is very similar to *Bacillus* and *Clostridium* genus spores in that it is composed of several protective layers to resist environmental stress (Figure 1-4) (Gil *et al.*, 2017).

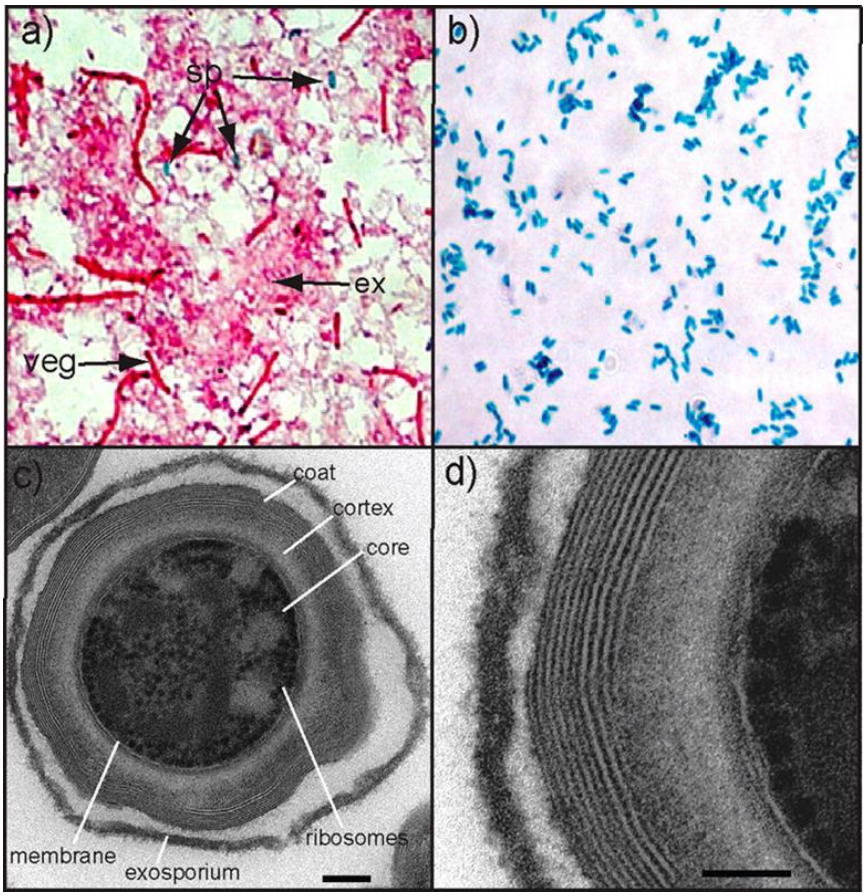


Figure 1-4 *C. difficile* spores.

a) Microscopic image of *C. difficile* endospores in green (sp) from a smear showing *C. difficile* vegetative cells (veg) and extracellular matrix (ex). **b)** Microscopic image of purified spores stained in green. **c)** Ultrastructure of *C. difficile* spore under electron microscope demonstrating the core with ribosomes and the layers: membrane, cortex, coat, and exosporium (bar 100 nm). **d)** Outer surface of the spore in higher magnification (bar 50 nm). Figure is taken from (Lawley *et al.*, 2009).

1.6 Virulence factors of *C. difficile*

C. difficile has several properties that enable it to colonize and cause infection including toxins, surface proteins, and hydrolytic enzymes.

1.6.1 Cytotoxins

The main virulence factor in *C. difficile* are two potent toxins, toxin A (TcdA) and toxin B (TcdB) (Taylor *et al.*, 1981). They belong to the family of large clostridial glucosylating toxins (LCGT) along with *C. sordellii* lethal (Tcsl) and hemorrhagic (TcsH) toxins, *C. novyi* α-toxin (TcnA), and *C. perfringens* type C large cytotoxin (Tcpl) (Genth & Just, 2015; von Eichel-Streiber *et al.*, 1996).

The genes encoding toxins A and B, *tcdA* and *tcdB* respectively, are located in the pathogenicity locus (PaLoc) with three additional genes *tcdE*, *tcdR*, and a *tcdC* (Figure 1-5A) (Hammond & Johnson, 1995). *C. difficile* strains that do not produce toxins, termed non-toxigenic *C. difficile* (NTCD), are incapable of causing CDI and have a non-coding region of 115 bp instead of the PaLoc region (Braun *et al.*, 1996; Brouwer *et al.*, 2012; Natarajan *et al.*, 2013). The *tcdE* encodes a bacteriophage holin-like pore-forming protein required for toxin secretion and *tcdR* encodes an RNA polymerase sigma factor which is a positive regulator of toxin expression (Govind & Dupuy, 2012; Mani & Dupuy, 2001). The role of *tcdC* in regulation of toxin production in *C. difficile* is still controversial. Some studies have suggested a negative regulatory effect because it is transcribed in the opposite direction to the other PaLoc genes (Figure 1-5A) (Cohen *et al.*, 2000; Hundsberger *et al.*,

1997; Matamouros *et al.*, 2007). Whereas subsequent studies are in favour of a modulatory effect of *tcdC* on toxin levels (Merrigan *et al.*, 2010). The ClosTron-based mutant of *tcdC* constructed in CD630Δerm showed minor differences in the transcription levels of the PaLoc genes and non-significant difference in the total toxin expression levels. Which suggest that in CD630Δerm TcdC is not a major regulator of toxin expression under the conditions tested (Bakker *et al.*, 2012). This was further confirmed in CD630 and R20291 as two mutants were created by allele exchange using the heterologous counter selection marker, *codA*. The CD630 mutant (with deletion of *tcdC*) and the R20291 mutant (with restoration of the 18-bp deletion and Δ117 frame-shift mutation) had no difference in cell growth and toxin production compared to their wild type (Cartman *et al.*, 2012).

The toxins are structurally composed of an N-terminal catalytic domain, a translocation domain, and a C-terminal binding domain (Figure 1-5B) (Jank *et al.*, 2007). Some studies have suggested a four-domain or ABCD structure for the toxins where the cysteine protease is considered the fourth domain (Albesa-Jové *et al.*, 2010; Jank & Aktories, 2008; Pruitt *et al.*, 2010). The binding domain is composed of repeating polypeptide units that fold into a screw-like superstructure to increase the protein binding surface to the epithelial receptor (von Eichel-Streiber *et al.*, 1992). TcdA uses glycoprotein 96 (gp96) as a co-receptor, whereas TcdB binds to a Wnt receptor (Wingless/Integrated), that belong to the frizzled family of proteins (FZDs), on the colonic epithelium with the aid of an endogenous FZD-bound fatty acid co-receptor (Chen *et al.*, 2018; Na *et al.*, 2008; Tao *et al.*, 2016). The binding

domain is also involved in the protection of toxin conformation (Olling *et al.*, 2014). After receptor binding, toxin A enters host cells via dynamin-dependent endocytosis, that is mediated by PACSIN2 (protein kinase C and casein kinase substrate in neurons protein 2), while toxin B enters via clathrin-dependent endocytosis (Chandrasekaran *et al.*, 2016; Papatheodorou *et al.*, 2010). In the endosome, low pH causes the toxin to form pores that will facilitate the transfer of the catalytic domain to the cytosol (Barth *et al.*, 2001; Orrell *et al.*, 2017). The cysteine protease will be activated by reducing conditions and the binding of cytosolic inositol hexakisphosphate (IP_6) leading to the cleaving of the catalytic domain which is the only part taken up to the cytosol (Egerer *et al.*, 2007; Pfeifer *et al.*, 2003). Another hypothesis describes an aspartate protease activity as being responsible for the processing of TcdB (Reineke *et al.*, 2007). Once inside the cytosol, the N-terminal glucosyltransferase domain targets the regulatory proteins (Rho-GTPases such as RhoA, Rac1, and Cdc42) of the eukaryotic actin cytoskeleton (Sehr *et al.*, 1998; Zeiser *et al.*, 2013). Inactivation of the Rho-GTPases causes disaggregation of the microfilament cytoskeleton leading to complete disruption in the integrity of cells, inflammation, damage to mucous membranes, cell rounding, and death (Figure 1-6) (Just *et al.*, 1995a; Just *et al.*, 1995b; Kasendra *et al.*, 2014). *C. difficile* toxins also cause the early release of proinflammatory cytokines and chemokines which in turn activate macrophages, enteric nerves and sensory neurons, leading to an influx of inflammatory cells, which add to the fluid secretion and intestinal inflammation (Cowardin *et al.*, 2016b; Popoff, 2018).

Figure 1-5 Schematic representation of the pathogenicity locus in CD630.

A. The PaLoc contain genes encoding toxins A and B (*tcdA* and *tcdB* respectively) in addition to three other genes (*tcdC*, *tcdE*, and *tcdR*) responsible for toxin genes regulation. **B.** The TcdB protein consist of three domains: N-terminal catalytic domain, the translocation domain, and C-terminal binding domain. The catalytic domain consists of a conserved tryptophan (Trp), a DXD motif responsible for the enzymatic activity, and a substrate specificity region. The translocation domain consists of the middle hydrophobic region, and the cysteine and aspartate proteases. The binding domain has combined repeats oligopeptides (CROP). Figure is taken with permission from (Rupnik *et al.*, 2009).

Figure 1-6 Schematic representation of the effect of TcdA and TcdB on colonic epithelial cells.

After the toxin binds to the colonic receptor, it is endocytosed. Then self-proteolytic enzymes cleave the catalytic glucosylating domain which in turn inactivate the host skeletal regulatory proteins GTPases (Rho and Rac) leading to loss of the tight junctions and disruption of the epithelial layer. GTD; glucosyltransferase domain, CPD; cysteine protease domain, CROP; combined repeats oligopeptides, IP_6; cytosolic inositol hexakisphosphate. Figure is taken with permission from (Abt *et al.*, 2016).

It is not clear whether one or both toxins A and B are responsible for the pathogenicity of *C. difficile.* Lyras *et al.* constructed isogenic toxin A and B mutants in a derivative of CD630 (JIR8094 or CD630E) by homologous recombination. The study suggests that toxin B is essential for virulence and that a strain producing toxin A only is avirulent (Lyras *et al.*, 2009). On the other hand, Kuehne *et al.* used ClosTron mutagenesis to construct three CD630Δerm mutants: A^-B^+ toxin A inactivated, A^+B^- toxin B inactivated, and A^-B^- both toxins inactivated. It was found that mutants producing one toxin were as virulent as the wild type A^+B^+, while the mutant with both toxins inactivated was not virulent. This study provided an insight on the importance of both toxins for *C. difficile* to cause infection (Kuehne *et al.*, 2010). The contradiction between these two studies might be related to the use of different mutagenesis methods and CD630 strains as well as inherent differences in hamster models (details of strain variation and mutagenesis methods are in section 1.7 and 1.8 respectively).

C. difficile strains are classified into groups, by toxinotyping, according to variations in PaLoc compared to the reference strain VPI 10463 in which toxin genes were sequenced first (toxinotype 0). There are 34 toxinotypes designated with roman numerals (I to XXXIV) (Rupnik & Janezic, 2016). The PaLoc is always present in the same location in the toxigenic strains belonging to clade 1 (Braun *et al.*, 1996). The work by Monot *et al.* has revealed some discrepancies between toxigenic and non-toxigenic *C. difficile* within clades 1, 5, and C-I regarding the location of the PaLoc (Figure 1-7). Three new variant strains RA09-70, SA10-050, and CD10-165 have a 75 bp

sequence replacing the PaLoc region followed by 5 open reading frames (ORF) whereas the alternative PaLoc was in another location in the genome. The atypical RA09-70 strain has a smaller PaLoc (10.5 Kb) and only produces TcdA (A+B-). In SA10-050 and CD10-165 strains, the PaLoc region was present next to the binary toxin genes (discussed below section 1.6.2) (Monot *et al.*, 2015).

The PaLoc was capable of transfer from a toxin-producing strain (CD630Δerm) to non-toxigenic *C. difficile* by horizontal gene transfer. This transfer caused CD37 to produce similar levels of toxin B as in CD630Δerm demonstrated by an *in vitro* cytotoxicity assay (Brouwer *et al.*, 2013).

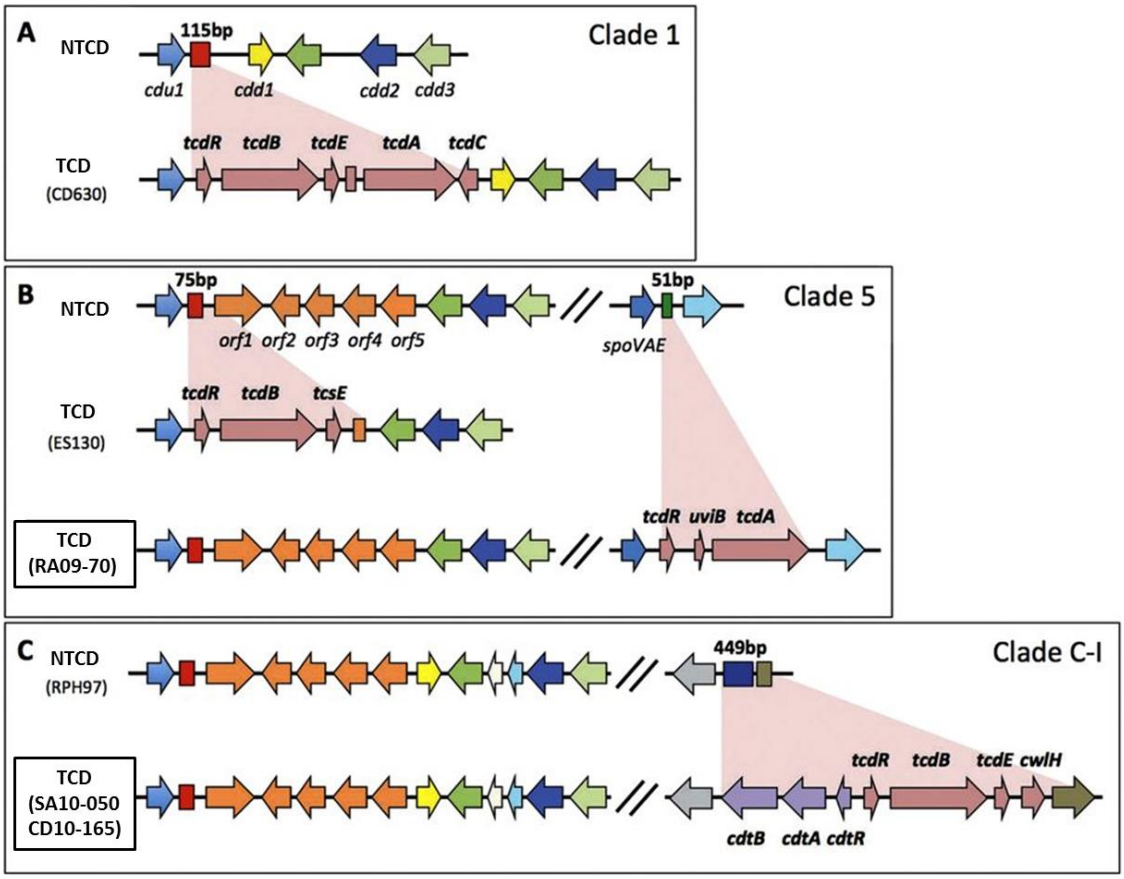


Figure 1-7 Schematic representation of PaLoc region in three *C. difficile* clades.

The PaLoc operon in toxigenic *C. difficile* (TCD) and the sequence replacing PaLoc in non-toxigenic *C. difficile* (NTCD) are represented in clades 1 (A), 5 (B), and C-I (C). The strains in the boxes are the new variants discovered with PaLoc present somewhere else in the genome. RA09-70 only produces TcdA (A^+B^-) and SA10-050 and CD10-165 have their PaLoc next to the CdtLoc. Figure is taken from (Monot *et al.*, 2015).

1.6.2 Binary toxin

A third toxin, *Clostridium difficile* transferase (CDT), is a member of the actin-ADP-ribosylating family of toxins along with *C. botulinum* C2 and *C. perfringens* E iota toxins (Popoff *et al.*, 1988). CDT, also known as binary toxin, is produced by 17-23% of *C. difficile* strains (Eckert *et al.*, 2015). CDT is a two subunit toxin: CDTa and CDTb encoded by genes *cdtA* and *cdtB* respectively, located outside the PaLoc region with a regulatory gene *ctdR* (Figure 1-8A) (Perelle *et al.*, 1997). The region is known as the CDT locus (CdtLoc) and it is present in toxigenic *C. difficile* strains only (Figure 1-9) (Carter *et al.*, 2007). CtdR is a LytTR response regulator that up-regulates the production of CDT to cytotoxic levels in *C. difficile* (Bilverstone *et al.*, 2017; Carter *et al.*, 2007). When *cdtA* and *cdtB* were introduced to CD37 (non-toxin producing strain) and CD630 (contain *cdtR*), the level of CDT was 15 times less in CD37 derivative. And when *ctdR* was introduced to the CD37 derivative, the level of CDT production increased 17-fold (Carter *et al.*, 2007).

The binary toxin components: CDTa, which is the active enzymatic component, and CDTb, the host binding component are illustrated in Figure 1-8B (Perelle *et al.*, 1997). CDTa has an N-terminus adaptor domain which is probably involved in the interaction with CDTb and the C-terminus enzymatic domain (Sundriyal *et al.*, 2009). The crystal structure of CDTb is not available but the general characteristics can be predicted from *C. botulinum* C2 toxin.

Figure 1-8 Schematic representation of the structure of the *C. difficile* binary toxin.

A. The genes encoding the binary toxin are located in the CDT locus outside the PaLoc: *cdtA* which encodes the catalytic domain and *cdtB* which encodes the binding domain. The CdtLoc contains a third gene (*cdtR*) which encode a protein that upregulate the expression of the toxin genes. Figure is taken with permission from (Rupnik *et al.*, 2009). **B.** The structural domains of CDTa and CDTb. CDTa consists of: adaptor and enzymatic domains responsible for binding to CDTb and the catalytic function respectively. CDTb consists of four domains: proteolytic activation, membrane insertion and pore formation, oligomerization, and receptor binding. Figure is taken from (Gerding *et al.*, 2014).

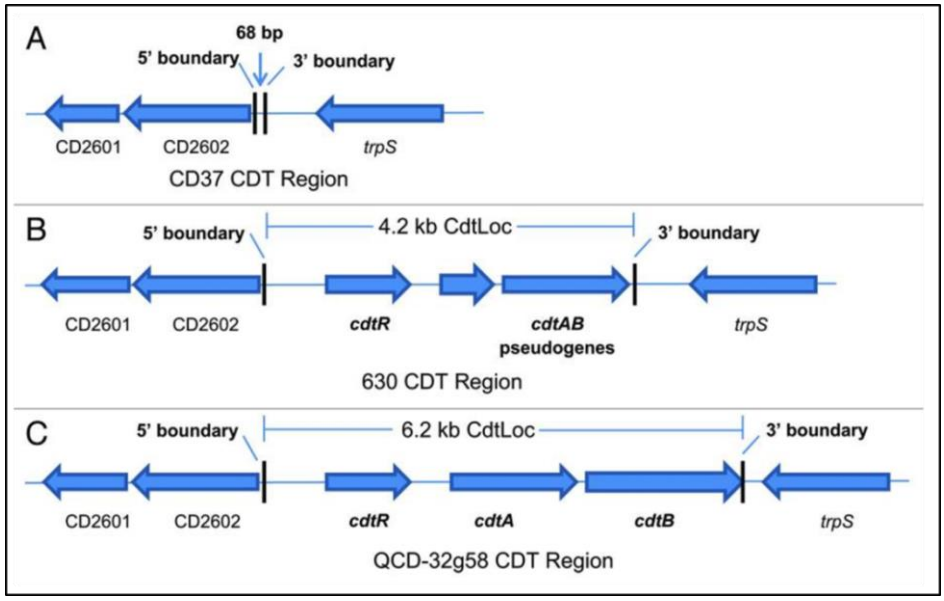


Figure 1-9 Schematic representation of CDT locus in *C. difficile*.

A. Non-toxigenic strain CD37 where the Cdt Loc is replaced by 68 bp. **B.** Toxigenic strain CD630 that does not produce the binary toxin but contains a truncated region with *cdtR* and *cdtAB* pseudogenes. **C.** Toxigenic strain QCD-32g58 that produces binary toxin and has the full length (6.2 Kb) region with all three intact genes. Figure is taken from (Gerding *et al.*, 2014).

The C2 toxin of *C. botulinum* binding component has four domains: the N-terminus which is the activation domain I, domain II is involved in host membrane insertion and pore formation, domain III is responsible for oligomerization, and domain IV (C-terminus) is involved in receptor binding (Schleberger *et al.*, 2006).

The binding component has to be activated by a serine-type protease (chymotrypsin) as for all ADP-ribosylating toxins (Barth *et al.*, 2004). When CDTb is activated, it binds to the host receptor, Lipolysis-stimulated Lipoprotein Receptor (LSR), forming a prepore-receptor complex (Papatheodorou *et al.*, 2011). After that CDTa binds to this complex and the toxin is engulfed and transported into the host cell by endocytosis. This triggers the formation of a membrane pore where CDTa moves to the cytosol. Then it catalyzes the ADP-ribosylation of G-actin and as a result disrupts the actin cytoskeleton and causes the host cell to die (Figure 1-10) (Hemmasi *et al.*, 2015).

The role of CDT in the development of CDI is not well understood. CDT could induce the formation of protrusions on the intestinal epithelium containing long microtubules, which causes adherence of *C. difficile* to Caco-2 cells and mouse epithelium (Schwan *et al.*, 2009). The protrusions contain trafficking vesicles and endoplasmic reticulum that reroutes and releases fibronectin which in turn enhances adherence of bacteria (Schwan *et al.*, 2014). CDT was found to be cytotoxic to Vero cells when expressed and purified from *Escherichia coli* (Sundriyal *et al.*, 2010). CDT was associated with increased inflammation in mouse model by supressing the eosinophilia that protects the

colon which might be related to hypervirulence observed in Ribotype 027 (Cowardin *et al.*, 2016a).

Figure 1-10 Schematic representation of the mechanism of action of CDT.

The binding component CdtB will recognise the Lipolysis-stimulated Lipoprotein Receptor (LSR) of the colonic cells leading to the endocytosis of the binary toxin into the cytosol. The enzymatic component CdtA will prevent the polymerisation of G-actin which will cause disruption of the G-actin-F-actin equilibrium in the cell and eventually the cytoskeleton. This results in excessive fluid loss, rounding and finally death of the target cell. Figure is taken with permission from (Abt *et al.*, 2016).

1.6.3 Surface proteins

Surface proteins are very important for the colonization of *C. difficile*. Some function as adhesins to attach *C. difficile* to receptors on the surface of host cells. These include: cell wall proteins (CWPs), fibronectin-binding protein (FBP), collagen-binding protein (Cbp), lipoprotein CD0873, heat shock protein, flagellar proteins, and pili (Le Monnier *et al.*, 2014).

CWPs are a family of surface proteins, which comprise 29 members, all contain a conserved binding domain (for host binding) and a second domain with a predicted function that varies between *C. difficile* strains (Fagan *et al.*, 2011; Karjalainen *et al.*, 2001). The most studied of these proteins are the SlpA, Cwp84, and Cwp66.

Surface layer proteins (SLPs or S-layer) are a large and diverse family of proteins that promote attachment of the bacterium to host cells (Calabi *et al.*, 2002; Merrigan *et al.*, 2013). They are made of two protein units: high molecular weight P47 and low molecular weight P36 proteins (Cerquetti *et al.*, 2000; Fagan *et al.*, 2009). They are encoded by the *slpA* gene and cleaved by Cwp84, a cysteine protease, from a single precursor SlpA (Calabi *et al.*, 2001; Kirby *et al.*, 2009). These proteins form a complex para-crystalline structure that covers the entire surface of the bacterium (Sleytr *et al.*, 2014). SLPs were found to also regulate immune responses by activating the production of cytokines and chemokines in addition to increased macrophage migration and phagocytotic activity *in vitro* (Ausiello *et al.*, 2006; Collins *et al.*, 2014). Moreover, SlpA was found to have important roles in

sporulation, resistance to innate immunity effectors, and toxin production (Kirk *et al.*, 2017).

Cwp66 is an adhesin that is encoded by *cwp66* found within the CWP gene cluster (Waligora *et al.*, 2001). The N-terminus anchor part is embedded in the *C. difficile* cell wall keeping the C-terminus exposed to the surface for host cell attachment (Pechine *et al.*, 2005). Cwp84 is discussed in more detail in section 1.6.4.

Fibronectin-binding protein Fbp68 (or FbpA) promotes attachment by binding to soluble fibronectin in plasma and to immobilized fibronectin and fibrinogen present on host cells (Barketi-Klai *et al.*, 2011; Hennequin *et al.*, 2003). Collagen-binding protein CbpA binds with type I and V collagen fibres produced by human fibroblasts and mouse intestinal tissues (Tulli *et al.*, 2013). CD0873 is a surface-associated lipoprotein that was characterized as adhesin by bioinformatic analysis and when tested on Caco-2 cells although *in vivo* analysis is needed (Kovacs-Simon *et al.*, 2014).

Heat shock protein, GroEL, is a member of the Hsp60 family of chaperons. The *groESL* operon consists of *groES* and *groEL* genes the latter of which encodes a surface protein that functions in adherence to cultured cells (Hennequin *et al.*, 2001).

Flagellum-mediated motility in bacteria is considered as an essential virulence factor rendering the pathogen capable of evading host immune response, stress, and harsh environmental conditions and moving towards

the site of colonisation or nutrients (Moens & Vanderleyden, 1996). The genes responsible for the assembly of flagella in *C. difficile* are organised into three operons F1, F2, and F3 (Stabler *et al.*, 2009). F1 comprises genes that encode late-stage flagellar proteins including filament protein FliC and cap protein FliD which are structural proteins essential to the formation of fully functional flagella (Tasteyre *et al.*, 2000; Tasteyre *et al.*, 2001). The F3 contains early flagellar genes such as *fliA* that encode FliA sigma factor (also known as SigD) which is thought to regulate the F1 genes (El Meouche *et al.*, 2013). The F2, interflagellar gene region, comprise of flagellar-biosynthetic glycan genes important for the assembly of the flagella and motility of the bacterium (Faulds-Pain *et al.*, 2014; Twine *et al.*, 2009). The role of flagellum and flagellar cap gene (*fliC* and *fliD*) in the colonization and virulence of *C. difficile* is not well understood and might be strain dependant. The ClosTron *fliC* and *fliD* mutants of 630Δerm showed loss of motility, increased adherence to Caco2 cells, and increased virulence in a hamster model (Dingle *et al.*, 2011). While *fliC* and *fliD* mutants of R20291 displayed impaired adherence to Caco2 cells and mouse caeca (Baban *et al.*, 2013). However, the observation by Dingle *et al.* was contradicted in mice as ClosTron *fliC* mutant of 630Δerm showed no difference in colonisation and adherence compared to the wild type (Baban *et al.*, 2013). Flagella can work in conjunction with toxins to induce epithelial inflammatory response and exert mucosal lesions in mice (Batah *et al.*, 2017).

C. difficile contains multiple surface appendages known as type IV pili (also called fimbriae) (Borriello *et al.*, 1988). The structural subunits are called pilin

and genes encoding all proteins required for pili formation are organised in a gene cluster (CD630_35030 to CD630_35130) (Sebaihia *et al.*, 2006; Varga *et al.*, 2006). In bacteria, type IV pili mediate adherence, colonization, DNA transfer, and twitching motility (Maldarelli *et al.*, 2014). Mutants in *pilA1* (encoding the major pilin) or *pilB1* (encoding the pilus biosynthesis ATPase) in CD630Δerm were tested for adherence to epithelial cell monolayers *in vitro* and mice epithelia. Type IV pili was not required for attachment but contributed to prolonged adherence under the control of cyclic di-GMP regulation (McKee *et al.*, 2018).

1.6.4 Enzymes

C. difficile is capable of producing hydrolytic enzymes that aid in colonization and pathogenicity during infection. Cwp84, a cell wall protease, is a member of the cysteine protease family present on the bacterial surface in association with S-layer proteins and mainly responsible for the maturation of SlpA (Bradshaw *et al.*, 2014; Kirby *et al.*, 2009). It was also found to be involved in the cleavage of some host mucosal extracellular matrix proteins such as fibronectin, laminin, and vitronectin which could affect the integrity of enterocytes thus assisting in the spread of CDI (Janoir *et al.*, 2007). Other hydrolytic enzymes produced by *C. difficile* that are thought to enhance colonization include hyaluronidase, chondroitin-4-sulphatase, and collagenase (Seddon *et al.*, 1990). A study suggested that high-temperature requirement A (HtrA) protease in *C. difficile* has a modulatory effect on virulence. Analysis of the ClosTron-based *htrA* mutant of CD630Δerm

revealed up-regulation of the *tcdA* gene, reduced spore formation, and decreased adherence to colonic cells (Bakker *et al.*, 2014).

1.7 *C. difficile*'s genome

CD630 was the first *C. difficile* to be sequenced and annotated followed by a number of strains including: CD37, R20291, M68, CF5, M120, G46, 196, 2007855, and BI1 (Brouwer *et al.*, 2012; Gaulton *et al.*, 2015; He *et al.*, 2010; Stabler *et al.*, 2009). CD630 has a circular chromosome of approximately 4.29 Mbp with a G+C content of 29% and 3,776 predicted coding sequences (CDSs). The genome also contains a plasmid (pCD630) of 7.88 Kb and 27.9% G+C content (Sebaihia *et al.*, 2006).

The genome of CD630 was re-sequenced twice and re-annotated three times (Roberts & Smits, 2018). It was re-sequenced by using single-molecule real-time and illumina sequencing technology that detected some structural differences compared to the original sequence. This include an additional rRNA gene cluster and additional tRNAs as well as loss of Tn*5397* and pCD630 (Riedel *et al.*, 2015). Re-annotation by Monot *et al.* revealed new coding sequences, corrected the start codon of 222 genes, and redefined the function of 500 genes (Monot *et al.*, 2011). The re-annotation by Pettit *et al.* identified gene products, updated gene nomenclatures, and changed the duplicate gene names by the addition of a number or letter (Pettit *et al.*, 2014). Also the manual re-annotation by Dannheim *et al.* changed the predicted function of 601 gene products, updated 1577 coding sequences, and reduced the number of hypothetical proteins (from 314 to 152)

(Dannheim *et al.*, 2017).

Two erythromycin sensitive derivatives of CD630, CD630Δerm and CD630E, were created by spontaneous deletion of one *erm*(B) gene present in the transposon-like structure Tn5398 (Hussain *et al.*, 2005; O'Connor *et al.*, 2006). CD630Δerm was sequenced and compared to CD630 where other than *erm*(B) deletion and the 18 bp duplication in *spo0A,* 71 differences including indels and single nucleotide polymorphisms (SNPs) in addition to relocation of CTn*5* were detected (van Eijk *et al.*, 2015). Whereas Dannheim *et al.* found three additional SNPs and three chromosomal non-coding inversions in comparison with van Eijk *et al.* (Dannheim *et al.*, 2017). The resequencing and comparison of CD630Δerm and CD630E with CD630 suggest that CD630Δerm is closer to the parental strain and thus it is more favourable for genetic manipulations (Collery *et al.*, 2017).

Other than the CDSs, the *C. difficile* genome consists of a large number of mobile genetic elements (MGEs) (Mullany *et al.*, 2015). In CD630, 11% of the genome is MGE composed of: 7 conjugative transposons (CTns), an element that is probably a mobilizable transposon (MTn) Tn*5398*, two prophages, a skin element, and IStrons (Sebaihia *et al.*, 2006).

1.8 Molecular manipulation of *C. difficile* genome

Our understanding of the biological function and pathogenicity of *C. difficile* is not yet as advanced as those for model organisms such as *E. coli* and *B. subtilis*. One of the limitations is that heterologous DNA could only be transferred to *C. difficile* by conjugation from *Bacillus subtilis* or *E. coli* donors

(Mullany *et al.*, 1994; Purdy *et al.*, 2002). Based on conjugation, cloning of genes in *C. difficile* using conjugative transposons or replicative plasmids was achievable (Haraldsen & Sonenshein, 2003; Herbert *et al.*, 2003; Mani *et al.*, 2002; Roberts *et al.*, 2003).

Directed gene inactivation in *C. difficile* was possible by the use of conjugative transposons. However *Tn916* and *Tn5397* were of limited use for random insertional mutagenesis because they insert at multiple sites or at a specific site within the genome respectively (Hussain *et al.*, 2005; Mullany *et al.*, 2012; Wang *et al.*, 2006). On the contrary, the mariner transposons and its derivatives have the ability to insert randomly into the genome of *C. difficile* (Cartman & Minton, 2010; Dembek *et al.*, 2015).

Most of the engineering tools for making targeted mutations were developed during the past decade and were based on single or double cross-over homologous recombination or by group II intron insertion. Mutants constructed by homologous recombination-mediated single cross-over event are inherently unstable as they can revert back to wild type. Also the unstable plasmid tends to 'loop out' of the insertion site in some of the cells in the same single cross-over mutant which results in a mixed population and might affect the phenotype tested (Dineen *et al.*, 2007; O'Connor *et al.*, 2006). Double cross-over event is more stable but rare and mutants are difficult to isolate and distinguish from single cross-over transconjugants. The use of allele exchange method based on counter-selection markers like cytosine deaminase (*codA*) and pyrimidine orotate phosphoribosyltransferase (*pyrE*)

encoding genes to create mutants is difficult to achieve (Cartman *et al.*, 2012; Heap *et al.*, 2012; Ng *et al.*, 2013). The protocol is time consuming as it involves two steps, a single cross-over event using selection marker then a second recombination utilising counter-selection marker. Single cross-over integrants should be re-streaked two to four times to distinguish them from transconjugate clones. Mutants constructed by *pyrE* double cross-over have to be screened by PCR to distinguish them from wild type reverent clones and grow very poorly which render them useful for phenotypic testing. Moreover this method needs additional steps of constructing a *pyrE*$^{-}$ strain and after constructing the desired mutant, converting it back to *pyrE*$^{+}$ before testing the phenotype (Heap *et al.*, 2012; Ng *et al.*, 2013).

The ClosTron gene knockout system, based on the Targetron, was adapted for the mutagenesis of *Clostridium sp.* by the insertion of mobile group II intron and the selection through a retrotransposable activated marker (RAM) (Heap *et al.*, 2007; Zhong *et al.*, 2003). The intron can be directed to insert into any sequence by nucleotide pairing between the intron RNA and the target site DNA (Heap *et al.*, 2010). ClosTron mutagenesis suffers from several limitations including the polar effect of the intron on the genes downstream of the insertion site (Kuehne & Minton, 2012). The mutation achieved by this method is not a complete deletion but rather an interruption of the gene which might still have a residual function and hence give variation in phenotype (Steiner *et al.*, 2011; Wang *et al.*, 2013). Also the mutant created by the ClosTron system might revert back as a result of the instability of the insertion (Frazier *et al.*, 2003; Wang *et al.*, 2016b).

The molecular genetic field expanded rapidly during the past few years with the adaptation of CRISPR-Cas9 system for the genetic manipulation of many bacteria (Doudna & Charpentier, 2014; Jiang *et al.*, 2013). This genetic tool is discussed in details in Chapter 6.

1.9 Treatment of *C. difficile*

CDI is generally controlled by antimicrobial therapy following the guidelines set by the Infectious Diseases Society of America (IDSA) and the Society for Health Care Epidemiology of America (SHEA). Vancomycin or fidaxomicin are recommended over metronidazole for the first incidence of CDI for severe or non-severe cases. For complicated CDI, vancomycin is the regimen of choice (McDonald *et al.*, 2018).

Treatment of the first recurrent CDI is with oral vancomycin if metronidazole was used for the initial episode, or use prolonged tapered and pulsed vancomycin regimen or fidaxomicin if a standard regimen was used for the initial CDI. In case of a second recurrent CDI, tapered or pulse regimen of vancomycin is recommended, or vancomycin followed by rifaximin, or fidaxomicin. In patients with multiple recurrences of CDI, faecal microbiota transplantation is recommended (McDonald *et al.*, 2018).

Due to the inefficiency of the current treatment regiments for CDI, especially recurrent cases, and the increased incidence of antimicrobial resistance, the search for alternative therapies is a priority.

Antibiotics with less harmful effect on the intestinal microbiota and better efficacy are difficult to manufacture. New antibiotics for the treatment of CDI are in different phases of clinical trial and have not been approved by the

food and drug administration (FDA). Two drugs, cadazolid (Actelion) and surotomycin (Merck), have completed phase III clinical studies (Actelion Ltd., 2017; Daley *et al.*, 2017). Two types of antibiotics, LFF571 (Novartis) and Ridinilazole (Summit Pharmaceuticals) have completed phase II studies. And CRS3123 (National Institute of Allergy and Infectious Diseases) is being evaluated in phase I studies. Reviewed in (Feher *et al.*, 2017; Peng *et al.*, 2018). Ramizol is under development and has been clinically tested in hamsters and compared to current antibiotic regimens (Rao *et al.*, 2016; Wolfe *et al.*, 2018).

Faecal microbiota transplantation (FMT) which involves giving faecal matter from a healthy donor to restore the intestinal microbiota in CDI patients is now being recommended as a treatment for multiple recurrence of CDI (McDonald *et al.*, 2018). This treatment option was reported to resolve more than 90% of recurrent CDI cases and is currently used in some medical centres in the US (Austin *et al.*, 2014). Additionally it can restore the intestinal microbiota to almost normal levels compared to a healthy donor (Song *et al.*, 2013). Although this treatment showed great results, patients still hesitate to use it (Wang *et al.*, 2016a). This motivated pharmaceutical companies to pack the microbiota in an oral capsule either frozen or freeze-dried which showed comparable results to fresh FMT in clinical trials (Dickson, 2018; Hecker *et al.*, 2016; Kao *et al.*, 2017).

Another alternative approach to treat CDI is monoclonal antibodies. In a study conducted in the US and Canada, the incidence of recurrent CDI was

decreased by 18% in patients who received fully human monoclonal antibodies CDA1 and CDB1 against *C. difficile* toxins A and B respectively during the antimicrobial treatment course (Lowy *et al.*, 2010). Another study found that two monoclonal antibodies MK3415 and MK6072 have the ability to neutralize *C. difficile* toxin A and B mediated proinflammatory cytokine expression and colonic tissue damage (Koon *et al.*, 2013). A recent study suggested the use of hyperimmune bovine colostrum after its success in treating piglets from CDI (Sponseller *et al.*, 2015). Actoxumab (MK-3415/GS-CDA1/CDA1) and bezlotoxumab (MK-6072/MDX-1388/CDB1) can bind and neutralize *C. difficile* toxins A and B, respectively (Wilcox *et al.*, 2017). Bezlotoxumab is FDA-approved agent that has proved effective in reducing the rate of first recurrence of CDI in clinical trial while actoxumab did not affect outcomes (Bartlett, 2017; Wilcox *et al.*, 2017).

Phage therapy was proposed shortly after bacteriophages were discovered in the early 1900s (D'Herelle, 1917; Summers, 1999). But the availability of antibiotics at that time made the research into using bacteriophages as therapeutic agents subside (Sulakvelidze *et al.*, 2001). The interest in using phage therapy rose again for many reasons, the emergence of highly virulent strains of *C. difficile*, the increased levels of antimicrobial resistance, and the lack of more efficient therapies (Burrowes *et al.*, 2011; Hargreaves & Clokie, 2014). The main advantage of using phage therapy is that phages are highly specific towards their targets without causing any harmful effect on their environment. Therefore, *C. difficile* would be attacked without affecting any of the microbiota in the human intestine. This specificity is also a disadvantage

because the *C. difficile* strain causing the infection should be known in order to use the specific phage to target it (Lin *et al.*, 2017). Another disadvantage is that all *C. difficile* phages available are lysogenic, which means that under certain circumstances they lysogenize the host instead of killing it. In that case lysogenic phages do not serve the purpose of treating CDI unless this method is accompanied by subsequent antibiotic therapy (Meader *et al.*, 2010; Meader *et al.*, 2013). The use of a cocktail of 7 phages was able to completely lyse *C. difficile in vitro* and reduce colonization in the hamster model (Nale *et al.*, 2015).

1.10 Haemolysins in Gram positive bacteria

1.10.1 Introduction to haemolysins

Haemolysins are proteins, and sometimes lipids, responsible for the lysis (rupture) of erythrocytes or red blood cells (RBCs) and the release of haemoglobin (Todar, 2012). It was as early as 1895 when β-haemolysis by *Streptococcus spp.* on blood agar plates was first described (Marmorek, 1895). And it took scientists more than four decades to identify the causative haemolysins, Streptolysin O and S (Todd, 1932; Weld, 1935).

In the laboratory, haemolysis is used as means of bacterial classification, mainly in *Streptococcus spp.,* according to its two different forms: alpha (α-) and beta (β-) while non-haemolytic is known as gamma (γ-) (Figure 1-11). Alpha-haemolysis, also called partial or incomplete, appears as greenish discolouration around bacterial colonies (Washington, 2012). This is due to the hydrogen peroxide produced by the bacterium oxidizing haemoglobin into methaemoglobin (Morgan & Neill, 1924). Beta-haemolysis, also known as complete, appears as clear areas around and under bacterial colonies caused by complete lysis of RBCs (Kumar, 2016).

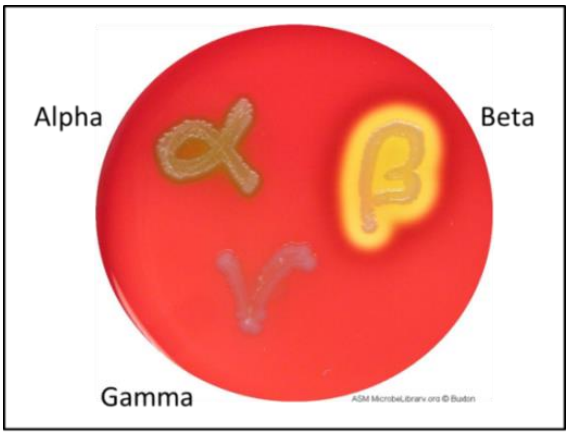


Figure 1-11 Types of bacterial haemolysis.

Alpha (α-) is partial haemolysis characterised by greenish discolouration. Beta (β-) is complete haemolysis characterised by clear areas under and around bacterial growth. Gamma (γ-) represents no haemolysis. Figure is taken from (Buxton, 2005).

1.10.2 Target cells

In the literature some cytolysins and toxins were initially described as haemolysins because their activities were first reported as haemolysis (Alouf, 2005). Although some bacterial toxins and cytolysins have multiple activities, alpha toxin for example, produced by *C. perfringens* possesses a lethal, necrotic, cytolytic, and haemolytic activity (McDonel, 1980). The lethal effect of haemolysins is not limited to erythrocytes but also includes various eukaryotic cells like epithelial cells, leukocytes, lymphocytes, and macrophages (Alouf, 2003; Manich *et al.*, 2008; Palmer, 2001). Some haemolysins are capable of causing haemolysis *in vitro* and *in vivo* while others can only show haemolysis *in vitro*. In a study to measure the *in vivo* effect of *C. tetani* tetanolysin on lab animals, the haemolysin caused death to the mice and intravascular haemolysis in rabbits and monkeys demonstrated by increased plasma haemoglobin levels (Hardegree *et al.*, 1971).

Another aspect that should be considered is that the sensitivity of erythrocytes to haemolysins is different from one species to another. Each haemolysin binds to a specific component (like cholesterol) in the membrane of target cells and hence if this component is more abundant in one erythrocyte species, then it would be more susceptible (Alouf, 2003). For example, erythrocytes of rabbit, guinea pig, and human are more susceptible to botulinolysin, which is a haemolysin produced by *C. botulinum.* Whereas sheep, mouse, rat, and chicken erythrocytes are much less susceptible (Haque *et al.*, 1992).

1.10.3 Mode of action

Generally, haemolysins act on the cell membrane by two mechanisms: pore formation and enzymatic activity reviewed in (Alouf *et al.*, 2015). Delta-toxin, produced by *S. aureus*, is a detergent-like toxin. It acts as a surfactant to disrupt the cell membrane leading to haemolysis (Dinges *et al.*, 2000).

Pore formation involves the insertion of amphipathic β-strands inside the cell membrane organised in a β-barrel cylinder (Figure 1-12). The channels formed by these β-barrel structures induce alterations in the membrane permeability leading to disruption of the integrity of the cell membranes (Benz, 2015). Almost one third of the clostridial haemolysins are pore forming cytolysins (Popoff & Bouvet, 2009). These can be divided according to the structure and mode of action into families, including the cholesterol dependent cytolysins (CDC) which require cholesterol in the membranes and form extraordinarily large pores (25-30 nm in diameter) (Farrand *et al.*, 2015). The CDC family consists of 21 toxins produced by 23 bacterial species of the

genera *Streptococcus, Staphylococcus, Bacillus, Clostridium, Listeria, Brevibacillus, Paenibacillus* and *Arcanobacterium* (Table 1-1) (Tweten, 2005). They share 40-70% primary structure homology and the crystal structures of six of them show high similarity in the tertiary structure (Wade *et al.*, 2015). They are extracellular haemolysins except pneumolysin that is only released after bacterial lysis (Alouf, 2003). Other pore forming toxins include *C. perfringens* ε-toxin and *C. septicum* α-toxin which are structurally related to aerolysins (Popoff & Bouvet, 2009).

Figure 1-12 Schematic representation of the pore formation process.

The soluble pore-forming toxins (PFT) attach to specific receptors in the membrane of target cells. Then oligomerization occurs by one of two mechanisms: surface pre-pore formation usually by β-PFT (referred to by number 1 in the figure) or partial pore formation by α-PFT inserted into the membrane (referred to by number 2 in the figure). Both mechanisms end by the formation of a transmembrane pore with different architecture, size, and shape. Figure is taken with permission from (Dal Peraro & van der Goot, 2016).

Bacterial sphingomyelinases (SMases) and phospholipases (PLases) are a group of lipolytic esterases surface-associated or secreted by a number of Gram positive bacteria (Table 1-1) (Flores-Díaz *et al.*, 2015). *C. perfringens*

α-toxin, *B. cereus* cerolysin AB, and *Listeria monocytogenes* can produce SMase C and PLase C which hydrolyse both sphingomyelin (SM) and phosphatidylcholine (PC) respectively (Gilmore *et al.*, 1989; Goldfine *et al.*, 1993; Urbina *et al.*, 2009). SM and PC are the most abundant components in the cell membrane of eukaryotic cells and RBCs (Carquin *et al.*, 2014). SMase C and PLase C bind to and hydrolyse SM and PC respectively leading to loss of the cell integrity and lysis (Figure 1-13) (Goni *et al.*, 2012). Some haemolysins have one enzymatic activity, like the *S. aureus* β-toxin which functions as SMase only (Huseby *et al.*, 2007). Other enzymes preferentially cleave other phospholipids like *C. novyi* α-toxin that hydrolyses phosphatidylglycerol (PG), phosphatidylethanolamine (PE), and phosphatidylinositol (PI) (Jepson & Titball, 2000).

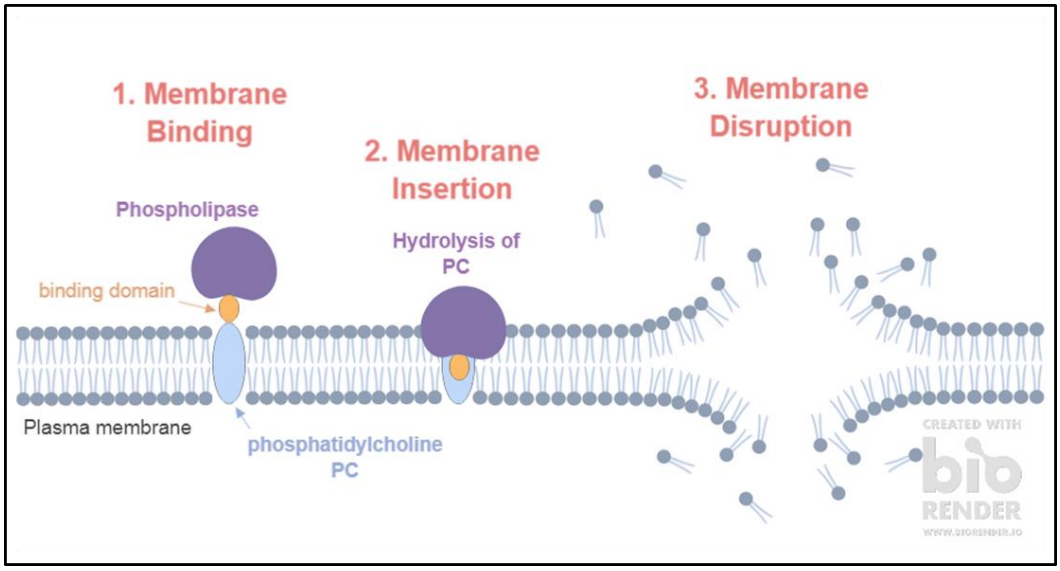


Figure 1-13 Schematic representation of the mechanism of α-toxin produced by *C. perfringens*.

The toxin binds to the phosphatidylcholine in the target membrane then cleaves it leading to disruption of the integrity of the membrane and lysis. Figure was created using BioRender.

1.10.4 Role in pathogenicity

One theory for the lytic activity of the bacterial haemolysins is the acquisition of nutrients such as iron, a very important compound for bacterial growth and virulence (Ratledge & Dover, 2000). Pore forming haemolysins are produced by many bacteria as virulence factors (Benz, 2015). Their role in causing disease may vary, as do the bacteria secreting them (Palmer, 2001). They induce various cellular effects but the exact contribution to disease progression is not fully understood (Eiffler *et al.*, 2016; Hotze & Tweten, 2012). CDC toxins can induce membrane-mediated toxicity, immune response, and regulate cell signalling and function (Cassidy & O'Riordan, 2013; Palmer, 2001). Recently, pneumolysin was found to cause DNA damage in alveolar epithelial cells by inducing DNA double stranded breaks (Rai *et al.*, 2016). SMases and PLases contribute to phagosomal escape, tissue colonization, infection and its progression, or immune response evasion (Flores-Diaz *et al.*, 2016). Mutants of bacteria such as *S. aureus, L. monocytogenes, C. perfringens,* and *C. septicum* with inactive haemolysin genes were less virulent compared to wild type (Awad *et al.*, 1995; Cossart *et al.*, 1989; Kennedy *et al.*, 2005; O'Reilly *et al.*, 1986).

1.10.5 Regulation of haemolysin genes

The genetic regulation of haemolysin expression is different from one bacterium to the other. Some are under positive control of a regulatory two-component system that also regulates the expression of other virulence genes. Perfringolysin O (encoded by *pfoA)* for example is under the control of the VirR/VirS system which also regulates the production of α-toxin (*plc*

gene), collagenase, and hemagglutinin (Ba-Thein *et al.*, 1996; Shimizu *et al.*, 1994). The VirR/VirS system is composed of two genes *virR*, which encodes a response regulator, and *virS*, encodes a sensory histidine kinase. Once VirS is activated, by environmental or other stimuli, it in turn activates the VirR by transferring a phosphate group. Thereafter, the VirR upregulates the expression of toxin genes (Ohtani & Shimizu, 2016). The mode of regulation under the control of VirR/VirS is different in perfringolysin O and α-toxin. Only the major promoter of *pfoA* gene is regulated by VirR/VirS and not the minor promoter whereas *plc* gene has only one promoter and is partially under the control of this regulatory system (Ba-Thein *et al.*, 1996). CodY which regulates the expression of many virulence genes in *C. perfringens* does not affect the production of perfringolysin O or α-toxin (Li *et al.*, 2013). The genes encoding perfringolysin O and other CDC toxins that are cloned and sequenced to date are all chromosomal (Wade *et al.*, 2015).

Table 1-1 Haemolysins of the most studied Gram positive bacteria.

Bacteria	Pore Forming haemolysin	Enzymatic haemolysin	Reference
Clostridium			
C. tetani	Tetanolysin	-	All Clostridial haemolysins are reviewed
C. botulinum	Botulinolysin	-	in
C. perfringens	Perfringolysin O, Δ- and ε-Toxin	α-Toxin	(Hatheway, 1990; Popoff & Bouvet, 2009)
C. sordellii	Sordellilysin	α-Toxin	
C. bifermentans	Bifermentolysin	α-Toxin	
C. novyi	Novyilysin	α-, β-, & Y-Toxins	
C. haemolyticum	-	β-Toxin	
C. chauvoei	Chauveolysin	α-Toxin	
C. septicum	Septicolysin O and α-Toxin	-	
C. histolyticum	Histicolysin O	ε-Toxin	

Bacteria	Pore Forming haemolysin	Enzymatic haemolysin	Reference
Bacillus			
B. cereus	Cerolysin O, haemolysin II, haemolysin III, and Cytotoxin K	Cerolysin AB	*B. cereus* haemolysins are reviewed in (Ramarao & Sanchis, 2013)
B. anthracis	Anthrolysin *[1]	-	*[1](Shannon *et al.*, 2003)
B. thuringiensis	Thuringiolysin O *[2]	-	*[2](Honda *et al.*, 1991)
B. alvei	Alveolysin *[3]	-	*[3](Geoffroy *et al.*, 1990)
Staphylococcus			
S. aureus	α-Toxin, Y-Toxin, and Leukocidin	β-Toxin	All haemolysins are reviewed in (Dinges *et al.*, 2000; Wiseman, 1975)
Streptococcus			
S. pyogenes	Streptolysin O *[4]	Streptolysin S*[6]	*[4](Duncan, 1974; Walev *et al.*, 1995)
S. pneumoniae	Pneumolysin *[5]	-	*[5](Lawrence *et al.*, 2015)
			*[6](Duncan & Mason, 1976)

Bacteria	Pore Forming haemolysin	Enzymatic haemolysin	Reference
Listeria *L. monocytogenes*	Listeriolycin O *[7]	Lecithinase *[8] Listeriolycin S *[9]	*[7](Jenkins *et al.*, 1964; Koster *et al.*, 2014) *[8](Huang *et al.*, 2016b) *[9](Cotter *et al.*, 2008)

α-, alpha; β-, beta; Y-, gamma; Δ-, delta; ε-, epsilon.

1.11 Aims of the study

Despite the fact that most *Clostridium* spp. including *C. sordellii and C. bifermentans,* which were moved with *C. difficile* to the *Peptostreptococcaceae* family, produce haemolysins, *C. difficile* is thought of as a non-haemolytic bacterium when grown on blood agar. This could be due to the lack of studies investigating possible *C. difficile* haemolysin production.

The aims of this work were to investigate haemolysis of *C. difficile* in the presence of glucose. Also to identify haemolysin genes from the genome of CD630, by creating a genomic library in *E. coli*. This work also aimed at using genetic manipulation tools and to specifically utilise the CRISPR-Cas9 system to study possible haemolysin genes in CD630.

Chapter 2

Materials and Methods

Chapter 2 General Materials and Methods

2.1 Bacterial strains, plasmids, and growth conditions

Bacterial strains and plasmids used in this work are listed in Table 2-1. *C. difficile* strains were cultured on brain heart infusion (BHI) agar (Oxoid) supplemented with 5% horse blood (E&O Laboratories) and appropriate antibiotic (erythromycin, Erm or tetracycline, Tet, 10 μg/ml) (Sigma) when required. And when desired *C. difficile* selective supplement (125 mg D-cycloserine and 4 mg cefoxitin) (Oxoid), 0.01% L-Cystein (Sigma), and 2% glucose (Sigma) were added. Cultures were incubated anaerobically in Macs-MG-1000 anaerobic cabinet (Don Whitely) (containing 80% nitrogen, 10% hydrogen, and 10% carbon dioxide) at 37°C for 48 hours. *C. difficile* in BHI broth (Oxoid) were grown in tissue culture flasks (Sarstedt) on a rotary shaker at 50 rpm in the anaerobic cabinet for 16-18 hours.

E. coli strains were grown in Luria-Bertani (LB) agar (Life Technologies) supplemented with ampicillin (Amp, 100 μg/ml), chloramphenicol (Cm, 30 μg/ml), or kanamycin (Km, 30 μg/ml) (all from Sigma) and incubated in 37°C incubator or on a rotary shaker at 200 rpm when cultured in LB broth (Sigma) for 16-18 hours.

All bacterial strains were stored in Microbank Bacterial and Fungal Preservation System (Fisher Scientific) at -80°C.

Table 2-1 List of bacterial strains and plasmids used in this work.

Bacterium or plasmid	Description	Reference or origin
***C. difficile* strains**		
CD630	Ribotype 012, TcdA+, TcdB+, and CDT- Tet^R Erm^R strain containing Tn*5397* and Tn*5398*	(Wust & Hardegger, 1983)
CD37	Ribotype 009, Non-toxigenic strain Tet^S Erm^S Rif^R	(Hachler *et al.*, 1987)
CD630Δerm	Erm^S mutant of CD630, Tet^R	(Hussain *et al.*, 2005)
CD630Δmgna-ClosTron	CD630Δerm mutant with a group II intron inserted within CD630_10340 gene	This work
CD630Δatlf-ClosTron	CD630Δerm mutant with a group II intron inserted within CD630_28300 gene	This work
CD630Δatlf-CRISPR-MI	CD630Δerm mutant with deletion of CD630_28300 gene and non-haemolytic phenotype	This work
CD630Δatlf-CRISPR-MII	CD630Δerm mutant with deletion of CD630_28300 gene and exhibiting haemolysis	This work
CD630Δatlf-CRISPR-MI::atlf	CD630Δatlf-CRISPR-MI complimented with pRPF185-atlf plasmid containing *atlf*	This work
CD630Δatlf-CRISPR-MII::atlf	CD630Δatlf-CRISPR-MII complimented with pRPF185-atlf plasmid containing *atlf*	This work
CD630Δatlf-CRISPR-MI::pRPF185	CD630Δatlf-CRISPR-MI mutant conjugated with pRPF185 plasmid used as a control strain in the haemolysis assay	This work

Bacterium or plasmid	Description	Reference or origin
***Bacillus* strains**		
B.subtilis CU2189	Tet^S	(Christie *et al.*, 1987)
***Streptococcus* strains**		
S. mutans UA159 (ATCC 700610)	Tet^R	(Ajdic *et al.*, 2002)
S. mutans ATCC 25175	Tet^R	(Coykendall, 1977)
S. australis FRStet12	Tet^R, Min^S	(Ciric *et al.*, 2011)
***E. coli* strains**		
α-select silver *E. coli*	Chemically competent transformation recipient	Bioline, UK
5-alpha *E. coli*	Chemically competent transformation recipient	NEB, UK
CA434	Chemically competent conjugation recipient	(Purdy *et al.*, 2002)
Ec-CD630-mgna	*E.coli* carrying pUC19 with CD630_10340 gene cloned in it	This work
Ec-pMTL007C-mgna	CA434 carrying pMTL007C-mgna that was designed to target CD630_10340 gene	This work
Ec-pUC-Lib-1.6	*E.coli* carrying pUC19 with 1.6 Kb insert cloned in it	This work
Ec-pUC-Lib-4.3	*E.coli* carrying pUC19 with 4.3 Kb insert cloned in it	This work
Ec-pUC-Lib-3.4	*E.coli* carrying pUC19 with 3.4 Kb insert cloned in it	This work
Ec-pUC-Lib-1.6Δ02921	*E.coli* mutant of Ec-pUC-Lib-1.6 with deletion of CD630_02921	This work
Ec-pUC-Lib-4.3 Δ11300	*E.coli* mutant of Ec-pUC-Lib-4.3 with deletion of CD630_11300	This work

Bacterium or plasmid	Description	Reference or origin
Ec-pUC-Lib-3.4 Δ28300	*E.coli* mutant of Ec-pUC-Lib-3.4 with deletion of CD630_28300	This work
Ec-CD630-02920	*E. coli* non-haemolytic clone carrying gene CD630_02920	This work
Ec-CD630-02921	*E. coli* haemolytic clone carrying gene CD630_02921	This work
Ec-CD630-02930	*E. coli* non-haemolytic clone carrying gene CD630_02930	This work
Ec-CD630-11280	*E. coli* non-haemolytic clone carrying gene CD630_11280	This work
Ec-CD630-11290	*E. coli* haemolytic clone carrying gene CD630_11290	This work
Ec-CD630-11300	*E. coli* haemolytic clone carrying gene CD630_11300	This work
Ec-CD630-11310	*E. coli* haemolytic clone carrying gene CD630_11310	This work
Ec-CD630-28290	*E. coli* non-haemolytic clone carrying gene CD630_28290	This work
Ec-CD630-28300	*E. coli* haemolytic clone carrying gene CD630_28300	This work
Ec-CD630-28310	*E. coli* haemolytic clone carrying gene CD630_28310	This work
Ec-CD630-28300-S1	*E. coli* mutant of Ec-CD630-28300 with substitution in start codon (ATG → GCG)	This work
Ec-CD630-28300-S166	*E. coli* mutant of Ec-CD630-28300 with substitution in ATG at nucleotide 166 (second ATG)	This work
Ec-CD630-28300-S1-S166	*E. coli* mutant of Ec-CD630-28300 with substitution in start codon and second ATG	This work
Ec-CD630-28300Δ58	*E. coli* mutant of Ec-CD630-28300 with deletion of 58 bp around the second ATG	This work
Ec-pET-atlf	α-select silver *E. coli* cells carrying pZK-ATLF for protein purification	This work

Bacterium or plasmid	Description	Reference or origin
BL21 (DE3)	Chemically competent T7 expression host	NEB
Ec-ATLF	BL21 competent *E. coli* cells carrying pZK-ATLF for protein purification	This work
Ec-pET-atlfΔSP	α-select silver *E. coli* cells carrying pZK-ATLFΔSP for protein purification	This work
Ec-ATLFΔSP	BL21 competent *E. coli* cells carrying pZK-ATLFΔSP for protein purification	This work
Ec-pMTL007C-atlf	CA434 carrying pMTL007C-atlf that was designed to target CD630_28300 gene	This work
Ec-CRISPR-Cas9	*E.coli* clone carrying pMTL83151-CRISPR-Cas9 plasmid	This work
Ec-CRISPR-Cas9-atlf	*E.coli* clone carrying pMTL83151-CRISPR-Cas9-atlf plasmid	This work
Ec-CRISPR-Cas9-atlfΔd	*E.coli* clone carrying pMTL83151-CRISPR-Cas9-atlfΔd plasmid	This work
CA434-CRISPR-Cas9-atlfΔd	CA434 carrying pMTL83151-CRISPR-Cas9-atlfΔd plasmid that was designed to delete CD630_28300 gene	This work
Ec-pRPR185	*E.coli* clone carrying pRPF185	This work
Ec-pRPR185-atlf	*E.coli* clone carrying pRPF185-atlf plasmid for complementation of CRISPR mutants	This work
CA434- pRPR185-atlf	CA434 carrying pRPF185-atlf plasmid for complementation of CRISPR mutants	This work
Plasmids		
pUC19	Amp^R high copy number cloning plasmid	NEB
pZK10340	CD630_10340 gene cloned in pUC19 in between *Hind*III and *EcoR*I	This work
pMTL007C-mgna	pMTL007C-E5 (catP) ClosTron group II intron designed to target the active site of CD630_10340 gene	ATUM (formerly DNA 2.0)

Bacterium or plasmid	Description	Reference or origin
pZK-Lib-1.6	Insert 1 (1.6 Kb) from CD630 genomic library cloned in pUC19 at *Hind*III site	This work
pZK-Lib-4.3	Insert 2 (4.3 Kb) from CD630 genomic library cloned in pUC19 at *Hind*III site	This work
pZK-Lib-3.4	Insert 3 (3.4 Kb) from CD630 genomic library cloned in pUC19 at *Hind*III site	This work
pZK-Lib-1.6Δ02921	pZK-Lib-1.6 with deletion of CD630_02921	This work
pZK-Lib-4.3 Δ11300	pZK-Lib-4.3 with deletion of CD630_11300	This work
pZK-Lib-3.4 Δ28300	pZK-Lib-3.4 with deletion of CD630_28300	This work
pZK02920	CD630_02920 gene cloned in pUC19 in between *BamH*I and *EcoR*I	This work
pZK02921	CD630_02921 gene cloned in pUC19 in between *Hind*III and *EcoR*I	This work
pZK02930	CD630_02930 gene cloned in pUC19 in between *BamH*I and *EcoR*I	This work
pZK11280	CD630_11280 gene cloned in pUC19 in between *BamH*I and *EcoR*I	This work
pZK11290	CD630_11290 gene cloned in pUC19 in between *Hind*III and *EcoR*I	This work
pZK11300	CD630_11300 gene cloned in pUC19 in between *Hind*III and *EcoR*I	This work
pZK11310	CD630_11310 gene cloned in pUC19 in between *BamH*I and *EcoR*I	This work
pZK28290	CD630_28290 gene cloned in pUC19 in between *BamH*I and *EcoR*I	This work
pZK28300	CD630_28300 gene cloned in pUC19 in between *Hind*III and *EcoR*I	This work
pZK28310	CD630_28310 gene cloned in pUC19 in between *BamH*I and *EcoR*I	This work
pZK28300-S1	pZK28300 with substitution at start codon of gene CD630_28300	This work
pZK28300-S1-S166	pZK28300-S1 with a second substitution in ATG at 166 nucleotide of gene CD630_28300	This work

Bacterium or plasmid	Description	Reference or origin
pZK28300-S166	pZK28300 with substitution in ATG at position 166 nucleotide of gene CD630_28300	This work
pZK28300Δ58	pZK28300 with deletion of 58 bp around second ATG of gene CD630_28300	This work
pET28a	Kan^R, T7 *lac* promoter, protein expression plasmid	Novagen
pZK-ATLF	CD630_28300 gene cloned in pET28a in between *Nde*I and *Hind*III	This work
pZK-ATLFΔSP	CD630_28300 gene without the signal peptide cloned in pET28a in between *BamH*I and *Xho*I	This work
pMTL007C-atlf	pMTL007C-E5 (catP) ClosTron group II intron designed to target the active site of CD630_28300 gene	ATUM (formerly DNA 2.0)
pMTL83151	pCB102, *catP*, ColE1 + *tra*, MCS	(Heap *et al.*, 2009)
pCas9	cas9 and tracrRNA expression in *E. coli*, Cm^R	AddGene, USA
pRPF185	pMTL960 with P_{tetM} -gusA and two transcriptional terminators: slpA and fdx	(Fagan & Fairweather, 2011)
pMTL83151-CRISPR-Cas9	Derived from pMTL83151 by inserting $P_{xyl/tetO}$ and *cas9* gene from pRPF185 and pCas9 respectively	This work
pMTL83151-CRISPR-Cas9-atlf	Derived from pMTL83151-CRISPR-Cas9 by inserting P_{tetM} and gRNA with 20 bp target sequence to delete CD630_28300 gene	This work
pMTL83151-CRISPR-Cas9-atlfΔd	Derived from pMTL83151-CRISPR-Cas9-atlf by inserting homology regions to delete CD630_28300 gene	This work
pRPF185-atlf	Derived from pRPF185 by inserting *atlf* in between *Sac*I and *BamH*I	This work

*TcdA, toxin A; TcdB, toxin B; CDT, binary toxin; Tet, Tetracycline; Erm, Erythromycin; Rif, Rifampicin; Amp, Ampicillin; Min, minocycline; Kan, kanamycin; pCB102, origin of replication from *C. butyricum* used as replicon for clostridial conjugation; *catP*, chloramphenicol gene; ColE1 + *tra*, origin of replication for *E. coli* transformation; MCS, multiple cloning site; Cm^R, chloramphenicol resistant.

2.2 Molecular biology techniques

2.2.1 Genomic DNA and plasmid DNA extraction

Genomic DNA from *C. difficile* was isolated using Puregene Yeast/Bact kit for Gram positive bacteria (Qiagene, UK). Plasmid DNA was isolated using QIAprep Spin Miniprep kit (Qiagene, UK). All the methods used were following the manufacturer's protocol.

2.2.2 DNA concentration

DNA concentration was measured using NanoDrop ND-1000 spectrophotometer (Labtech UK) where 1 µl of sample was analysed. The concentration is measured as ng/ µl.

2.2.3 Oligonucleotide synthesis

The oligonucleotides (primers) used in this work are listed in Table 2-2. These primers were designed either manually or by using the online tool Primer 3 (http://simgene.com/Primer3), checked for dimer formation by Multiple Primer Analyzer (https://www.thermofisher.com), and synthesized by Sigma Aldrich.

2.2.4 Polymerase Chain Reaction (PCR) amplification

PCR amplification reaction was prepared using 25 µl DNA polymerase mix [MyTaq Red mix (Bioline Reagents Ltd) or Q5® High-Fidelity 2x Master Mix (NEB,UK)], 2.5 µl of 10 µM of each primer, 1 µl of 100 ng/µl DNA, and 19 µl of molecular grade water (Sigma) to make up a total volume of 50 µl. The reaction was carried out in a Biometra T3000 Thermocycler (Biometra) using

this program: initial denaturation at 94°C for 4 minutes, followed by 25 cycles, each consisting of denaturation at 94°C for 30 seconds, annealing at 55-65°C for 30 seconds, and extension at 72°C for 1-3 minutes, and a final extension step of 72°C for 4 minutes ending with a rapid thermal ramp and hold at 4°C until analysis. The temperature of the annealing step is 5°C lower than the lowest melting temperature of the primers used. The extension step time is usually 1 minute for each 1 Kb of the product.

2.2.5 Agarose gel electrophoresis

Gel electrophoresis was used to separate DNA according to molecular mass on 1% (w/v) agarose gel (Bioline) prepared in 1x TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA) (Severn Biotech) with 5 µl/100 ml of Gel Red (10,000x in water) (Cambridge Bioscience) or Ethidium Bromide (Life Technologies). DNA was mixed with 5x loading buffer (Bioline) and processed in parallel to a 1 µl hyperladder 1Kb (Bioline) for 1 hour at 90 voltage. The gel then was visualized by UV illumination using Alpha Imager (Alpha InnoTech) and the image was captured on Alpha View (version 1.0.1.10) (Alpha InnoTech).

2.2.6 DNA purification

DNA, PCR products or plasmid DNA, was purified using QIAquick PCR purification kit (Qiagen) or QIAquick gel extraction kit (Qiagen) following the manufacturer's instructions.

2.2.7 Restriction endonucleases

Restriction endonucleases or enzymes (RE, from NEB UK) used in this study were used in their designated buffers and under the conditions recommended by the manufacturer. When two RE were used together for double digest of the product, the NEB web tool: (https://nebcloner.neb.com/#!/redigest) was used to find the best buffer and conditions for the reaction.

2.2.8 Dephosphorylation

Calf Intestinal Alkaline Phosphatase (CIAP) (NEB) was used to remove the 5' phosphate group from the plasmid DNA after it was digested with single RE to prevent self-ligation. The reaction was performed according to the manufacturer's protocol.

2.2.9 DNA ligation

T4 DNA ligase (NEB) was used to ligate DNA according to the manufacturer's protocol with some modifications. The insert and plasmid DNA added to the ligation mixture was determined using in silico web tool called ligation calculator (http://www.insilico.uni-duesseldorf.de/Lig_Input.html). And the ligation mix was incubated at 16°C for 16 hours.

2.2.10 Preparation of CA434 competent cells

E. coli CA434 competent cells were chemically prepared according to (Sambrook *et al.*, 1989).

2.2.11 Transformation

Transformation was performed in *E. coli* CA434 or α-select silver competent cells (Bioline, UK) according to the manufacturer's instruction. Transformants were selected on LB plates containing appropriate antibiotic depending on the selective marker present in the plasmid.

2.2.12 Blue/White screening for recombinant plasmids

For cloning plasmid (pGEM T-Easy vector or pUC19) that contain LacZ-a fragment, blue/white screening was performed. Transformants were selected on LB plates containing Amp (100 µg/ml), 50 µg/ml X-Gal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside), and 0.1 mM IPTG (isopropyl β-D-thiogalactopyranoside) (Sigma). White colonies, containing inserts, were selected for further investigation.

2.2.13 Conjugation

Conjugation was performed by following the protocol described previously (Heap *et al.*, 2009). Media were supplemented when required with antibiotics at the following concentrations: chloramphenicol (Cm, 15 or 30 µg/ml) and thiamphenicol (Tm, 15 or 30 µg/ml) (both from Sigma). The phosphate buffered saline (PBS) used was from Sigma.

2.3 DNA sequencing and analysis

DNA was sequenced by GeneWiz UK (formerly known as Beckman Coulter Genomics) using Sanger sequencing method. The sequence results were analysed using NCBI: BlastN to identify the homologous nucleotides between

the sequences and BlastX to identify homologous proteins from nucleotide sequences. Bioedit (version 7.2.0) and GeneDoc (v2.7) were utilized for the sequence alignment.

2.4 Cloning of genes in *E. coli*

DNA cloning and manipulation was carried out as described previously (Sambrook *et al.*, 1989). The gene was amplified using extracted genomic DNA (section 2.2.1) as a template with a primer pair in which two restriction enzyme sites were included. The PCR product was visualised on 1% agarose gel (section 2.2.5) to check for the correct band size then purified by PCR purification kit if one band was seen or by gel extraction kit if multiple bands were shown (section 2.2.6). The purified PCR product and plasmid were digested (section 2.2.7) with the same restriction endonucleases incorporated in the primers. After purification by PCR purification kit, both the insert and plasmid were ligated together (section 2.2.9) then transformed into *E. coli* α-select silver competent cells (section 2.2.11). Plasmid DNA was extracted from transformants (section 2.2.1), digested, and separated on 1% agarose gel to visualise the insert. Cloning of the gene was further confirmed by Sanger sequencing (section 2.3).

2.5 Site-directed mutagenesis (SDM)

Site-directed mutagenesis was conducted using Q5 Site-Directed Mutagenesis Kit (NEB, UK). The principle of the method is based on a PCR reaction using custom primers to create substitution, deletion or insertion in

the target plasmid. The primers can be designed more accurately through the NEB online primer design software (NEBaseChanger.neb.com). A substitution mutagenesis is created by changing the nucleotides at the centre of the primers. A deletion can be created by designing primers that flank the region to be deleted. And an insertion is created by adding the nucleotides to the 5´ end of the primers. PCR was performed as described in section 2.2.4 except that the denaturation cycle was at 98°C instead of 94°C. The PCR was followed by treatment with triple enzymes: kinase, ligase, and DpnI (KLD), in order to phosphorylate the 5´ end of the amplicon, re-circularise the amplicon, and remove the DNA template, respectively. The mixture was then transformed into chemically competent α-select silver *E. coli* cells and plated on LB agar supplemented with the appropriate antibiotic marker specific for the plasmid. The correct transformants were verified by RE digest and Sanger sequencing.

Table 2-2 List of oligonucleotides/primers used in this work.

Primer name	Sequence	Reference
mgna-F	5’ -GGCGG**AAGCTT**AAGGGGGAGTAATATGGCTAG -3’	This work
mgna-R	5’ -GGCGC**GAATTC**GTTAACAATTATTTTGTCTC -3’	This work
Clostron-mgna-F	5’ -TGAAGCAGGAGATGGAAAGGT -3’	This work
Clostron-mgna-R	5’ -TTTGCCTCAAACAACCCATG -3’	This work
EBS	5’ - CGAAATTAGAAACTTGCGTTCAGTAAAC -3’	(Heap *et al.*, 2009)
pUC19-M13-F	5’ -CGCCAGGGTTTTCCCAGTCACGAC -3’	This work
pUC19-M13-R	5’ -TCACACAGGAAACAGCTATGAC -3’	This work
ZK1-F	5’ -TGGATATGAGTATGGATGGTCTTTTTTA -3’	This work
ZK1-R	5’ -CACCCAATTGTTTTGGTGAATT -3’	This work
ZK2-F	5’ -TGTTCCAAAGGCGAAAGAAT -3’	This work
ZK2-R	5’ -TGGTGCATTCATTGCAAATC -3’	This work
ZK3-F	5’ -TGTCCATTTTGGGCAAAGTAT -3’	This work
ZK3-R	5’ -TTCTACCAATATTGCCGCATC -3’	This work
ZK4-F	5’ -GGGGCACTATTAATGGAAAGC -3’	This work
ZK4-R	5’ -GTTCCTCTGCTCCCCAATCT -3’	This work
ZK5-F	5’ -TTCAAGTTTACAAAGTCAAAGTTCG -3’	This work
ZK5-R	5’ -GAGTGTGCAATTGCTTTTCTAA -3’	This work
ZK6-F	5’ -CATCATGCCCTTTACCTTTG -3’	This work

Primer name	Sequence	Reference
ZK6-R	5’ -ACACACTTAGTCAAATAGTA -3’	This work
ZK7-F	5’ -TATCCCTTTGGAACTTTTTT -3’	This work
ZK7-R	5’ -AGAAACTCCTGTTAAACAGA -3’	This work
ZK8-F	5’ -TTCAGTTTTTCATATCCACG -3’	This work
ZK8-R	5’ -AGATGGAAAAAATACAATAA -3’	This work
ZK9-F	5’ -ACTAATTACCTTTAATGTAT -3’	This work
ZK9-R	5’ -TAAGAAGATACAAGAAGATG -3’	This work
ZK13-F	5’ -GAATATTGTCTCAATAAATGGATTATCAAAAGG -3’	This work
ZK13-R	5’-CTTATTCCTCCAATAAAGATAATTTTTCTATAAGAAATTTG -3’	This work
ZK14-F	5’ -GAAGAGAATAAAAGTCTTAACAGTCGTTTTAGC -3’	This work
ZK14-R	5’ -GCCAATTCTCCTTCATGTACTCAATAAATTTATG -3’	This work
ZK15-F	5’ -CCCTCCTTTGAGTTAGGCCG -3’	This work
ZK15-R	5’ -GAGATGCTATTATAATGGCATCTCTTTTATATTGC -3’	This work
ZK16-F	5’ -GGCG**GGATCC**AAGGAGGGATTGACATGAATG -3’	This work
ZK16-R	5’ -GCGC**GAATTC**CCATAATCTTATTCCTCCAATAAAG -3’	This work
ZK17-F	5’-GGCG**AAGCTT**GGAGGAATAAGATTATGGAAAATTTG -3’	This work
ZK17-R	5’ -GCGC**GAATTC**CAATATTCATCATTCTTCACCTCTTC -3’	This work
ZK18-F	5’ -GGCG**GGATCC**GGAAGAGGTGAAGAATGATGAATATTG -3’	This work
ZK18-R	5’ -GCGC**GAATTC**CTATGTATTCTCTCACTTTACTACCTCC -3’	This work

Primer name	Sequence	Reference
ZK19-F	5’ -GGCG**GGATCC**AGGAGTGATATAATTTGGAAAAAAAA -3’	This work
ZK19-R	5’ -GCGC**GAATTC**AAATCACCTACTTTGTCTCATACC -3’	This work
ZK20-F	5’ -GGCG**AAGCTT**AGGTGATTTTATGTTGATATTAGGAC -3’	This work
ZK20-R	5’ -GCGC**GAATTC**CTTTCTTACTATTCACGCCAATTC -3’	This work
ZK21-F	5’ -GGCG**AAGCTT**GAAGGAGAATTGGCGTGAATAGTAA -3’	This work
ZK21-R	5’ -GCGC**GAATTC**TCTCTTCATTAGTTTTTTCCTTTCTTG -3’	This work
ZK22-F	5’ -GGCG**GGATCC**AGAAAGGAAAAAACTAATGAAGAG -3’	This work
ZK22-R	5’ -GCGC**GAATTC**TTAATCATTTAATTTGATTCATTCG -3’	This work
ZK23-F	5’ -GGCG**GGATCC**AGGAGGTGTTTGACT TGAAAAAAGG -3’	This work
ZK23-R	5’ -GCGC**GAATTC**ATCAATTACTACTCCTTATACAAGGCTTC -3’	This work
ZK24-F	5’ -GGCG**AAGCTT**AGGAGGGAATAATATGAGACC -3’	This work
ZK24-R	5’ -GCGC**GAATTC**AAAACTCTATTTAGCTAAATTTTGC -3’	This work
ZK25-F	5’ -GGCG**GGATCC**AGGGGTGAGATTTACATGAAGAAAG -3’	This work
ZK25-R	5’ -GCGC**GAATTC**GTTAGGGTCTAATTTGTATTTTTATTTC -3’	This work
ZK26-F	5’ -AGGGAATAATGCGAGACCAAGTAAAAAATTATTAATAG -3’	This work
ZK26-R	5’ -CCTAAGCTTGGCGTAATC -3’	This work
ZK27-F	5’ -AGCTAATGCTGCGGTAAATAGATTAGC -3’	This work
ZK27-R	5’ -TCGTTTTTATCGTAATTTCC -3’	This work
ZK28-F	5’ -CG**CATATG**AGACCAAGTAAAAAATTATTAATAGC -3’	This work

Primer name	Sequence	Reference
ZK28-R	5’ -GC**AAGCTT**TTTAGCTAAATTTTGCAAAAAG -3’	This work
ZK29-F	5’ -GGCG**GGATCC**ATGGTAAATAGATTAGC -3’	This work
ZK29-R	5’ -GCGC**CTCGAG**TTTAGCTAAATTTTGC -3’	This work
Clostron-atlf-F	5’ -GTGGAAAGTTAACAGATGAAAAAG -3’	This work
Clostron-atlf-R	5’ -GAGATGCCATTATAATAGCATCTC -3’	This work
Pxyl/tetO-F	5’- GT**CCCGGG**TTAAGACCCACTTTCAC -3’	This work
Pxyl/tetO-R	5’-CCTATTGAGTATTTCTTATCCATGCTTATTTTAATTATACTCTATC -3’	This work
Cas9-F	5’-GATAGAGTATAATTAAAATAAGCATGGATAAGAAATACTCAATAGG-3’	This work
Cas9-R	5’- CA**GTCGAC**TCAGTCACCTCCTAGCTGAC -3’	This work
gRNA-F	5’ - CA**CCCGGG**CAGCAGACCAAAGCAACG -3’	This work
gRNA-R	5’ - CA**GCGGCCGC**AAAAAAAGCACCGACTCG -3’	This work
ZK32-F	5’ - GA**GGCGCGCC**TGATACTTTGGTAAATCCACCAG -3’	This work
ZK32-R	5’ - GATGCCATTATAATAGCATCTTCCTTTGAGTTAGGCCGGGAG -3’	This work
ZK33-F	5’ - CTCCCGGCCTAACTCAAAGGAAGATGCTATTATAATGGCATC -3’	This work
ZK33-R	5’ - GT**GGCGCGCC**GTGGATATACCTTACACAATATG -3’	This work
ZK34-F	5’ - CACCAAATACAGATGAACCAATAG -3’	This work
ZK34-R	5’ - GGAATTATAGTTATAATAAATCCATTGC -3’	This work
ZK35-F	5’ - GTCCATGGAGATCTCGAGGC -3’	This work

Primer name	Sequence	Reference
ZK35-R	5’ - GGTAGCTTAATATATAAGAGCTGAGGAC -3’	This work
C-atlf-F	5’ - CA**GAGCTC**AGGAGGGAATAATATGAGACC -3’	This work
C-atlf-R	5’ - GC**GGATTC**CTCTATTTAGCTAAATTTTGC -3’	This work
pRPF-fdx-F	5’ - GCATCAAGCTAGCATAAAAATAAG -3’	This work

*The highlighted sequences represent the restriction endonuclease site and the underlined sequences represent a mismatch compared to the original sequence.

Chapter 3

C. difficile is haemolytic in the presence of 2% glucose

Chapter 3 *C. difficile* is haemolytic in the presence of 2% glucose

3.1 Introduction

In 2010, Williams *et al.* discovered that the addition of 2% glucose to blood agar induces *C. difficile* to produce haemolysis (Rachel Williams, Peter Mullany, and Adam P. Roberts, unpublished data). A number of *C. difficile* strains were tested for haemolytic activity including: CD630, CD37, CD196, CD79685, R20291, CD1457, CD1458, CD1481, CD1490, and CD1523. All strains showed haemolysis on 2% glucose-containing agar and no haemolysis in the absence of it except CD1523. Also different concentrations of glucose (0.0%, 0.5%, 1.0%, 1.5%, and 2%) were tested and 2% was found to cause the most haemolysis in the majority of strains.

In 2012, Dasgupta *et al.* created a CD630 genomic library in *E. coli* to identify genes responsible for haemolysis (Protima Dasgupta, Rachel Williams, and Adam P. Roberts, MSc project). The preliminary investigation resulted in four independent haemolytic *E. coli* clones which all contained the same genes: CD630_10330 and CD630_10340. Three of the haemolytic clones showed reduced haemolysis on blood agar compared to the clear haemolysis of the wild type. This observation was investigated and appeared to be because of an *E. coli* insertion sequence (IS) element within CD630_10340 gene.

In this chapter, haemolysis of CD630 with and without 2% glucose was investigated. In addition, gene CD630_10340 was considered as a possible haemolysin and was designated *mgna* (mannosyl-glycoprotein endo-beta-N-acetylglucosamidase). The *mgna* gene was studied to determine whether it encodes a haemolysin in *C. difficile*.

3.2 Materials and methods

3.2.1 Testing haemolysis on blood agar plate

C. difficile strains were cultured on BHI agar (Oxoid, UK) supplemented with 5% defibrinated horse blood (E&O laboratories) and 2% glucose prepared from 50% glucose stock (Sigma). Plates were incubated anaerobically at 37°C between 72 and 96 hours or until haemolysis was visible. *E. coli* clones were grown on LB agar (Life Technologies) supplemented with 100 µg/ml Amp and 5% defibrinated horse blood and incubated aerobically at 37°C between 24 and 48 hours.

3.2.2 pH assay

CD630 was grown in 10 ml BHI broth (Oxoid, UK) in 25 cm^2 tissue culture flask (T-25) with a ventilating cap (SARSTEDT) with and without 2% glucose (Sigma). The flasks were incubated anaerobically at 37°C for 18 hours. The pH of the overnight culture was measured using Jenco 6230N pH meter and the readings were recorded. The average of the pH measurements were calculated for the glucose-containing and normal broth then compared to each other. The difference between the pH of the two growth conditions was statistically measured by unpaired two-tailed t-test using GraphPad Prism v8.1.2. A p-value of 0.05 or below indicates a significant difference between the values tested while a p-value of above 0.05 indicates no significant difference.

3.2.3 Analysis of the four haemolytic *E. coli* clones

The four *E. coli* clones were grown from stock (provided by Protima Dasgupta) on LB agar (Life Technologies) supplemented with 100 µg/ml Amp and 5% defibrinated horse blood. Plates were incubated aerobically at 37°C between 40 and 48 hours. Plasmid DNA from the clones was extracted, digested with *Hind*III, and visualised on a 1% agarose gel. The inserts DNA was analysed by Sanger sequencing and BLAST.

3.2.4 Constructing *E. coli* clone carrying *mgna*

CD630 genomic DNA was used to amplify *mgna* (975 bp) by PCR using mgna-F and mgna-R primer pair that contains *Hind*III and *EcoR*I restriction sites respectively (Figure 3-1). The PCR product was cloned into pUC19 by digesting both with *Hind*III and *EcoR*I and ligating them with T4 DNA ligase. The ligation product was transformed into chemically competent α-select silver *E. coli* cells. Transformants (white colonies) were picked from LB agar supplemented with Amp, X-Gal, and IPTG. Plasmid DNA was extracted, digested with *Hind*III and *EcoR*I, and visualised on a 1% agarose gel. The construct was further confirmed by Sanger sequencing. The detailed protocols for each step of the cloning are in Chapter 2.

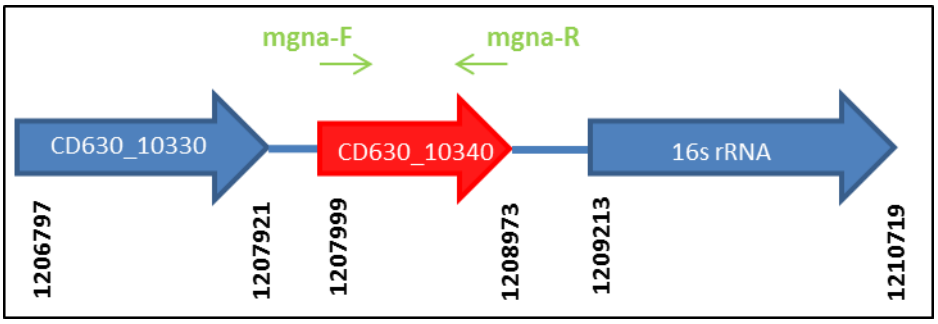


Figure 3-1 Schematic representation of CD630_10340 gene (*mgna*) and primers used to clone it.

The gene (red arrow) encoding a putative mannosyl-glycoprotein endo-beta-N-acetylglucosamidase was cloned using primers mgna-F and mgna-R (small green arrows). The blue arrows represent the gene upstream (encoding UDP-N-acetylglucosamine 2-epimerase protein) and downstream (rRNA-16S ribosomal RNA) of the *mgna*. The numbers under the arrows are the gene ranges in GenBank which represent the position of these genes in CD630 genome.

3.2.5 ClosTron mutagenesis

The Clostron mutagenesis system was used for insertional knock-out of *mgna* (CD630_10340). The plasmid used for the knockout was designed using the ClosTron website (www.clostron.com). The Perutka algorithm available through the intron design tool was used to identify intron insertion target sequences within *mgna*. The intron insertion sequence was selected to have the highest score among the sequences that are at the region that encodes the protein active site. The plasmid (pMTL007C-mgna) was ordered from ATUM (formerly known as DNA 2.0) and recovered from the GFC filter according to the manufacturer's instructions. Plasmid pMTL007C-mgna was transferred via conjugation into CD630Δerm as described previously (Heap *et al.*, 2010; Heap *et al.*, 2009). Transconjugants were selected for by streaking onto BHI agar containing *C. difficile* selective supplement and Tm (15 µg/ml) and then were re-streaked on Erm (10 µg/ml) containing plates.

PCR utilizing Clostron-mgna-F and Clostron-mgna-R primer pair was performed to screen for group II intron insertion. While PCR utilizing Clostron-mgna-F and EBS primer pair was used to confirm that the group II intron was inserted at the correct position within the gene.

3.2.6 Growth curve for CD630 with glucose

The CD630 growth curve was performed with the addition of glucose. A single colony of CD630 was used to inoculate 10 ml BHI broth supplemented with 2% glucose anaerobically in a 25 cm^2 tissue culture flask (T-25) with a ventilating cap (SARSTEDT) for 18 hours. This culture was used to inoculate 15 ml BHI broth supplemented with 2% glucose in a 75 cm^2 tissue culture flask (T-75) with a ventilating cap (SARSTEDT) to a starting optical density OD600 of 0.1. The flask was incubated anaerobically on a 50 rpm shaker. From the flask, 1 ml sample was taken hourly and the absorbance at OD600 was measured for up to 11 hours using WPA CO8000 cell density meter (Biochrom Ltd, UK). The mean of three biological repeats were plotted into a line chart using Microsoft Excel 2010. Then the standard deviation was calculated to add the error bars to the readings.

3.2.7 Contact haemolysis assay

CD630 was tested for haemolytic activity as described previously (Deshpande *et al.*, 1997) (Stachowiak *et al.*, 2012) with the following modifications. CD630 was grown in 15 ml BHI broth supplemented with 2% glucose in T-75 flask anaerobically for 18 hours. This was used for the haemolysis assay at stationary phase. To test the haemolysis at mid-

exponential phase, the overnight culture was used to inoculate 15 ml BHI broth supplemented with 2% glucose to a starting OD600 of 0.1. Then the flask was incubated anaerobically on a 50 rpm shaker for 5 to 6 hours (OD600 of 0.6). Both cultures were harvested at 4500 x *g* for 10 minutes (5000 rpm in Eppendorf centrifuge 5804 R) and re-suspended in 0.5 ml phosphate-buffered saline (PBS, Oxoid UK) supplemented with 2% glucose (Sigma). The concentration of the CD630 suspension was then adjusted to OD600 of 10. Horse RBCs were gently sedimented at 1125 x *g* for 5 minutes at 4°C (2,500 rpm in Eppendorf centrifuge 5804 R) then washed twice with PBS supplemented with 2% glucose and 0.5% bovine serum albumin (BSA, Sigma UK). Inside the anaerobic cabinet, 100 µl of CD630 suspension was mixed with an equal volume of 1% RBCs suspension in 96-well round-bottom microtiter plate (SARSTEDT). In the same assay, PBS with 2% glucose was used as a blank and RBCs suspension was used as negative control. Two bacterial cell suspensions, CD630 and CD37, were included as a negative control without adding RBCs suspension to them. Also 1% lysed RBCs in sterile distilled water was used as a positive control. The plate was sealed with parafilm M (Alpha Laboratories UK) to prevent exposure to oxygen while it was centrifuged at 1200 x *g* (3000 rpm in Fisher Scientific Jouan C412 centrifuge) for 15 minutes. It was then incubated anaerobically at 37°C for 18 hours. After incubation, the plate was centrifuged as mentioned above and 100 µl of the supernatant was collected in another microtiter plate. Haemoglobin released in the supernatant was measured by reading the OD at 540 with CLARIOstar reader (BMG LABTECH UK). Several concentrations of CD630 cell suspension were tested, and the concentration which showed

the highest haemolytic activity was selected for the assay. The defibrinated horse blood was always used fresh from E&O laboratories.

3.3 Results

3.3.1 Haemolysis of *C. difficile* when 2% glucose is added to blood agar

CD630 was non-haemolytic on blood agar while addition of 2% glucose resulted in clear zones of haemolysis (Figure 3-2). It was perceived that the presence of glucose in the agar might cause haemolysis either by a direct effect such as changing the pH of the media which can exert pressure on the membrane of RBCs leading to their rupture. Or that glucose might have an effect on the bacteria itself causing it to produce haemolysins.

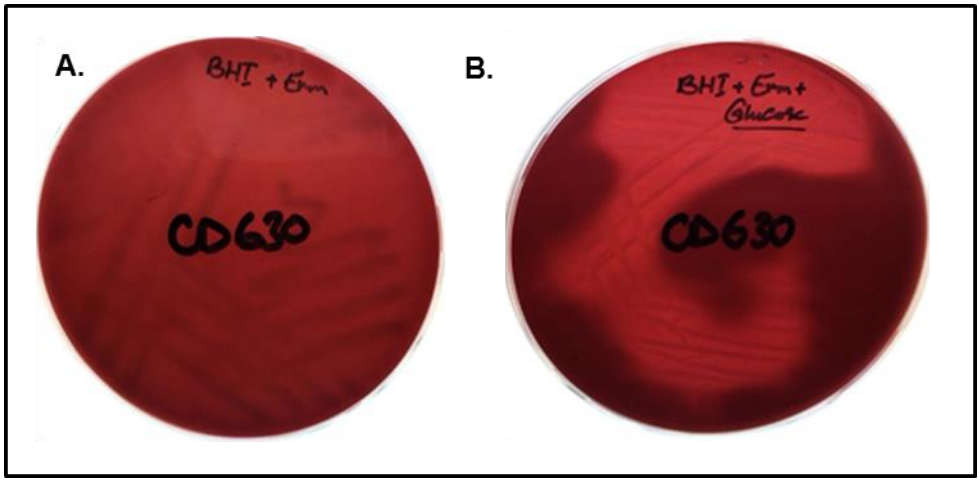


Figure 3-2 CD630 growth with and without 2% glucose.

A. CD630 showed no haemolysis on blood agar without glucose. **B.** CD630 showed clear zone of haemolysis (β-haemolysis compared to **Figure 1-11**) when the agar was supplemented with 2% glucose.

3.3.2 pH change when *C. difficile* was grown in 2% glucose

To determine the pH of the overnight culture, CD630 was grown anaerobically with and without 2% glucose and the pH measured (Table 3-1). The average of the pH of the culture without glucose was 6.08 and with glucose was 5.87. The pH difference was statistically significant between the two culture conditions. But this decrease in pH is not likely to cause the large zones of haemolysis observed as RBCs membrane defect and haemolysis start at a pH of 5 and below which was not the pH measured in both conditions. Therefore the second hypothesis that glucose might induce CD630 to produce a haemolysin is more likely and should be tested.

Table 3-1 pH of CD630 bacterial cultures with and without glucose.

	CD630 in BHI	CD630 in BHI + 2% Glucose
1	6.14	5.80
2	6.13	5.91
3	6.15	5.98
4	6.15	5.98
5	6.03	5.82
6	6.05	5.79
7	6.03	5.75
8	6.04	5.90
9	6.03	5.90
Avg	**6.08**	**5.87**
p-value		**< 0.0001**

*Avg; is the average of the readings.

3.3.3 Sequence analysis of the *E. coli* clones retrieved from a previous CD630 library

The *E. coli* clones retrieved from the CD630 library done by Protima Dasgupta were grown and plasmid DNA was extracted and inserts DNA sequenced and analysed by BLAST. All the four clones contained two genes: CD630_10330 and CD630_10340.

The first gene (GenBank AM180355.1, range 1206797 to 1207921, CD630_10330) was predicted to encode a UDP-N-acetylglucosamine 2-epimerase protein (374 residues). The protein had an epimerase_2 domain (PF02350, residues 24-365) that catalyses the production of UDP-N-acetyl-D-mannosamine (UDP-ManNAc), from UDP-N-acetyl-D-glucosamine (UDP-GlcNAc), which is a capsule precursor (Hinderlich *et al.*, 1997). In Enterobacteriaceae, this enzyme is involved in the enterobacterial common antigen biosynthesis, which is part of bacterial outer membrane biogenesis. Some enzymes in this family are bifunctional as they possess UDP-N-acetylglucosamine 2-epimerase and the N-acetylmannosamine kinase functions (Stasche *et al.*, 1997). The bifunctional enzymes are involved in the biosynthesis of sialic acid to form N-acetylneuraminic acid (NeuAc). Sialic acids are utilised and incorporated in many bacterial surface glycoconjugates in order to evade host immune response (Chen *et al.*, 2016). The N-acetylmannosamine kinase belongs to the family of transferases which are responsible for transferring phosphorous containing groups (ATP and N-acyl-D-mannosamine to ADP and N-acyl-D-mannosamine 6-phosphate). This protein belongs to the family of ROK (IPR000600) (Repressor, ORF, Kinase)

which include transcriptional repressors, uncharacterised ORF, and sugar kinases. In *E. coli* strain k-12, the N-acetylglucosamine repressor (nagC) controls fimB expression which is involved in switching on-and-off type 1 fimbriation (Sohanpal *et al.*, 2007). The second gene (GenBank AM180355.1, range 1207999 to 1208973, CD630_10340) was predicted to encode a protein identified as putative mannosyl-glycoprotein endo-beta-N-acetylglucosamidase (324 residues). It had a glucosaminidase domain (PF01832, residues 173-314) and a transmembrane region (residues 20-42). The domain is distantly related to the eukaryotic lysozymes and is known to hydrolyse peptidoglycan.

Three of the haemolytic clones had an *E. coli* insertion sequence (IS) element within CD630_10340 gene at different positions which explain the reduced haemolysis of these clones on blood agar observed previously.

3.3.4 Cloning of a possible haemolysin gene

To study whether *mgna* was implicated in haemolysis, it was cloned in *E. coli*. Ec-CD630-mgna clone carrying *mgna* (975 bp) was grown on blood agar to detect haemolysis. The plates were further incubated when the clone did not show haemolysis after overnight incubation (16-18 hours). After up to 48 hours incubation, the clone showed very clear zones of haemolysis compared to the non-haemolytic controls: wild type *E. coli* and *E. coli* carrying the empty vector (Figure 3-3). Therefore, *mgna* encodes a haemolysin in *E. coli*.

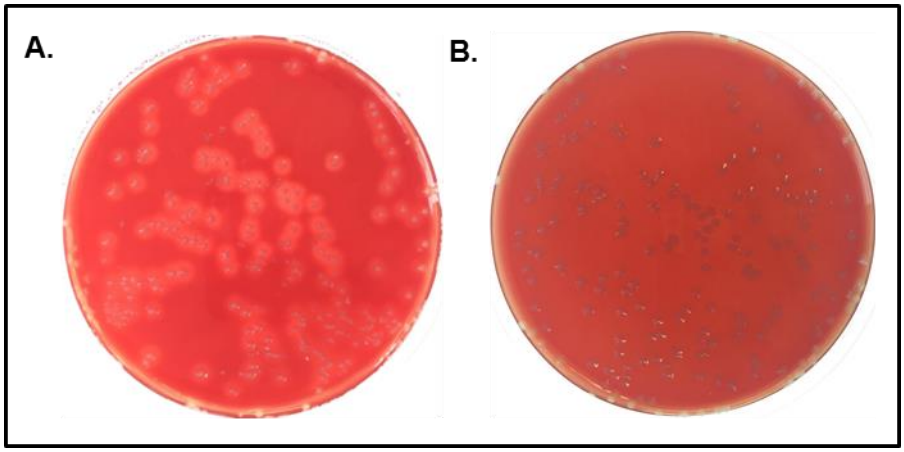


Figure 3-3 Haemolysis of *E. coli* clone carrying *mgna* on blood agar plate.

A. transformation plate of *E.coli* carrying pUC19 with *mgna* cloned from CD630. **B.** transformation plate of *E.coli* carrying the empty plasmid (pUC19) as a control.

3.3.5 Constructing a ClosTron mutant

To determine if *mgna* is responsible for haemolysis in CD630, a knockout mutant was constructed. Plasmid pMTL007C-mgna was designed to target *mgna* in the sense orientation immediately after nucleotide 675 in the sequence encoding the catalytic domain (Figure 3-4). Transconjugants were screened by PCR to confirm the group II insertion location. Genomic DNA from CD630Δerm was also subjected to PCR with the same primer pairs as a control. The PCR product of mutant CD630Δmgna-ClosTron with primers Clostron-mgna-F and Clostron-mgna-R was 2424 bp compared to the 624 bp amplicon from CD630Δerm (1800 bp is the size of group II intron). The PCR product using Clostron-mgna-F and EBS universal (695 bp) was detected for the ClosTron mutant only and not CD630Δerm. This confirms that group II intron had inserted into the correct predicted site within *mgna*. The mutation was further confirmed by sequencing (Figure 3-5).

The CD630Δmgna-ClosTron mutant was cultured on blood BHI plates with and without 2% glucose and compared to the wild type. The mutant showed haemolysis identical to the wild type (Figure 3-6). Another PCR screen for the haemolytic colonies growing on the plate containing 2% glucose confirmed the presence and position of the group II intron.

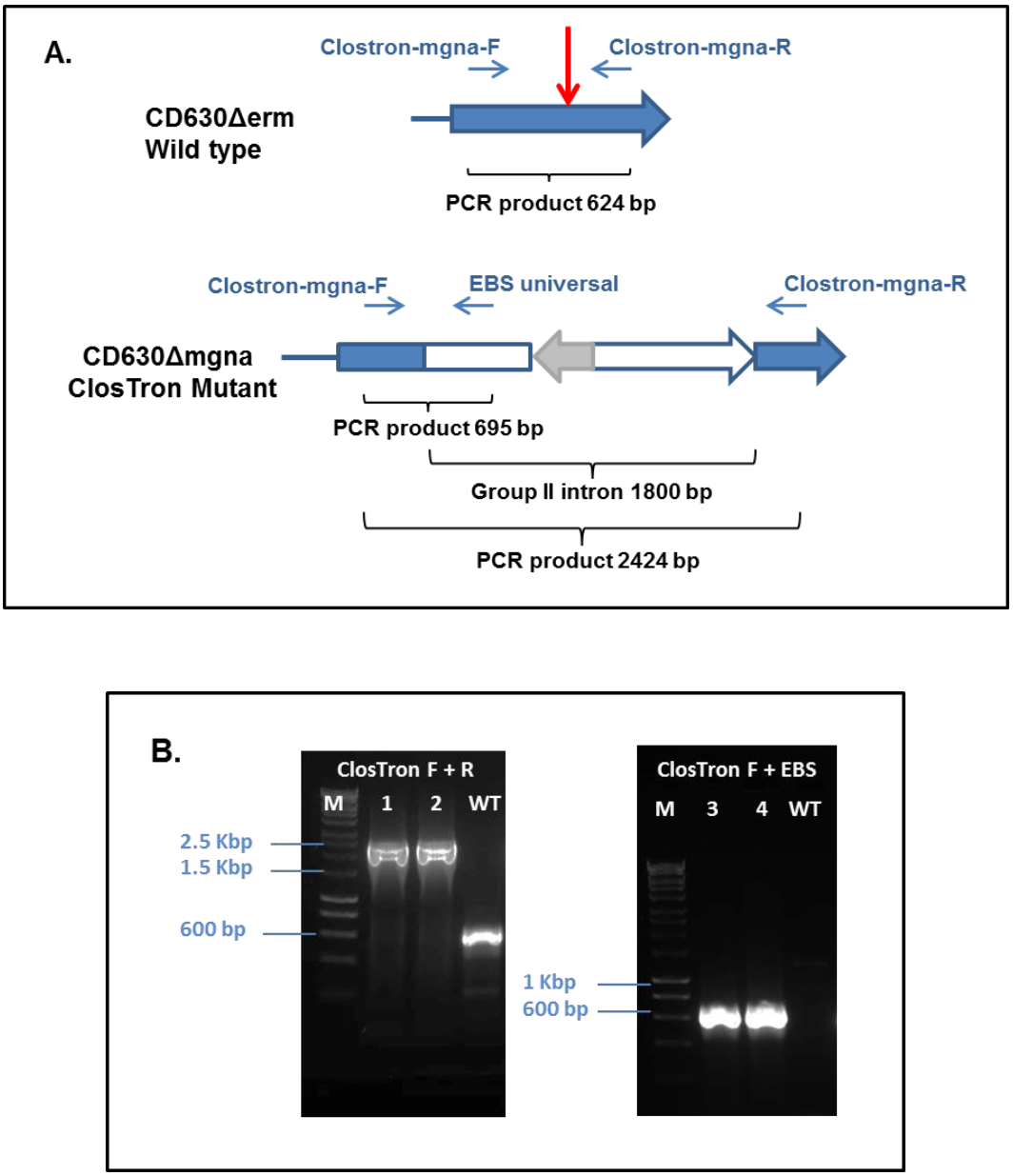


Figure 3-4 PCR screens to confirm deactivation of *mgna* by insertion of group II intron.

A. Schematic representation of CD630_10340 gene (*mgna* in blue coloured arrow) where the group II intron (white arrow encoding a region of *ermB* in grey arrow) was inserted. The small blue arrows represent the primers used to screen for the insertion and the red arrow represents the position of the insertion within the targeted gene. **B.** 1% Agarose gel showing the PCR products amplified with Clostron-mgna-F/Clostron-mgna-R primers should give 2424 bp and 624 bp band sizes in the mutant and the wild type respectively. PCR product utilizing Clostron-mgna-F/EBS should give 695 bp band in the mutant and no product with the wild type. M, 1 Kb DNA ladder; 1&2, CD630Δmgna-ClosTron mutant genomic DNA with Clostron-mgna-F/Clostron-mgna-R primers; WT, wild type (CD630Δerm) genomic DNA with Clostron-mgna-F/Clostron-mgna-R primers; 3&4, CD630Δmgna-ClosTron mutant genomic DNA with Clostron-mgna-F/EBS primers; WT, wild type (CD630Δerm) genomic DNA with Clostron-mgna-F/EBS primers.

Figure 3-5 DNA sequence of the *mgna* in CD630Δmgna-ClosTron mutant.

The full sequence of *mgna* is presented with the exact insertion site of the group II intron. The sequence highlighted in grey is the group II intron and within the EBS primer is highlighted in yellow. Other primers designed to retarget the ClosTron insertion are in blue. The target sequence within *mgna* chosen for group II intron insertion is in red.

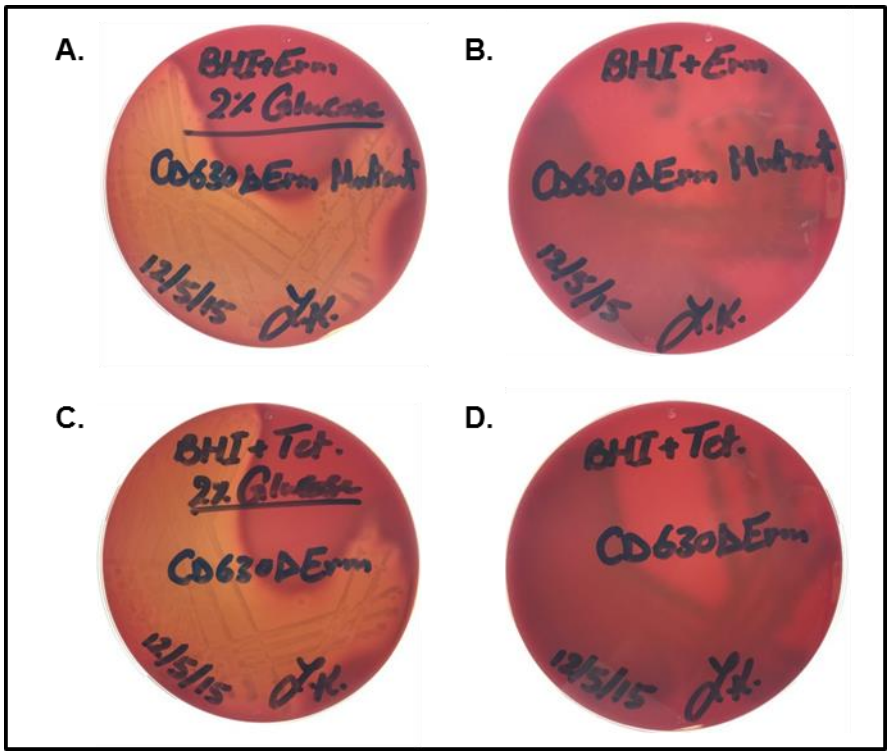


Figure 3-6 Testing haemolysis of the CD630Δmgna-ClosTron mutant.

CD630Δmgna-ClosTron mutant grown on BHI agar plates with (A.) and without 2% glucose (B.). CD630Δerm grown as control with (C.) and without 2% glucose (D.). The mutant showed haemolysis similar to wild type.

3.3.6 Growth curve for CD630 with and without glucose

Two growth curves were performed to determine the growth phases of CD630 with and without 2% glucose for the purpose of developing a quantitative assay to measure haemolysis. The OD600 measurements were plotted into a line chart (Figure 3-7). There was a statistical difference between the two growth curves starting at the exponential phase ($p < 0.001$ and $p < 0.0001$ represent at the smallest and largest difference between the curves respectively). The growth curve indicated that the mid-exponential phase for CD630 in the presence of 2% glucose was at OD600 of 0.6 which is between 5 to 6 hours.

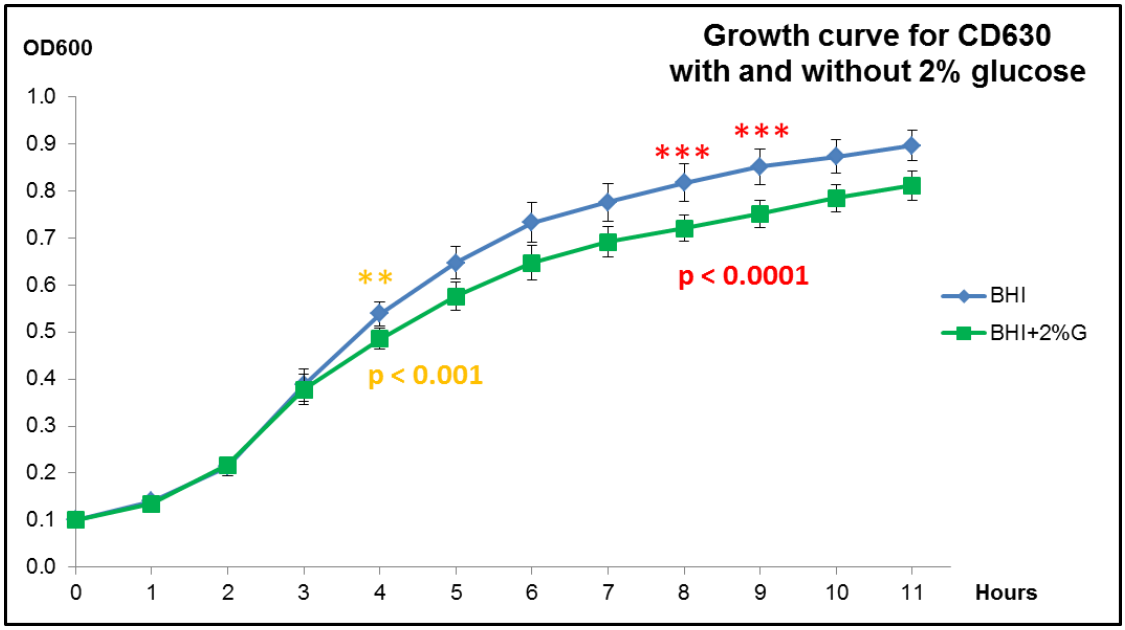


Figure 3-7 Growth curve for CD630 with and without 2% glucose.

This chart represents the growth of CD630 in BHI with 2% glucose (green) and without glucose (blue) over 11 hours. The mean of the readings were calculated and plotted as markers in the chart. The standard deviation was also calculated and presented as error bars. Statistical significance is indicated by three asterisks ***, $p < 0.0001$ and two asterisks **, $p < 0.001$ (unpaired two-tailed t-test measured by GraphPad Prism v8.1.2).

3.3.7 Contact haemolysis assay for CD630

A contact haemolysis assay was developed for the quantitative measurement of haemolysis caused by the addition of 2% glucose to CD630 while growing in BHI. The principle of the assay was based on measuring the haemolysis of the RBCs after they were incubated in direct contact with the bacterial cells. To know the best concentration of bacterial suspension to use for the assay, two-fold serial dilution of the bacterial cells were tested with RBCs suspension. CD630 cell suspensions at stationary and mid-exponential phases were used. BSA was included in the assay to maintain the shape and integrity of RBCs. RBCs suspension was incubated in the same conditions without *C. difficile* to demonstrate that they remain intact during the assay (negative control). Also RBCs suspended in water were included to demonstrate maximum haemolysis (positive control). Two *C. difficile* bacterial suspensions without the addition of RBCs were used to highlight that the OD540 measured was related to haemolysis. After incubation, the supernatant was measured at OD540 which correspond to the haemoglobin released by the lysis of RBCs. The OD540 measurements for CD630 were plotted into a bar chart (Figure 3-8). The highest level of haemolysis was caused by the neat suspension of CD630 which corresponds to OD600 of 10.

The OD540 measurements using neat CD630 and controls were plotted into a bar chart (Figure 3-9). The difference between the haemolysis caused by CD630 and the negative control was statistically significant with a p-value of 0.05 which indicated that CD630 grown in mid-exponential (MEP) or stationary (SP) phases can haemolyse RBCs.

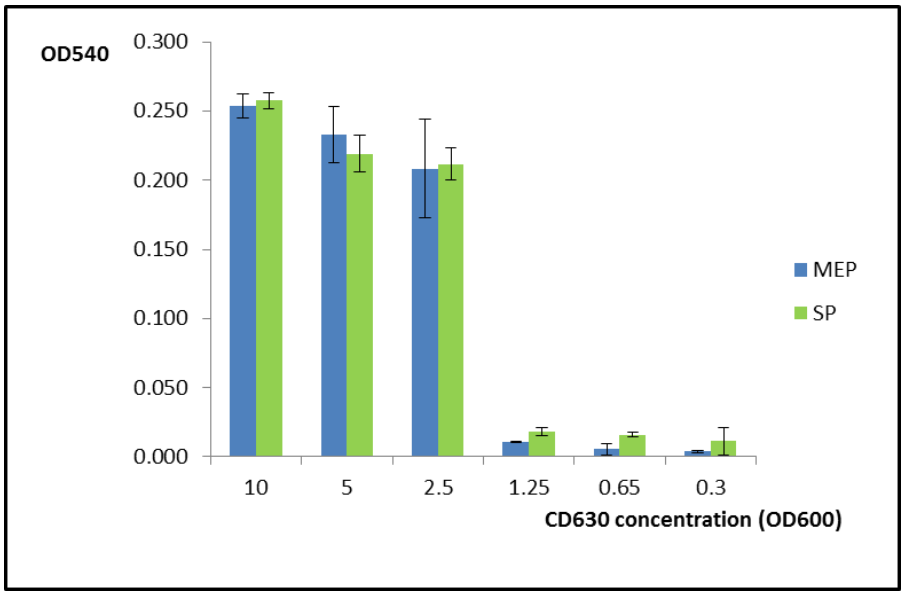


Figure 3-8 Testing the best concentration of CD630 overnight culture for the haemolysis assay.

The bars represent the OD540 of the supernatants from incubating two-fold serial dilution of CD630 with 1% RBCs suspension. The two growth phases used for the assay are mid-exponential (MEP) in blue bars and stationary phase (SP) in green. In both MEP and SP, the highest level of haemolysis was reached when the OD600 of the CD630 suspension was 10.

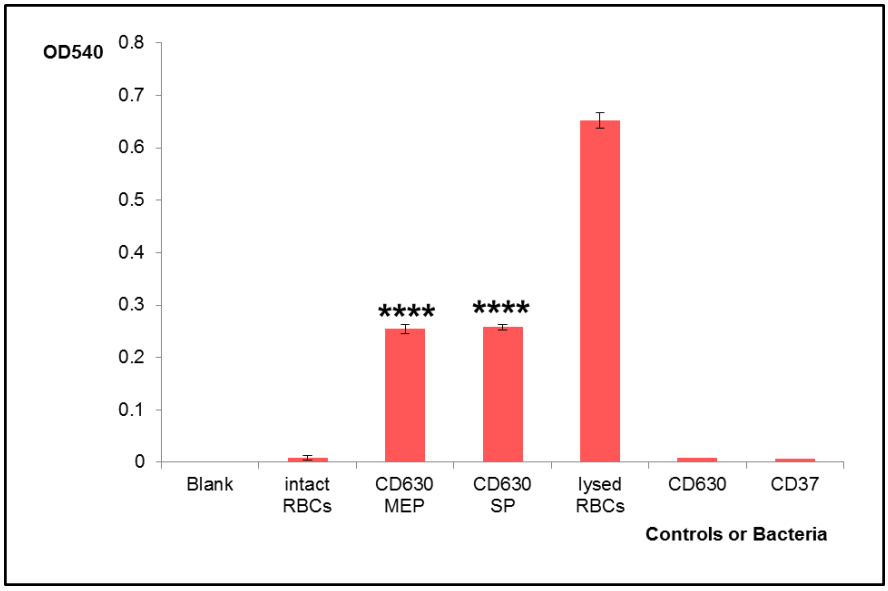


Figure 3-9 Contact haemolysis assay for CD630.

The bars represent the haemolysis of RBCs when added to CD630 at MEP and SP in comparison to intact RBCs (as a negative control) and lysed RBCs (as a positive control). CD630 and CD37 bacterial suspension without RBCs were used as negative controls. The haemolysis of CD630 at MEP and SP show significant increase in haemolysis compared to the negative control. Statistical significance is indicated by four asterisks ****, $p < 0.0001$ (unpaired two-tailed t-test measured by GraphPad Prism v8.1.2).

3.4 Discussion

C. difficile is a major health problem in hospitals and the community. It has many virulence factors that contribute to its pathogenic ability. A possible virulence factor that has not yet been reported is haemolysis. It was discovered that *C. difficile* has the ability to haemolyse horse blood in agar when supplemented with 2% glucose (Rachel Williams, Peter Mullany, and Adam P. Roberts, unpublished data). Some Gram positive bacteria were also reported to produce haemolysins after the addition of glucose. *B. cereus* produced 20 to 50 times more enterotoxins in fermenter-grown cultures that contain 1% glucose compared to cultures that do not (Garcia-Arribas & Kramer, 1990). Also *B. anthracis*, which was thought to be non-haemolytic in LB media, can cause haemolysis when grown in BHI or LB supplemented with glucose (Shannon *et al.*, 2003). A study showed that *S. aureus* produces alpha-haemolysin at an optimal level when 0.2% of glucose was added to the agar; omitting glucose causes a drop in haemolysin production while glucose concentrations higher than 0.5% tend to be inhibitory (Duncan & Cho, 1972). There is no explanation to why glucose can induce haemolysis in Gram positive or negative bacteria.

C. difficile ferments carbohydrates (such as fructose, glucose, mannitol, and mannose) with an end product of acids (acetic, butyric, propionic, valeric, isobutyric, isovaleric, isocaproic, and isoheptanoic) and minor quantities of alcohol (ethanol or butanol) (Karlsson *et al.*, 2000; Nakamura *et al.*, 1982). It was reported that when L. monocytogenes consume high concentrations of glucose, it causes a drop in pH which inhibits the production of Listeriolysin O

(Datta & Kothary, 1993). A phenomenon called “transient repression” because the bacteria resume haemolysin production when there is no more glucose left (Duncan & Cho, 1972). In a normal physiological pH of 7.4, RBCs remain intact and maintain their integrity. Haemolysis starts to occur below pH 6, where it is a slow process between pH 5.8 to 5.0, and starts to be faster between pH 5.0 and 4.0 where changes in the RBCs membrane take place. When the changes were examined by electron microscopy, RBCs membrane showed defects that eventually developed into openings (Arvinte *et al.*, 1989). The pH assay performed in this work suggests that the addition of glucose during the growth of CD630 did not cause accumulation of acidic metabolites. And hence no impact on RBCs unlike the work presented by Duncan and Cho.

Glucose was reported to affect the expression of 18% of the whole genome (667 genes) in *C. difficile* JIR8094. The transcriptome analysis showed that 349 and 318 genes were up or down regulated, respectively (Antunes *et al.*, 2012). Glucose can repress the metabolism of carbohydrates through a pleiotropic regulator, catabolic control protein CcpA, which mediates the global transcriptional response (Antunes *et al.*, 2011). Glucose can also repress the transcription of toxins A and B genes resulting in decrease in the production of these toxins (Dupuy & Sonenshein, 1998). The effect of glucose on the production of haemolysins might be explained as an adaptive mechanism exerted by the bacteria to overcome drastic environmental changes. As an example, CodY is a global regulator protein in *C. difficile* that represses genes that are not necessary when nutrients are abundant (like

toxin genes) and reverses the repression when nutrients become scarce (Dineen *et al.*, 2007). CodY in *C. difficile* was found to overexpress 146 and repress 19 genes that are involved in transport of nutrients, biosynthesis of amino acids, fermentation pathways and many more (Dineen *et al.*, 2010). These findings suggest a relationship between metabolism and virulence. Quorum sensing has been also described in *C. difficile* and was found to regulate toxin production (Carter *et al.*, 2005; Darkoh *et al.*, 2015). *C. difficile* might overexpress haemolysin genes through quorum signalling to acquire nutrients such as iron. The quorum sensing molecule, encoded by *sagA*, from *S. pyogenes* was found to overexpresses *siaA* in response to low-iron conditions (Salim *et al.*, 2007). The *siaA* is part of the sia operon that controls the haemolytic activity of *S. pyogenes* to acquire iron from the environment (Bates *et al.*, 2003; Eichenbaum *et al.*, 1996). Contrarily, Acyl homoserine lactones (AHLs) which are quorum signalling molecules produced by many microbiota have been reported to abolish haemolysis mediated by Streptolysin-S but not Streptolysin-O in *S. pyogenes* (Saroj *et al.*, 2017). This example demonstrates the complex dynamics of bacterial populations and the numerous gene regulation systems that are involved in colonization and virulence. The biochemical effect of glucose on CD630 to exhibit haemolysis was not studied in this work but rather the genes that were thought to encode haemolysins were investigated.

The CD630 library constructed by Dasgupta *et al.* in 2012 revealed a clone, containing CD630_10340 gene (*mgna*) that showed haemolysis (unpublished data). And when the gene was interrupted by an *E.coli* IS element, there was

a reduction in haemolysis. This chapter focused on determining if *mgna* was responsible for haemolysis by constructing a clone in *E. coli* containing only *mgna*. The Ec-CD630-mgna clone was haemolytic leading to the conclusion that this gene is responsible for haemolysis in *E. coli*. However, the mutant of this gene in CD630 (CD630Δmgna-ClosTron) had exactly the same haemolytic activity as the wild type CD630Δerm when grown on blood agar containing 2% glucose. These results suggested that there might be more than one gene in *C. difficile* that can encode proteins causing haemolysis. Another hypothesis is that testing haemolysis on blood agar was not sensitive enough to detect slight decrease in the haemolysis of mutants. The haemolytic activity of bacterial haemolysins in Gram positive and negative bacteria is usually tested by a quantitative assay (Ochi *et al.*, 2004; Oda *et al.*, 2010; Sansonetti *et al.*, 1986; Tay *et al.*, 1995). The measurement of haemoglobin released into the supernatant due to the breakdown of RBCs utilised in this work is the most accurate method to estimate haemolysis (Ishii, 1960; Rowe & Welch, 1994). Therefore, a haemolysis assay for *C. difficile* was developed.

3.5 Conclusions

This chapter described the haemolysis of *C. difficile* in blood agar when 2% glucose was added. Gene CD630_10340 (*mgna*) was found to be responsible for haemolysis in *E. coli* and further investigation into its role in *C. difficile* is needed. Identification of other genes in CD630 encoding haemolysins is of great importance.

Chapter 4

CD630 genomic library identifies six genes implicated in haemolysis

Chapter 4 CD630 genomic library identifies six genes implicated in haemolysis

4.1 Introduction

In this chapter, screening of the CD630 genome to find haemolysin encoding genes is described. This was done by constructing a genomic library in a host that is non-haemolytic when grown on blood agar. A number of bacterial hosts were tested for haemolytic activity for example *E.coli*, *Bacillus*, and *Streptococcus*. *E.coli* was chosen because it was not haemolytic, easy to grow in the lab, competent cells are available commercially, and is known as a model bacteria for scientific research (Cooper, 2000; Idalia & Bernardo, 2017). Colonies of *E. coli* containing the recombinant plasmid that were showing haemolysis zones were investigated. The insert DNA containing genes expressed in the haemolytic colonies were sequenced to know which gene was responsible for haemolysis. These genes were studied by nucleotide analysis tools to learn about the protein product and its function.

The use of nucleotide database search tools like BLAST will often give a long list of homologies for the genes being investigated. These are often derived from whole genome sequencing studies or homologies to genes experimentally studied. This means that the tools sometimes only provide limited information on gene functions. Therefore, new bioinformatics tools are essential for the annotation of novel genes and the prediction of protein structure and function (Hernandez *et al.*, 2009). Proteins consist of one or more structural units that fold more or less independently of each other,

commonly termed domains (Alberts *et al.*, 2014). Domains are responsible for a specific function or interaction contributing to the overall protein role. One of the tools which enables extensive annotation of protein domains with respect to function, phyletic distribution, structure, and functionally important residues is SMART (Simple Modular Architecture Research Tool). It can detect more than 500 domain families found in extracellular, signalling, and chromatin-associated proteins (Doerks *et al.*, 2002). The SMART database has more than 200 million annotated domains and other protein features through synchronisation with UniProt, Ensembl, and STRING (Letunic & Bork, 2018). The SMART also has a function to include analysis from other databases like Pfam, SCOP, TMHMM, SEG, GO, and Interproscan. Pfam (protein family) was produced by the European Bioinformatics Institute using a sequence database based on UniProt. It contains a large collection of protein families each represented by multiple sequence alignments and hidden markov models (HMMs) (Finn *et al.*, 2016). For the structure and evolutionary relationships in proteins, SCOP (Structural Classification of Proteins) is also available within SMART (Andreeva *et al.*, 2014). The tool was developed at the MRC Laboratory of Molecular Biology and Centre of Protein Engineering in Cambridge. TMHMM (Tied Mixture Hidden Markov Model) and SignalP are tools developed by DTU bioinformatics unit at Technical University of Denmark. TMHMM is a database to predict transmembrane helices in proteins (protein topology) (Moller *et al.*, 2001). SignalP is a server for the prediction of the presence and location of signal peptide cleavage sites in amino acid sequences based on a combination of several artificial neural networks (Nielsen, 2017;

Petersen *et al.*, 2011). SEG, developed by the bioinformatics group at Research Institute of Molecular Pathology (IMP) in Austria, is used to predict low complexity regions. It is a method to divide the globular and non-globular regions of protein sequences fully automatically (Wootton, 1994). InterProScan was constructed through the European Molecular Biology Laboratory and European Bioinformatics Institute (EMBL-EBI). It is a database for the classification of proteins into families and for the prediction of domains and important sites (Finn *et al.*, 2017). The Gene Ontology (GO) project is for finding the biological functions of genes at the molecular, cellular and tissue levels (Ashburner *et al.*, 2000; GO, 2017).

4.2 Materials and methods

4.2.1 Finding a non-haemolytic host for creating a genomic library

Two *E. coli* DH5α strains: α-select silver *E. coli* (Bioline, UK) and 5-alpha *E. coli* (NEB, UK) were grown on LB agar containing 5% horse blood and incubated at 37°C for 16-18 hours. Also *B. subtilis* CU2189, *S. mutans* (ATCC 700610 and ATCC 25175), and *S. australis* FRStet12 were grown on BHI supplemented with 5% horse blood and incubated at 37°C in the presence of 5% CO_2 (LEEC MkII proportional temperature controlled incubator) for 16-18 hours.

4.2.2 Construction of a genomic library

A *C. difficile* genomic library was created by partial *Hind*III digest (10 minutes using 10µl of 20,000 units/ml of *Hind*III) of CD630 genomic DNA. The digested fragments of 1-10 Kb in length were then ligated to the *Hind*III-cleaved and dephosphorylated pUC19. The recombinant DNA was introduced to chemically competent α-select silver *E. coli* cells (Bioline, UK) by transformation. The size of the library was calculated by multiplying the average insert size by the number of colonies (colony forming unit, cfu) in 1 ml of the transformation mix. The transformation mix was plated on 24 LB agar plates containing blood and screened for haemolytic colonies. Plasmid DNA was extracted from all the haemolytic colonies then pooled, re-transformed, and the ten most haemolytic colonies were picked for further analysis. Inserts in the plasmids were sequenced using universal pUC19-M13-F and pUC19-M13-R primer pair. DNA inserts larger than 1 Kb were

additionally sequenced with multiple primers (ZK1-F to ZK9-F and ZK1-R to ZK9-R) to determine the whole sequence. All standard molecular techniques are described in details in Chapter 2.

4.2.3 Bioinformatic analysis

Bioinformatic analysis was performed utilizing the following protein prediction tools: SMART v7.0 that includes analysis from Pfam v30.0, SCOP v1.75, TMHMM v2.0, SignalP v4.1, and SEG. For protein molecular function and biological process prediction UniProtKB data base was used.

4.2.4 Site-directed mutagenesis (SDM)

Site-directed mutagenesis was performed using the Q5 Site-Directed Mutagenesis Kit (NEB UK, section 2.5) to delete genes: CD630_02921, CD630_11300, and CD630_28300 (201 bp, 555 bp, and 663 bp respectively). Two custom primers for each gene were designed in opposite direction with their 5’ ends adjacent to the nucleotide sequence to be deleted. PCRs were performed utilizing the plasmid DNA of the three inserts and primer pairs ZK13-F with ZK13-R, ZK14-F with ZK14-R, and ZK15-F with ZK15-R, respectively. Transformation with subsequent plasmid DNA extraction, restriction analysis, and sequencing was conducted as in Chapter 2.

4.2.5 Constructing clones carrying ten genes encoding haemolysins

Cloning each of the ten genes discovered from CD630 library was performed. Detailed protocol was mentioned previously (section 3.3.3) and in Chapter 2. The genes were amplified from CD630 genomic DNA utilising primer pairs (ZK16-F and ZK16-R, ZK17-F and ZK17-R … to ZK25-F and ZK25-R) which contain *BamH*I or *Hind*III in the forward primer and *EcoR*I in the reverse together with a GC clamp. The amplicons were cloned into similarly digested pUC19 and transformed into α-select silver competent *E. coli*. Plasmid DNA was isolated and subjected to restriction analysis and sequencing to confirm the cloning of each gene.

4.2.6 Haemolysis assay for *E.coli*

The haemolysis of *E. coli* clones and their mutants were tested quantitatively using an assay similar to the one developed for *C. difficile* in section (3.2.6) with the following modifications. *E. coli* was cultured overnight (16-18 hours) in 15 ml LB broth supplemented with 10 µg/ml Amp in a 37°C shaker incubator (200 rpm). The assay plate was incubated aerobically at 37°C for 18 hours.

4.3 Results

4.3.1 Screening for a non-haemolytic host in which to construct *C. difficile* genomic library

A bacterial host with non-haemolytic phenotype should be used to construct the library. On blood agar, *E. coli* 5-alpha (NEB) showed β-haemolysis around colonies while α-select silver *E. coli* (Bioline) was non-haemolytic. *B. subtilis* CU2189 and *S. mutans* (ATCC 700610 and ATCC 25175) showed β-haemolysis whereas *S. australis* FRStet12 was α-haemolytic (Figure 4-1). Therefore α-select silver *E. coli* was used to create the library.

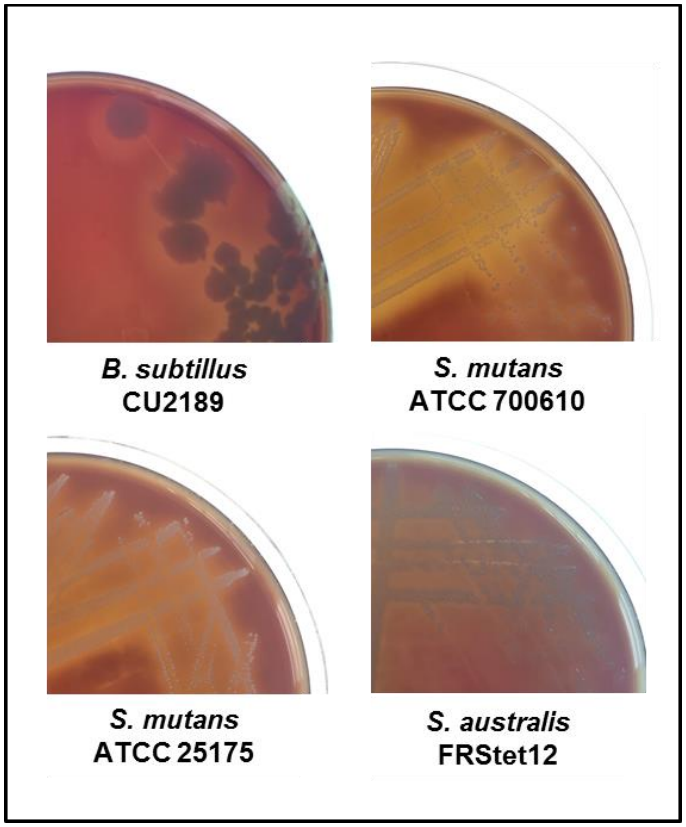


Figure 4-1 Growth of possible genomic library hosts on blood agar.

B. subtilis CU2189 and *S. mutans* (ATCC 700610 and ATCC 25175) were β-haemolytic whereas *S. australis* FRStet12 was α-haemolytic compared to **Figure 1-11**.

4.3.2 Constructing and screening of a *C. difficile* genomic library

A CD630 genomic library was constructed in *E. coli* to look for genes that encode haemolysis (methods section 4.2.4). The genomic library had an average insert size of 1,780 bp and there were a total of 22 x 10^6 clones. The number of clones required to screen one whole genome of CD630 was 2,410 clones (the size of CD630 genome 4,290,252 bp divided by the average insert size 1,780 bp). The genomic library yielded 61 haemolytic colonies (1.3%) out of 4,800 that were screened. The high number of haemolytic clones can be explained by the fact that sometimes multiple clones contain the same insert. The plasmid DNA of the 61 haemolytic colonies were pooled and retransformed in parallel with a control containing the plasmid only without insert which was non-haemolytic (Figure 4-2). Ten colonies with largest zone of haemolysis were selected for further analysis. Plasmid DNA extraction followed by *Hind*III digest showed that three out of the ten haemolytic clones had different insert sizes (Figure 4-3). That was confirmed with another digest of the plasmid DNA using *EcoR*I. The plasmid DNA of the three clones was retransformed three times to rule out spontaneous mutation in *E. coli* and always showed haemolysis.

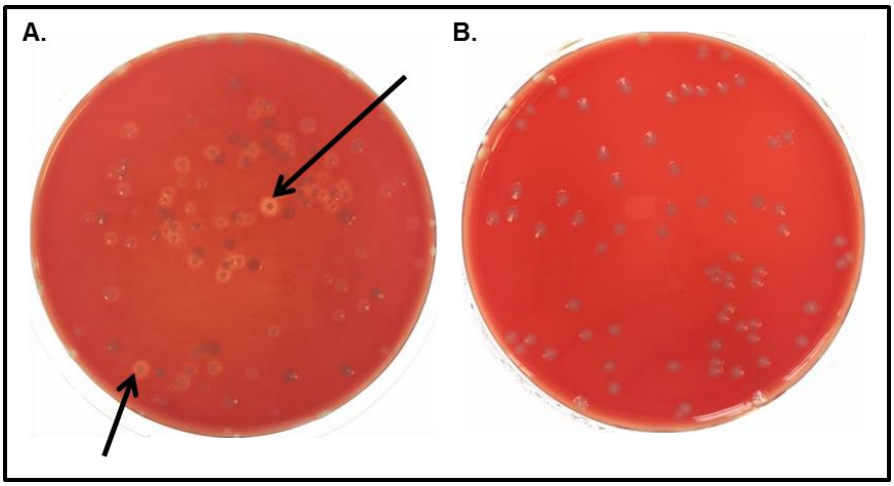


Figure 4-2 Blood agar plates from the CD630 library.

A. *E. coli* carrying pUC19 with CD630 DNA from the library showing haemolytic and non-haemolytic colonies. Arrows pointing to the colonies showing more haemolysis that were selected for further analysis. **B.** *E. coli* carrying the empty plasmid used as control showing non-haemolytic colonies.

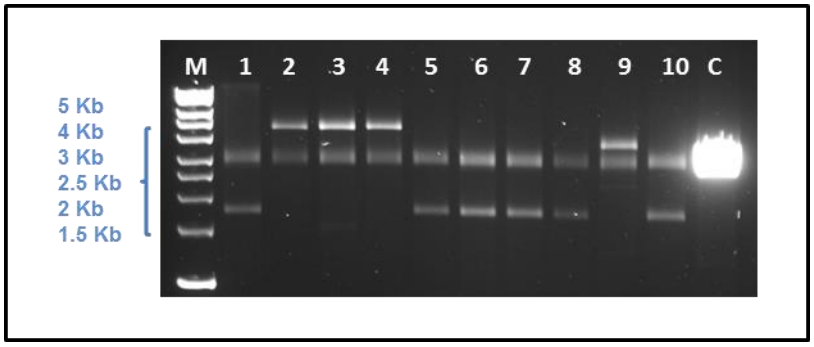


Figure 4-3 *Hind*III digest of the DNA of 10 haemolytic *E. coli* clones.

The digest was showing three different insert sizes. Clones 1, 5, 6, 7, 8, and 10 are the same (1.8 Kb) and clones 2, 3, 4 are the same (5 Kb) and clone 9 was 4 Kb. M, 1 Kb DNA ladder; 1-10, ten haemolytic clones digested with *Hind*III; C, pUC digested with *Hind*III (2.68 Kb).

4.3.3 Sequence analysis of the haemolytic *E. coli* clones

Sequence analysis of the three haemolytic *E. coli* clones: Ec-pUC-Lib-1.6, Ec-pUC-Lib-4.3, and Ec-pUC-Lib-3.4 revealed 1,630 bp, 4,257 bp, and 3,433 bp insert, respectively (Figure 4-4). The latter two inserts were too large, as a result the initial sequence data only revealed 1 Kb from the 3' and 5' ends leaving the middle part not sequenced. Multiple primers were designed to sequence about 2,257 bp and 1,433 bp of Ec-pUC-Lib-4.3 and Ec-pUC-Lib-3.4 respectively which confirmed the genes within the inserts (data not shown).

The insert in Ec-pUC-Lib-1.6 had three genes: one complete and two partial ones at each end of the insert (Figure 4-4A). The genes from upstream to downstream were predicted to encode: an HTH-transcriptional regulator (CD630_02920, range 353054 to 354148), an uncharacterized protein (CD630_02921, range 354153 to 354353), and an ABC transporter ATP-binding protein (CD630_02930, range 354353 to 355270).

The insert in Ec-pUC-Lib-4.3 had four genes: two partial and two complete (Figure 4-4B). The genes were predicted to encode DNA polymerase I (POLI) (CD630_11280, range 1326574 to 1329222), dephospho-coenzyme A kinase (CD630_11290, range 1329231 to 1329833), putative lytic transglycosylase-like protein (CD630_11300, range 1329826 to 1329380), and putative solute-binding lipoprotein (ABC transporter substrate-binding protein) (CD630_11310, range 1330380 to 1332095).

And the insert in Ec-pUC-Lib-3.4 had three genes: two partial and one complete (Figure 4-4C). The genes were predicted to encode a putative membrane protein (CD630_28290, range 3301529 to 1302491), uncharacterized protein (CD630_28300, range 3302613 to 1303275), and a putative adhesin (CD630_28310, range 3303646 to 1306564).

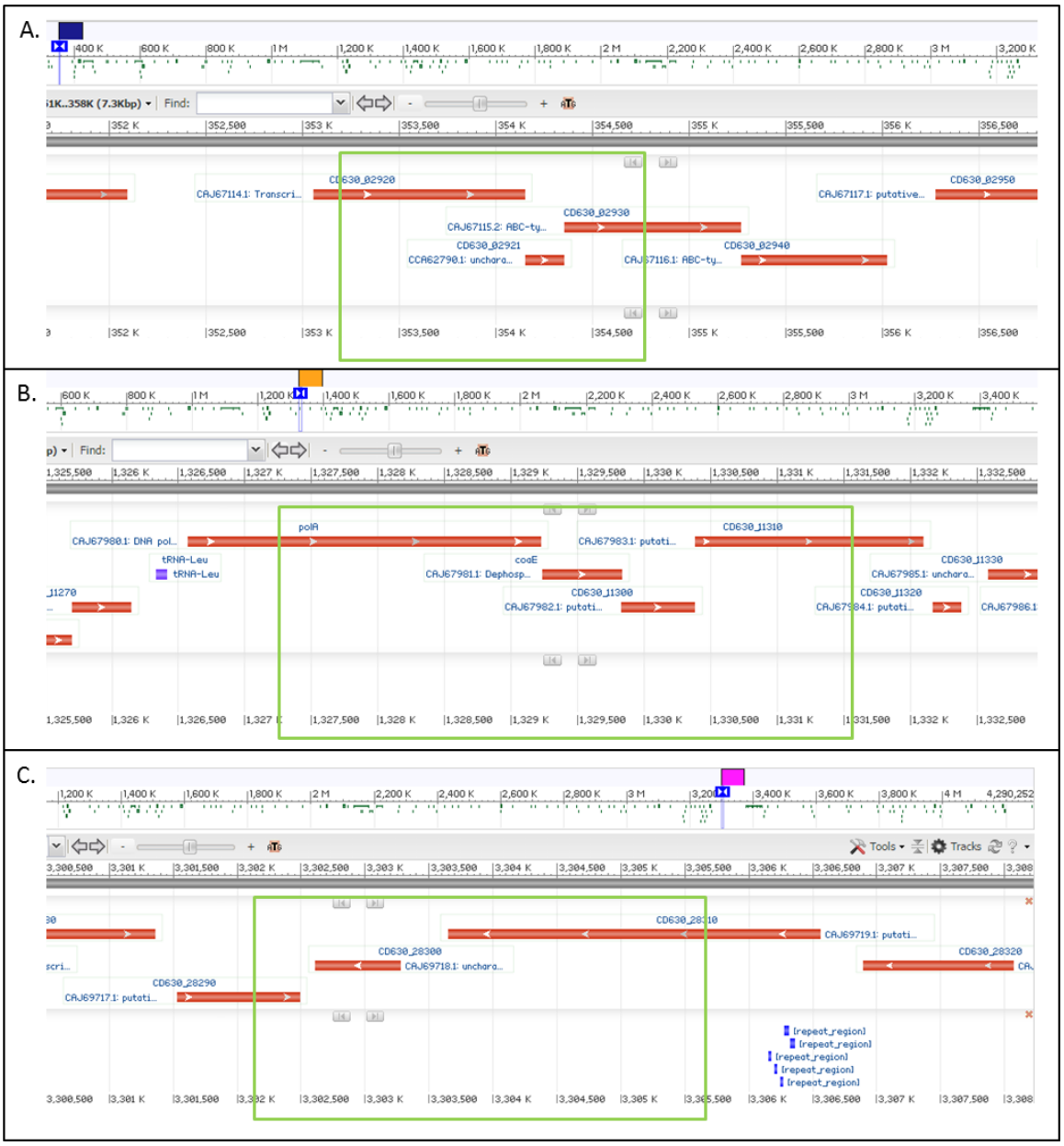


Figure 4-4 Diagram of the *C. difficile* genome region that had been cloned into the haemolytic clones from the 630 library.

The figures were taken from NCBI graphics where the green boxes are an approximate representation of the inserts. A. Insert in Ec-pUC-Lib-1.6 had three genes (inside the green box) encoding from left to right: HTH-transcriptional regulator, uncharacterized protein, and ABC transporter ATP-binding protein. B. Insert in Ec-pUC-Lib-4.3 had four genes encoding: DNA polymerase I (polA), dephospho COA kinase, putative lytic transglycosylase-like protein, and ABC transporter substrate-binding protein. C. Insert in Ec-pUC-Lib-3.4 had three genes encoding: putative membrane protein, uncharacterized protein, and putative adhesin.

4.3.4 Bioinformatic analysis of the genes isolated from the library

Further analysis was done using protein prediction tools that will give details about the family and domains of the proteins as well as the function, topology, and structure if available. The properties of the proteins gleaned from bioinformatic analysis are summarised in Table 4-1.

The predicted transcriptional regulator protein (364 residues) belonged to the helix-turn-helix (HtH) xenobiotic response element (XRE) family of transcriptional regulators with a Cro/C1-type HTH domain (PF01381, residues 9-64). The Cro/C1 domain name is derived from the transcriptional repressors Cro and C1 of temperate bacteriophages 434 and Lambda, respectively. Proteins with an HtH motif bind DNA in a sequence-specific manner through the binding of its recognition helix with the major groove of target DNA. Hence, these proteins are primarily involved in gene regulation. The uncharacterised protein in Ec-pUC-Lib-1.6 was predicted to have a phospholipase D-nuclease N-terminal domain (PLDc_N) (PF13396, residues 20-63). This domain is often found at the N-terminus of the phospholipase D-nuclease family of proteins. The protein was very short (66 residues) with 2 transmembrane α-helix regions (residues 7-29 and 42-61). UniProtKB alignment of this protein revealed 34-70% homology to 32 proteins from different species (mainly *Lactobacillus spp.* and *Clostridium spp.*) and all were annotated as membrane or uncharacterised proteins with unknown functions. The ABC transporter ATP-binding protein (305 residues) has a domain that belongs to the P-loop containing nucleoside triphosphate hydrolases (P-loop NTPase) superfamily and ATPases associated with

diverse cellular activities family (AAA domain, residues 28-208). This superfamily is found in all kingdoms of living organisms and is involved in the hydrolysis of ATP to energise diverse cellular processes. The protein alignment only detected a fraction of this vast family and the poorly conserved N-terminal helix is missing. Another domain predicted which overlaps with the AAA domain is the ABC transporter-like domain which belongs to the ATP-binding cassette superfamily (PF00005). Both superfamilies have the same biological function. UniProtKB and GO databases predicted a ferric-transporting ATPase activity that catalyses the transfer of ferric ions across the membrane.

The DNA polymerase I protein encoded by Ec-pUC-Lib-4.3 had 882 amino acid residues and was predicted to have three domains. The N-terminus and internal 5' to 3' exonuclease (53EXOc_N) domains (PF02739, residues 3-261) which are commonly found together and are associated with 5' to 3' nuclease activities. The second domain was the 3' to 5' exonuclease (35EXOc) domain (PF01612, residues 299-473) which is responsible for the 3' to 5' exonuclease proofreading activity (hydrolysis of unpaired or mismatched nucleotides). And finally the DNA polymerase 1 domain (PF00476, residues 504-880) which belong to the DNA polymerase A family. The family has a DNA-directed DNA polymerase activity responsible for DNA replication. SMART predicted a helix-hairpin-helix class 2 (HhH2) motif (SM000279, residues 173-208) that is similar to, but distinct from, HtH and the helix-loop-helix (HLH) motifs. These motifs have two helices (H1 and H2) connected by a short turn. The HhH motif is present in non-sequence-specific

DNA binding proteins where binding occurs through hydrogen bonds between nitrogen atoms in the protein backbone and phosphate groups in DNA. The second protein (200 residues) has a dephospho-coenzyme-A kinase (CoaE) domain (PF01121, residues 2-182) belonging to the P-loop containing nucleoside triphosphate hydrolases (P-loop NTPase) superfamily. The CoaE family catalyses the final step in CoA biosynthesis, the phosphorylation of the 3'-hydroxyl group of ribose using ATP as a phosphate donor. The third protein (184 residues) had a soluble lytic transglycosylase (SLT) domain (PF01464, residues 37-152) and a transmembrane region (residues 7-26). This domain belongs to the lysozyme-like superfamily which degrade murein via cleavage of the β-1,4-glycosidic bond between N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG), with the concomitant formation of a 1,6-anhydrobond in the muramic acid residue. SLT domains are found mainly in proteins from phages and type II, type III and type IV secretion systems. These proteins have hydrolase activity acting on glycosyl bonds. The transmembrane region contained one α-helix indicating that the rest of the protein is outside of the membrane. The last protein (571 residues) contained a bacterial extracellular solute-binding protein family 5 (SBP_bac_5) domain (PF00496, residues 84-495). It belongs to the periplasmic binding protein-like II superfamily which has high affinity transport systems involved in active transport of solutes across the cytoplasmic membrane. The protein also contained a signal peptide region (residues 1-19) that is known to be present at the N-terminus of the majority of newly synthesized proteins that are destined towards the secretory pathway. One of the members of the solute-binding protein family 5 is the

Haem-binding lipoprotein from *Haemophilus influenza* which binds heme *in vitro* and promotes the binding of this compound by *E. coli* recombinants expressing this protein (Hanson *et al.*, 1992). Another family member is the Sap (sensitivity to antimicrobial peptides) of *H. influenzae* which has a transport function that promotes the utilization of heme (Mason *et al.*, 2011). Also NikA (nickel-binding periplasmic protein) which is a protein known to bind to nickel in *E. coli* has a heme binding function (Shepherd *et al.*, 2007).

The putative membrane protein (320 residues) encoded by Ec-pUC-Lib-3.4 had two EamA domains (named after the O-acetyl-serine/cysteine export gene in *E. coli*) with 10 α-helices (PF00892, residues 2-139 and 147-302). Many members of this family are classed as drug or metabolite transporters but have unknown function. The second protein (uncharacterized, 220 residues) was predicted to have a signal peptide (residues 1-26) and an anthrax toxin lethal factor (ATLF) domain (PF07737, residues 122-220) homologous to that found in *B. anthracis.* The protein is homologous to the C-terminal domain found in both the lethal (LF) and edema factor (EF) proteins that usually bind to the protective antigen (PA) protein in the toxin. The three component proteins self-assemble at the surface of host cell receptors, causing a series of toxic complexes that can produce shock-like symptoms and death (Young & Collier, 2007). InterProScan predicted a metallopeptidase catalytic domain (IPR024079, residues 51-218), which overlaps the ATLF domain. There are over 50 metallopeptidase families known and all need a divalent cation (usually zinc but it can be any other) for the hydrolysis of peptide bonds which are essential for all major physiological

processes. The last protein was a putative adhesin (972 residues) that had two collagen binding domains (PF05737, residues 329-439 and 456-598). The adhesin protein family function as receptors for target cell binding which is important for the pathogenesis of bacteria. The protein also had a collagen adhesin B-type (Cna_B) domain (PF05738, residues 750-818), one transmembrane (residues 946-965), one signal peptide (residues 1-23), and three low complexity regions. Both domains were found in *S. aureus* protein (Cna) important for adherence to cartilage.

Table 4-1 Summary of the bioinformatic analysis for the proteins encoded by ten putative genes obtained from the CD630 genomic library.

Gene	Gene ID	Protein predicted	Size of protein	Family	Domain	Other Features	Function
1-1	CD630_02920	HtH- transcriptional regulator	364	HtH-XRE transcriptional regulators	Cro/C1 type HtH	-	gene regulation
1-2	CD630_02921	Uncharacterised protein	66	PLDc_N	PLDc_N	two transmembrane α-helix regions	unknown
1-3	CD630_02930	ABC transporter / ATP binding protein	305	P-loop NTPase and ATP-binding cassette	AAA and ABC transporter-like	-	hydrolyse ATP to energise cellular activities
2-1	CD630_11280	DNA polymerase I	882	DNA polymerase A	53EXOc_N & 35EXOc & DNA polymerase 1	HhH2 motif	DNA replication
2-2	CD630_11290	Dephospho-CoA kinase	200	P-loop NTPase CoaE	CoaE	-	catalyses the final step in CoA biosynthesis
2-3	CD630_11300	Putative lytic transglycosylase-like protein	184	lysozyme-like	SLT	transmembrane α-helix region	hydrolase activity
2-4	CD630_11310	Putative solute-binding lipoprotein	571	periplasmic binding protein-like II	SBP_bac_5	signal peptide	transport of solutes across the cytoplasmic membrane

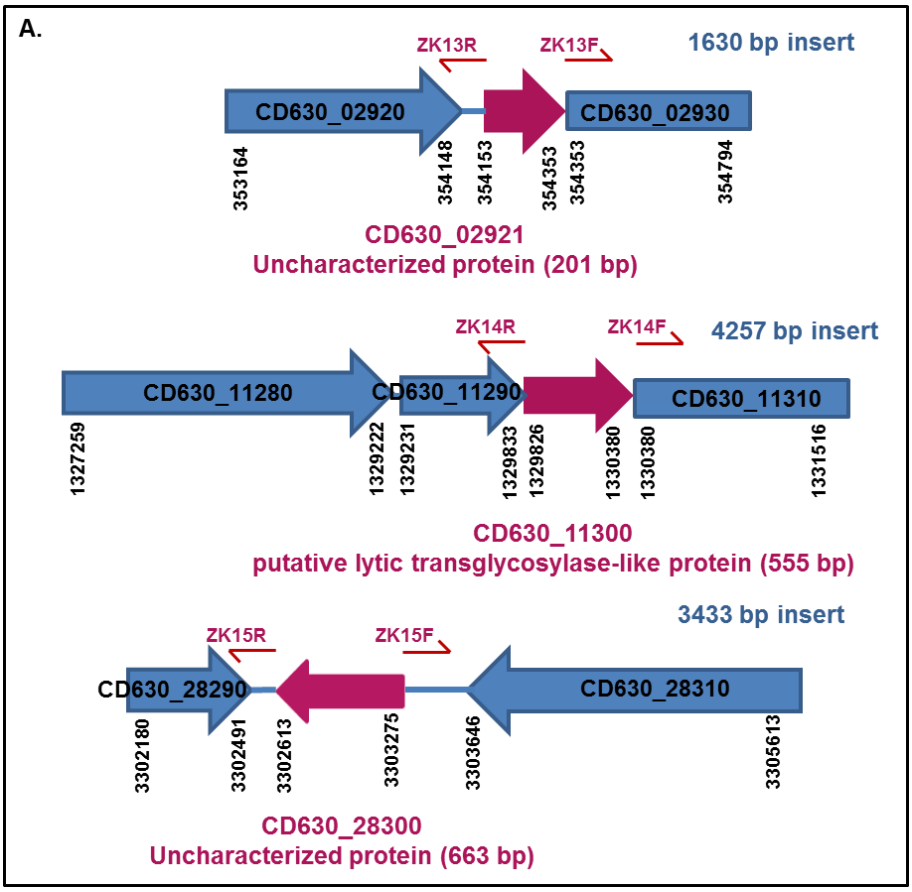
Gene	Gene	Protein predicted	Size of protein	Family	Domain	Other Features	Function
3-1	CD630_28290	Putative membrane protein	320	EamA	EamA	-	efflux mechanism
3-2	CD630_28300	Uncharacterised protein	220	metallopeptidase	ATLF and metallopeptidase	signal peptide	anthrax toxin or breaking down peptide bonds
3-3	CD630_28310	Putative adhesin	972	adhesin proteins	collagen binding Cna_B	signal peptide, transmembrane, 3 low complex regions	adherence

*Genes in clone Ec-pUC-Lib-1.6 are highlighted in grey, genes in Ec-pUC-Lib-4.3 in orange, and genes in Ec-pUC-Lib-3.4 are in green. HtH, helix-turn-helix; XRE, xenobiotic response element; Cro/C1, transcriptional repressors Cro and C1 of bacteriophage 434 and lambda respectively; PLDc_N, phospholipase D-nuclease N-terminal domain; NTPase, nucleoside triphosphate hydrolases; AAA, ATPases associated with diverse cellular activities; 53EXOc_N, N-terminus and internal 5' to 3' exonuclease; 35EXOc, 3' to 5' exonuclease; HhH2, helix-hairpin-helix class 2; CoaE, dephospho-coenzyme-A kinase; SLT, soluble lytic transglycosylase; SBP_bac_5, bacterial extracellular solute-binding proteins family 5; EamA, named after the cysteine and O-acetyl-L-serine gene in *E. coli*; ATLF, anthrax toxin lethal factor; Cna_B, collagen adhesin B-type

4.3.5 Deletion of three putative genes from the haemolytic clones

The genes that were intact in the haemolysin producing clones *i.e.* those encoding the uncharacterized proteins (CD630_02920 and CD630_28300) and the putative lytic transglycosylase-like protein (CD630_11300) were further studied. The three genes (201 bp, 555 bp, and 663 bp respectively) were deleted by site-directed mutagenesis using PCR and custom primers (Figure 4-5A). Restriction analysis of the plasmid DNA of the mutants revealed the correct inserts (1,429 bp, 3,702 bp and 2,770 bp respectively) after the deletion (Figure 4-5B). Three *E. coli* mutants: Ec-pUC-Lib-1.6Δ02921, Ec-pUC-Lib-4.3Δ11300, and Ec-pUC-Lib-3.4Δ28300 were constructed. Sequence analysis confirmed that the mutants had a deletion in the correct position (Figure 4-6, Figure 4-7, and Figure 4-8).

In Ec-pUC-Lib-4.3, CD630_11300 gene overlapped with both genes upstream and downstream of it. Deletion of this gene has caused an accidental deletion of (5'-GTGAATAG-3') including the stop codon of the upstream gene (CD630_11290) and deletion of AT nucleotides from the start codon of the downstream gene (CD630_11310) (Figure 4-7). This was non-avoidable in order to delete the full gene. It was postulated that this might have caused generation of a new open reading frame (ORF). But examination of the predicted amino acid sequence showed that this did not occur (Figure 4-9).



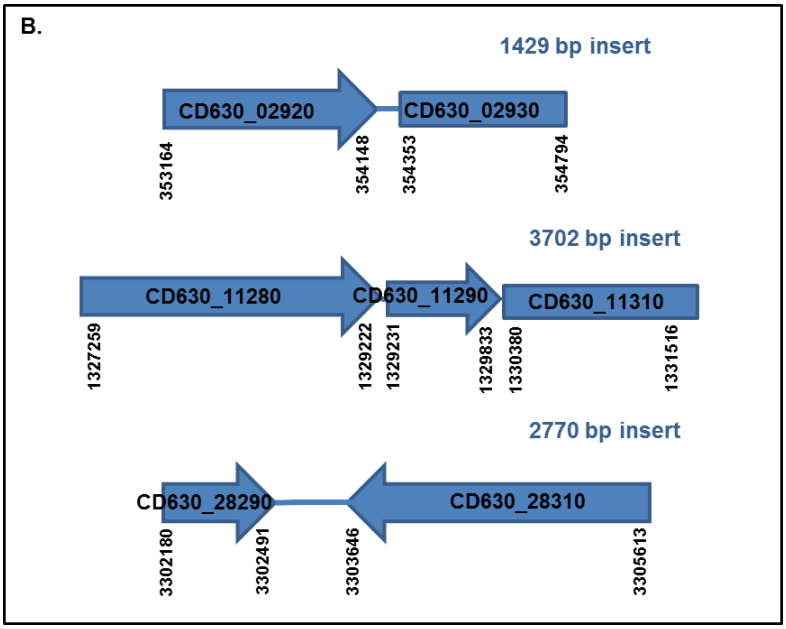


Figure 4-5 Site-directed mutagenesis of the three clones found from CD630 library.

A. An illustration of the three clones with the genes found in each (Blue arrows). The genes in pink are the ones to be deleted and the small pink arrows represent the position of the primers used in the SDM. The numbers underneath the gene represent their position within CD630 genome. **B.** An illustration of the mutants after the deletion of genes (the drawings are not to scale)

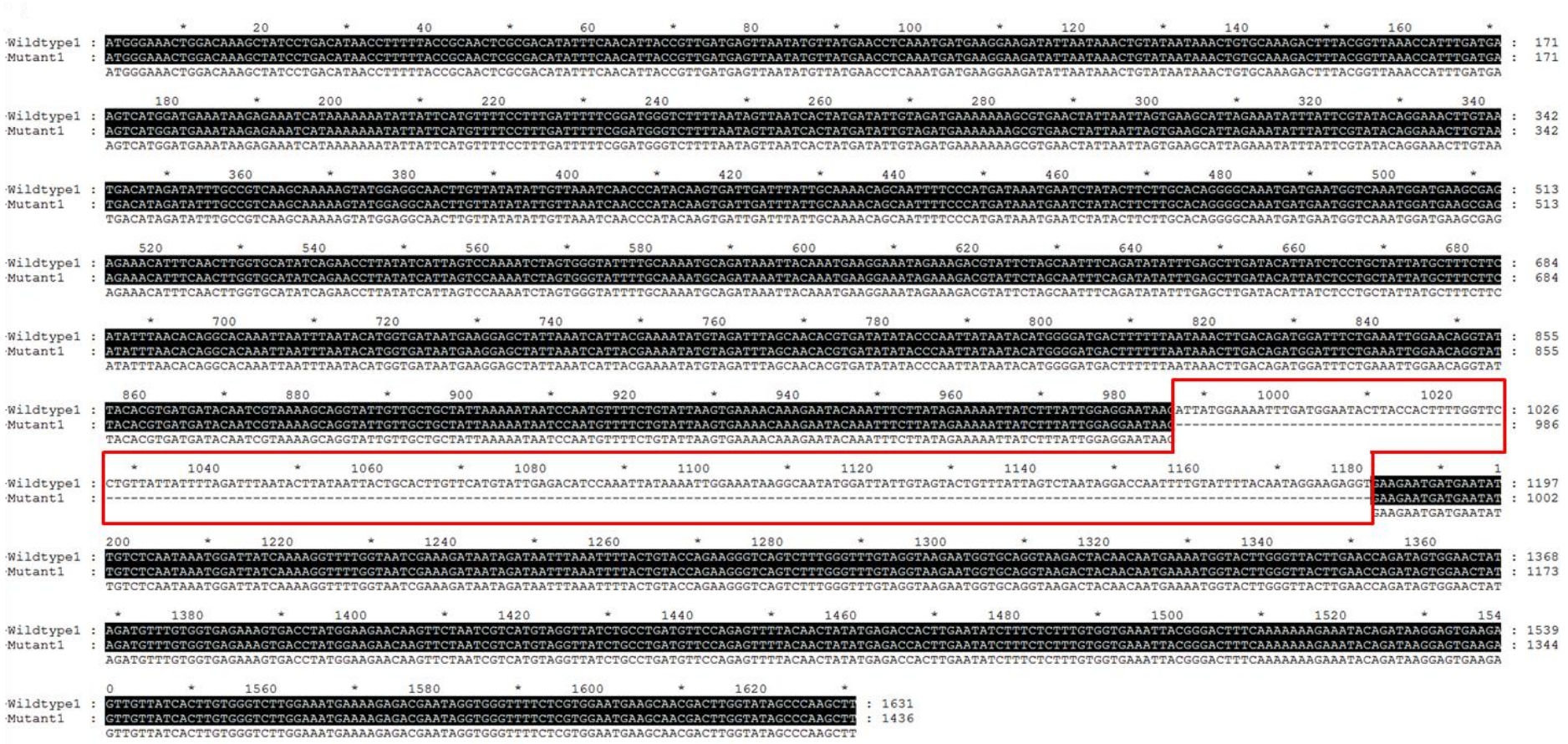


Figure 4-6 Alignment of the sequence of SDM mutant Ec-pUC-Lib-1.6Δ02921 with the wild type.

The sequence of the SDM mutant was aligned with the wild type clone Ec-pUC-Lib-1.6. The gene deleted (CD630_02921) in the mutant is represented by the red box. This alignment was performed using GeneDoc.

Figure 4-7 Alignment of the sequence of SDM mutant Ec-pUC-Lib-4.3Δ11300 with the wild type.

The sequence of the SDM mutant was aligned with the wild type clone Ec-pUC-Lib-4.3. The gene deleted (CD630_11300) in the mutant is represented by the red box. The arrows point to the deleted sequence of the upstream (CD630_11290 deletion of 5'-GTGAATAG-3') and downstream (CD630_11310 deletion of 5'-AT-3') genes. This alignment was performed using GeneDoc.

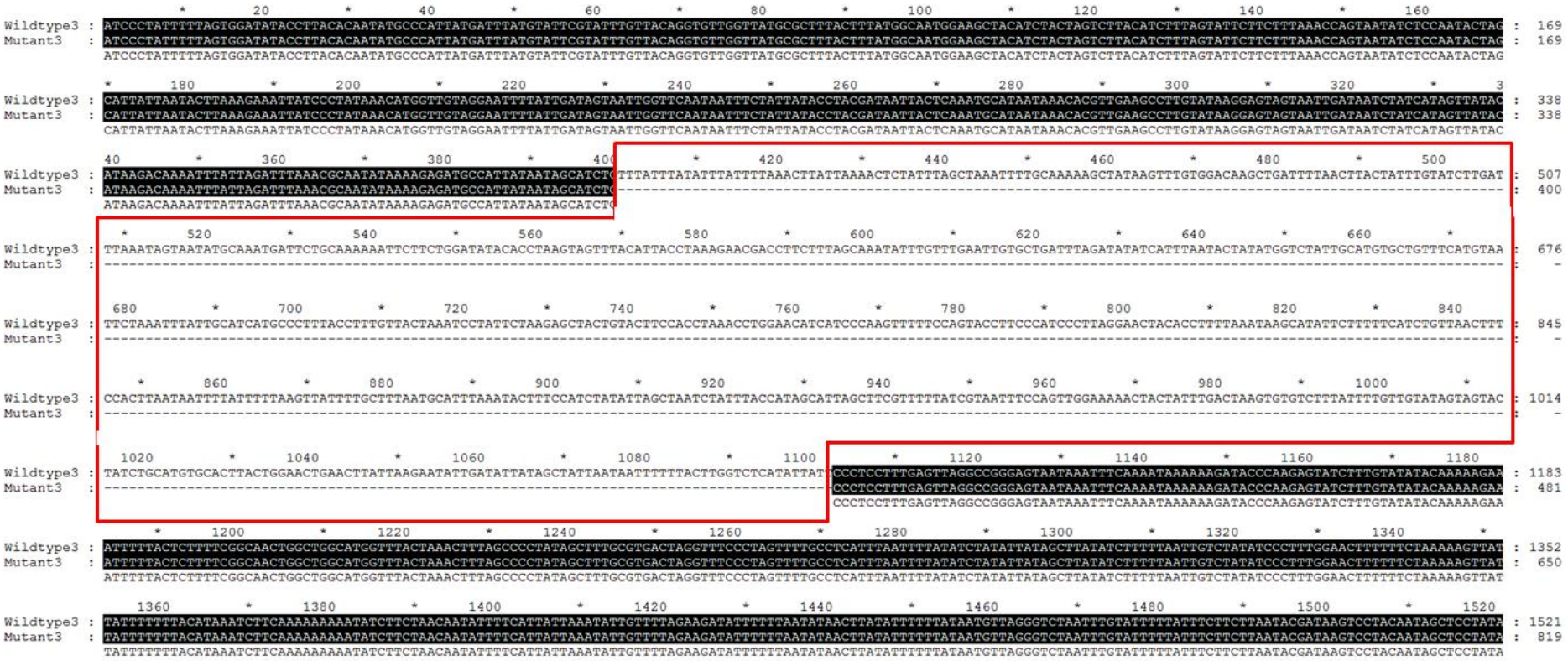
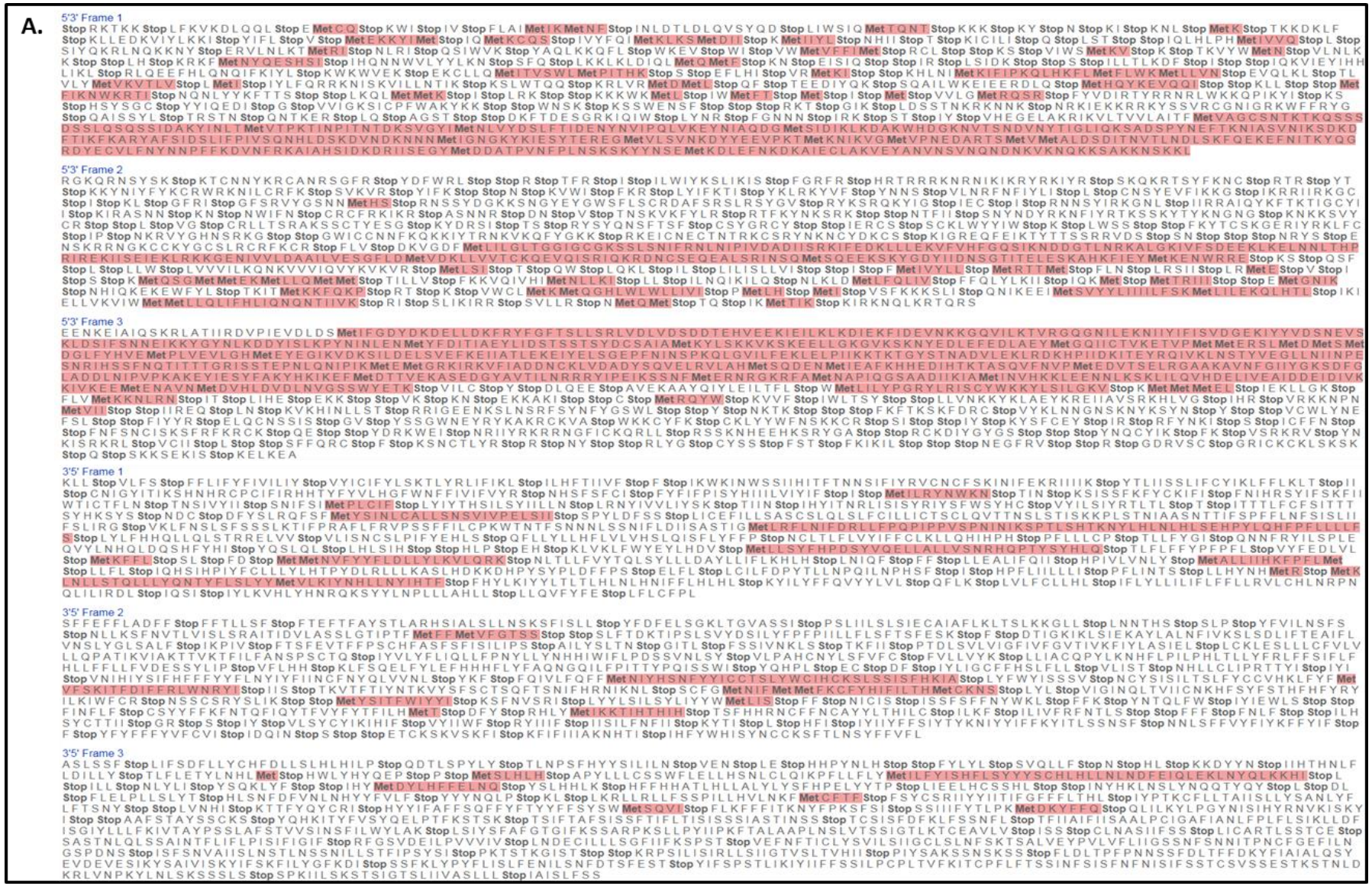


Figure 4-8 Alignment of the sequence of SDM mutant Ec-pUC-Lib-3.4Δ28300 with the wild type.

The sequence of the SDM mutant was aligned with the wild type clone Ec-pUC-Lib-3.4. The gene deleted (CD630_28300) in the mutant is represented by the red box. This alignment was performed using GeneDoc.


A.
5'3' Frame 1
5'3' Frame 2
5'3' Frame 3
3'5' Frame 1
3'5' Frame 2
3'5' Frame 3

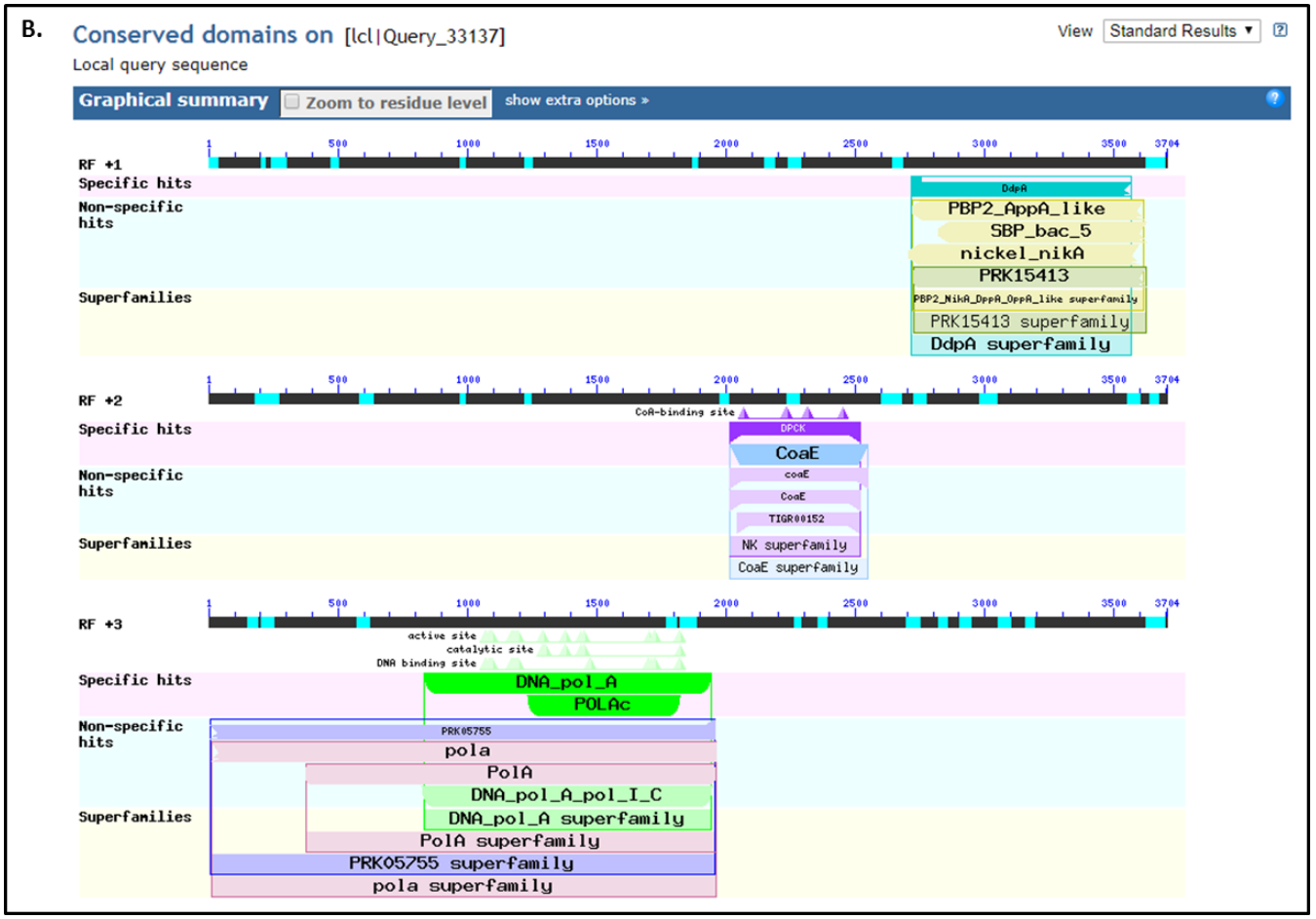


Figure 4-9 Translation of the DNA sequence of mutant Ec-pUC-Lib-4.3Δ11300.

A. The translation was performed using ExPasy and all the possible ORFs were highlighted in pink. There were three long ORFs from 5' to 3' named in figure 5'3' Frame 1, 2, and 3. **B.** Analysis by BLASTX of these ORFs encoded the same proteins found in Ec-pUC-Lib-4.3: DNA polymerase I, dephospho-coenzyme-A kinase, and bacterial extracellular solute-binding protein, respectively. Therefore, the SDM did not cause the accidental formation of a new protein.

4.3.6 Testing the haemolysis of the SDM mutants

The mutants were cultured on LB agar containing blood and compared to the parental clones (wild type). *E. coli* and *E. coli* carrying empty pUC19 were grown as well as non-haemolytic controls. After 48 hours of incubation, Ec-pUC-Lib-1.6Δ02921 was non-haemolytic while mutants Ec-pUC-Lib-4.3Δ11300 and Ec-pUC-Lib-3.4Δ28300 showed haemolysis that was identical to the wild type (Figure 4-10). The results are summarized in Table 4-2.

It was noticed that few haemolytic colonies (3 out of 160 colonies, Figure 4-10) appeared on the agar every time mutant Ec-pUC-Lib-1.6Δ02921 was grown. PCR and sequencing confirmed that the colonies had a deletion in gene CD630_02921. The controls which were *E. coli* and *E. coli* carrying pUC19 without insert never produced haemolytic colonies. Regarding the other two clones, Ec-pUC-Lib-4.3 and Ec-pUC-Lib-3.4, deleting one gene (CD630_11300 and CD630_28300 respectively) in each did not affect haemolysis probably because the remaining genes are encoding haemolysis. As a result, it was decided to clone each of the ten genes individually to localise the gene responsible for haemolysis.

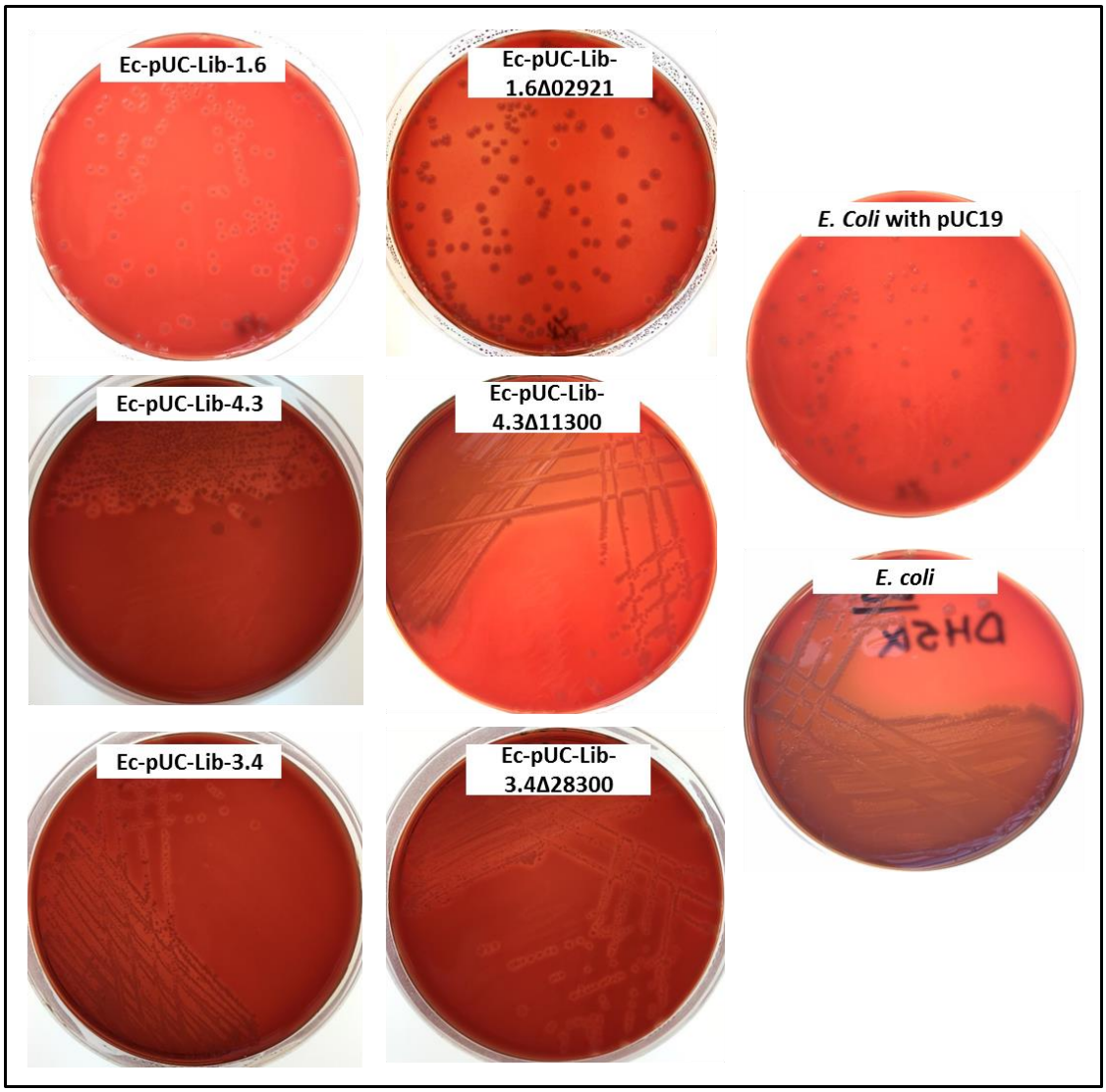


Figure 4-10 Haemolysis of the CD630 genomic library clones and their mutants.

The three clones (Ec-pUC-Lib-1.6, Ec-pUC-Lib-4.3, and Ec-pUC-Lib-3.4) to the left of the figure showed haemolysis. Mutant Ec-pUC-Lib-1.6Δ02921 was non-haemolytic while mutants Ec-pUC-Lib-4.3Δ11300 and Ec-pUC-Lib-3.4Δ28300 were haemolytic (the middle plates). Two non-haemolytic controls, *E. coli* and *E. coli* carrying empty plasmid, were also cultured (right side plates).

Table 4-2 Summary of the haemolytic activity of the clones and their mutants.

	Gene	Presence	Proteins encoded by the genes	Haemolysis
Clone Ec-pUC-Lib-1.6	CD630_02920	Partial	HtH- transcriptional regulator	Haemolytic
	CD630_02921	Complete	Uncharacterised protein	
	CD630_02930	Partial	ABC transporter / ATP binding protein	
Clone Ec-pUC-Lib-4.3	CD630_11280	Partial	DNA polymerase I	Haemolytic
	CD630_11290	Complete	Dephospho-CoA kinase	
	CD630_11300	Complete	Putative lytic transglycosylase-like protein	
	CD630_11310	Partial	Putative solute-binding lipoprotein	
Clone Ec-pUC-Lib-3.4	CD630_28290	Partial	Putative membrane protein	Haemolytic
	CD630_28300	Complete	Uncharacterised protein	
	CD630_28310	Partial	Putative adhesin	
Ec-pUC-Lib-1.6Δ02921	CD630_02920	Partial	HtH- transcriptional regulator	Not haemolytic
	CD630_02930	Partial	ABC transporter / ATP binding protein	
Ec-pUC-Lib-4.3Δ11300	CD630_11280	Partial	DNA polymerase I	Haemolytic
	CD630_11290	Complete	Dephospho-CoA kinase	
	CD630_11310	Partial	Putative solute-binding lipoprotein	
Ec-pUC-Lib-3.4Δ28300	CD630_28290	Partial	Putative membrane protein	Haemolytic
	CD630_28310	Partial	Putative adhesin	

The three clones retrieved from the genomic library are haemolytic. Deletion of the middle gene from each clone resulted in loss of haemolysis in mutant Ec-pUC-Lib-1.6Δ02921 and no change in haemolysis in the other two mutants. Clone Ec-pUC-Lib-1.6 and its SDM mutant Ec-pUC-Lib-1.6Δ02921 are highlighted in grey, Ec-pUC-Lib-4.3 and mutant Ec-pUC-Lib-4.3Δ11300 are in orange, and Ec-pUC-Lib-3.4 and mutant Ec-pUC-Lib-3.4Δ28300 are in green.

4.3.7 Cloning of ten genes retrieved from the CD630 library

To determine which gene from the genomic library was responsible for haemolysis, each of the ten genes was amplified using primer pairs that contained *BamH*I or *Hind*III and *EcoR*I then cloned into pUC19 and transformed into α-select silver *E. coli*. The constructed *E. coli* clones: Ec-CD630-02920, Ec-CD630-02921, Ec-CD630-02930, Ec-CD630-11280, Ec-CD630-11290, Ec-CD630-11300, Ec-CD630-11310, Ec-CD630-28290, Ec-CD630-28300, and Ec-CD630-28310 are listed in Table 4-3. Each clone was cultured on LB agar containing blood and compared to non-haemolytic *E. coli* carrying the empty plasmid. Out of the ten clones, four (Ec-CD630-02921, Ec-CD630-11290, Ec-CD630-11300, and Ec-CD630-28310) showed slight haemolysis observed only when the colonies were scraped from the plate by a loop. One clone (Ec-CD630-11310) that contained a gene encoding a putative solute-binding lipoprotein showed haemolysis around the colonies. And one clone (Ec-CD630-28300) that contained a gene encoding a protein with ATLF domain showed a large zone of haemolysis (Figure 4-11).

Table 4-3 Haemolysis of the 10 genes retrieved from the CD630 genomic library which were cloned individually.

Clone	Gene	Proteins encoded by the genes	Domain	Signal peptide	Haemolysis
Ec-CD630-02920	CD630_02920	HtH- transcriptional regulator	Cro/C1 type HtH	-	Not haemolytic
Ec-CD630-02921	CD630_02921	Uncharacterised protein	PLDc_N	-	Haemolysis under the colonies
Ec-CD630-02930	CD630_02930	ABC transporter / ATP binding protein	AAA ABC transporter-like	-	Not haemolytic
Ec-CD630-11280	CD630_11280	DNA polymerase I	53EXOc_N, 35EXOc DNA polymerase 1	-	Not haemolytic
Ec-CD630-11290	CD630_11290	Dephospho-CoA kinase	CoaE	-	Haemolysis under the colonies
Ec-CD630-11300	CD630_11300	Putative lytic transglycosylase-like protein	SLT	-	Haemolysis under the colonies
Ec-CD630-11310	CD630_11310	Putative solute-binding lipoprotein	SBP_bac_5	Yes	Haemolytic
Ec-CD630-28290	CD630_28290	Putative membrane protein	EamA	-	Not haemolytic
Ec-CD630-28300	CD630_28300	Uncharacterised protein	ATLF metallopeptidase	Yes	Very haemolytic
Ec-CD630-28310	CD630_28310	Putative adhesin	collagen binding Cna_B	Yes	Haemolysis under the colonies

Clones that showed haemolysis under the colonies are highlighted in blue, the haemolytic are in pink, and the non-haemolytic are not highlighted. Cro/C1, transcriptional repressors Cro and C1 of bacteriophage 434 and lambda; HtH, helix-turn-helix; PLDc_N, phospholipase D-nuclease N-terminal; AAA, ATPases associated with diverse cellular activities; 53EXOc_N, N-terminus and internal 5' to 3' exonuclease; 35EXOc, 3' to 5' exonuclease; CoaE, dephospho-coenzyme-A kinase; SLT, soluble lytic transglycosylase; SBP_bac_5, bacterial extracellular solute-binding proteins family 5; EamA, named after the cysteine and O-acetyl-L-serine gene in *E. coli* responsible for efflux mechanism; ATLF, anthrax toxin lethal factor; Cna_B, collagen adhesin B-type.

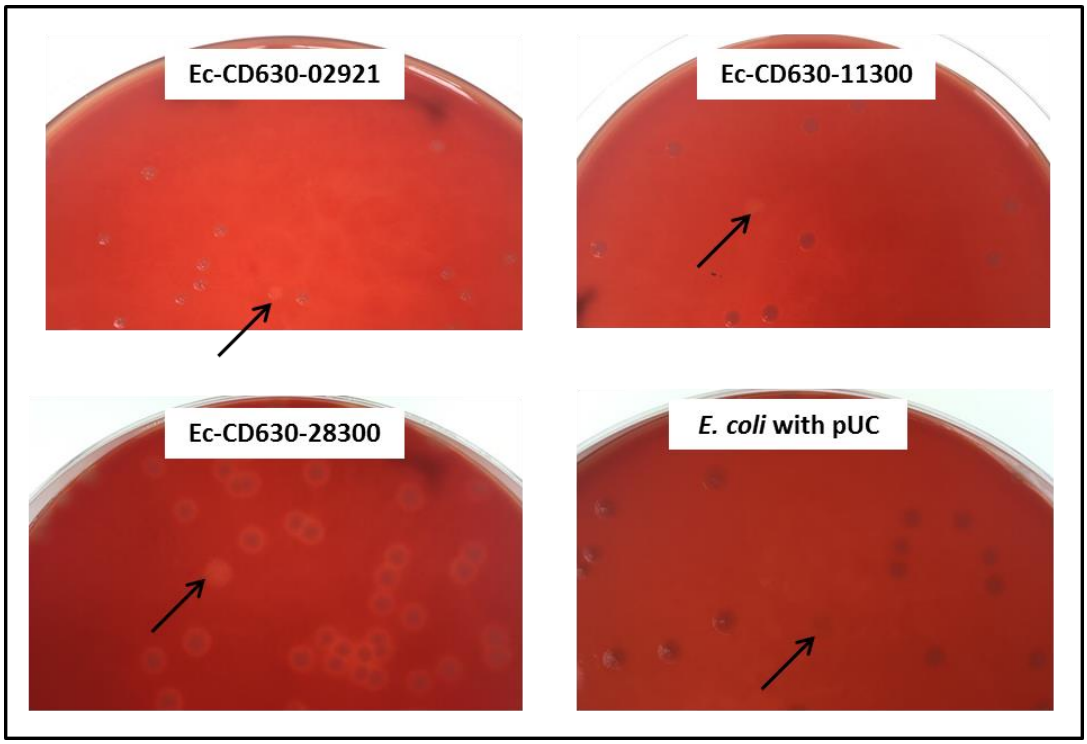


Figure 4-11 Haemolytic activity of *E. coli* clones containing single *C. difficile* genes.

Clones Ec-CD630-02921and Ec-CD630-11300 showed haemolysis under the colonies. Ec-CD630-28300 was very haemolytic compared to the others. The control *E. coli* carrying the empty plasmid was non-haemolytic. The black arrows points at the position where the colonies were growing before they were picked from the plate by a loop.

4.3.8 Analysis of a possible operon that contain genes encoding haemolytic activity

The three genes (CD630_11290, CD630_11300, and CD630_11310) that encoded haemolysis in clone Ec-pUC-Lib-4.3 were overlapping which suggested that they might be part of an operon. The promoter site was detected using BPROM (Softberry) and the transcriptional terminator with Arnold (NCBI). Many promoters were predicted within the sequence upstream of the gene encoding DNA polymerase I (CD630_11280). There was a predicted transcriptional terminator 340 bp downstream of gene CD630_11310 just after gene CD630_11320 which encodes a putative heavy-metal transport/detoxification protein (Figure 4-12). This analysis suggested that the three haemolytic genes were part of an operon and requires experimental verification.

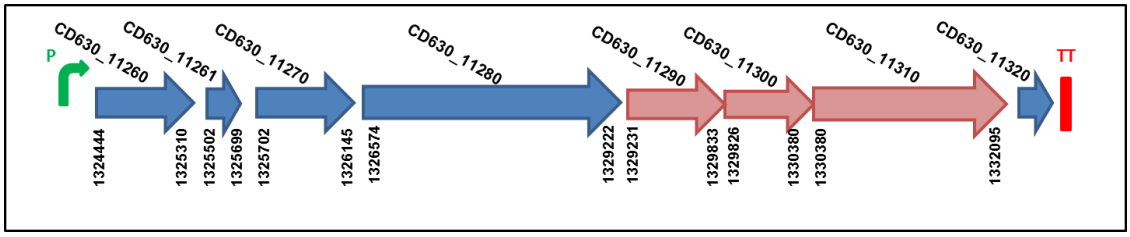


Figure 4-12 Schematic representation of a possible haemolytic operon in CD630.

The predicted promoter is represented as green arrow. Genes (CD630_11260 and CD630_11261) encoded transcriptional regulators AraC and HTH respectively. Genes (CD630_11270 and CD630_11280) encoded putative membrane protein and DNA polymerase I respectively. The genes (CD630_11290, CD630_11300, and CD630_11310) (red arrows) are haemolytic when cloned into *E. coli*. The gene downstream (CD630_11320) encoded a putative heavy-metal transport/ detoxification protein. The predicted transcriptional terminator was represented as red bar.

4.3.9 Developing a haemolysis assay for *E. coli*

To measure the haemolysis of the mutants more accurately, a quantitative assay was developed which was similar to the one used for *C. difficile* (section 3.2.6 and 3.3.6). The haemolytic clone (Ec-pUC-Lib-1.6) and its mutant with a deletion in gene CD630_02921 (Ec-pUC-Lib-1.6Δ02921) and the clone containing gene CD630_02921 (Ec-CD630-02921) were tested along with two non-haemolytic *E. coli* controls. The haemolysis OD540 measurements were plotted into a bar chart (Figure 4-13). The negative and positive controls were included in the assay to ensure its reliability but not added to the chart because they were not diluted in a similar way as the bacterial suspensions. It was surprising that the non-haemolytic controls (*E. coli* and *E. coli* carrying pUC19) showed haemolysis in the assay. There was no difference between the haemolysis of the *E. coli* strains even when the concentration of the cells were reduced. It is possible that the assay was too sensitive for *E. coli*. Therefore a new assay should be developed and optimised.

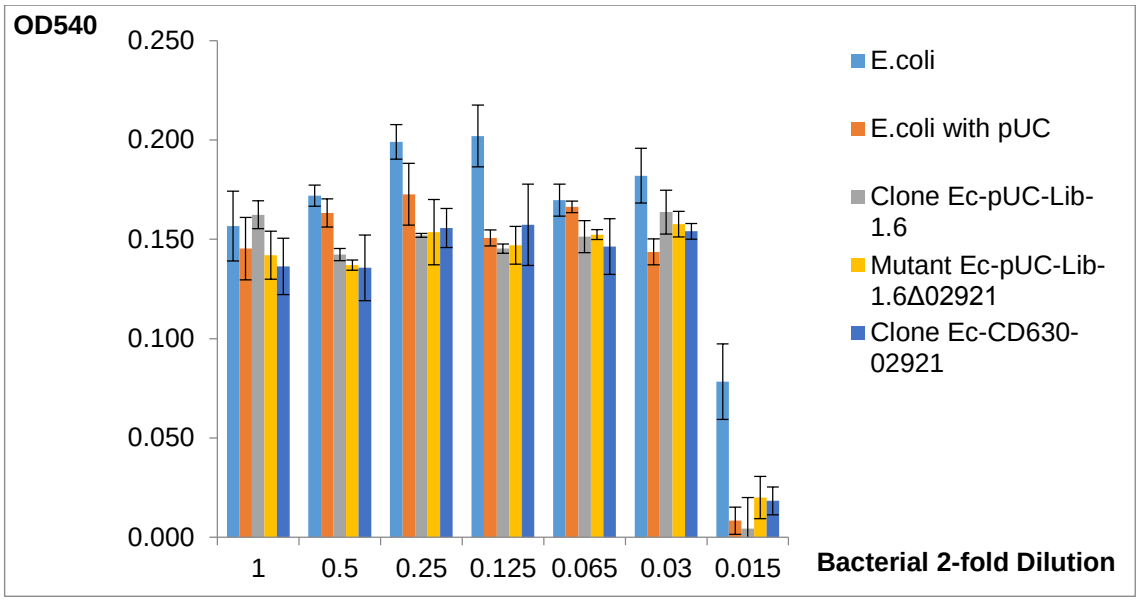


Figure 4-13 Haemolysis assay for *E. coli* clones and mutant.

The bars represent the OD540 of the supernatants after incubation of 1% RBCs suspension with two-fold serial dilution of *E. coli* starting from 1 (neat which is OD600 of about 17) to 0.015. *E. coli* and *E. coli* carrying an empty plasmid (pUC19) were used as a negative control in the assay. Two clones: one carrying insert one and the other carrying gene 2-1, and mutant of the first clone were tested. The negative and positive controls were included in the assay but not added to the chart because they were not serially diluted as with the *E. coli* suspensions. All the *E. coli* strains showed haemolysis even the non-haemolytic controls which render this assay unfit for testing.

4.4 Discussion

The CD630 library identified three haemolytic clones: Ec-pUC-Lib-1.6, Ec-pUC-Lib-4.3, and Ec-pUC-Lib-3.4 containing three, four, and three genes respectively (Table 4-1). Bioinformatic analysis has revealed that some of these genes encode proteins implicated in replication, catabolic process, or regulation of other genes. Other genes encode membrane proteins responsible for iron and other metals transport or binding. One gene (CD630_28300) was a possible haemolysin (Table 4-3). Bioinformatic analysis is a good tool to give an indication about the function of proteins but the data is predictive and requires experimental confirmation. The three genes that were chosen to be deleted by SDM were the ones present in full length and two of these were registered as uncharacterised in the GenBank database which mean that their function in *C. difficile* might be novel. *E. coli* strains Ec-pUC-Lib-4.3Δ11300 and Ec-pUC-Lib-3.4Δ28300, with a complete deletion of CD630_11300 and CD630_28300 genes respectively, were still haemolytic indicating that the deleted genes were not the ones causing haemolysis. Or that there is more than one gene in these inserts responsible for haemolysis. In other bacteria more than one gene is needed for haemolysis. *B. cereus* produces a haemolysin, Cerolysin AB, which is encoded by two genes. Individually *cerA* and *cerB* express two proteins with different enzymatic activity: phospholipase C and sphingomyelinase, respectively but they act synergistically to cause haemolysis. Clones carrying *cerA* or *cerB* only could not cause haemolysis (Gilmore *et al.*, 1989). Ec-

pUC-Lib-1.6Δ02921 produced mainly non-haemolytic colonies but few haemolytic colonies (1.9 %) still appeared on the agar.

Therefore isolating and cloning each of the potential haemolysins on their own was necessary to determine which genes encoded haemolysis. Six of which showed haemolysis either slight, under the colonies, or around them. All of these genes encode possible haemolysins and require further investigation. But in the next chapters the work focused on CD630_28300 gene (designated *atlf*), which encodes an uncharacterized protein that contained an anthrax toxin lethal factor (ATLF) domain, because it demonstrated the highest haemolytic activity when cloned into *E. coli*.

The *lef* in *B. anthracis* (GenBank: M29081.1) encodes an anthrax toxin lethal factor (LF, 809 residues) which is part of the tripartite anthrax toxin (Bragg & Robertson, 1989). The LF protein is predicted to have two ATLF domains (PF07737, residues 69-286 and 587-808) and Anthrax-tox_M domain (PF09156, residues 297-583). The protein also has a transmembrane region (residues 7-29) and a low complexity region (residues 45-63). The putative ATLF protein encoded by CD630_28300 gene has a 22% amino acid sequence homology to the C-terminus of LF from *B. anthracis* (Figure 4-14). Both proteins have a signal peptide at the N-terminus indicating that they are secreted, and the catalytic domain at the C-terminus. Although the protein sequence homology is low and the putative ATLF is much smaller in size than the LF from *B. anthracis,* both have similar general components and can cause haemolysis. *B. anthracis* was initially thought to be a non-haemolytic

bacterium as this was one of the features to distinguish it from other *Bacillus* spp. in the laboratory (Klichko *et al.*, 2003). Later it was demonstrated that it can cause haemolysis to human and not horse RBCs (Papaparaskevas *et al.*, 2004). The haemolytic activity of LF, EF and PA, individually and in combination, was tested to demonstrate their effect on human RBCs (Wu *et al.*, 2003). It was found that PA in combination with LF and EF can induce haemolysis *in vitro*. This confirms that LF needs PA to be transported into target cells in order to be active. These proteins are non-toxic *in vivo* when administered individually but LF in combination with PA can kill experimental rats, mice, and guinea pigs (Ezzell *et al.*, 1984). When *lef* was cloned along with *pa* into *E. coli*, the recombinant LF protein was biochemically active and showed the same level of lethality in the mouse macrophage assay compared to LF secreted by *B. anthracis* (Robertson & Leppla, 1986). These studies provide strong evidence that LF can be purified and used *in vitro* and *in vivo* to demonstrate its lethal activity. This suggests that ATLF produced by CD630 should be purified and tested directly on RBCs to determine its function.

The attempts to develop a haemolysis assay that can be used for *E. coli* have not been fruitful. It is postulated that the incubation period might have been too long for *E. coli* as most of the studies mention 1 to 6 hours of incubation with RBCs (Ishii, 1960; Short & Kurtz, 1971). The haemolysis assay for *E. coli* was not optimised as it was not the main focus of this work.

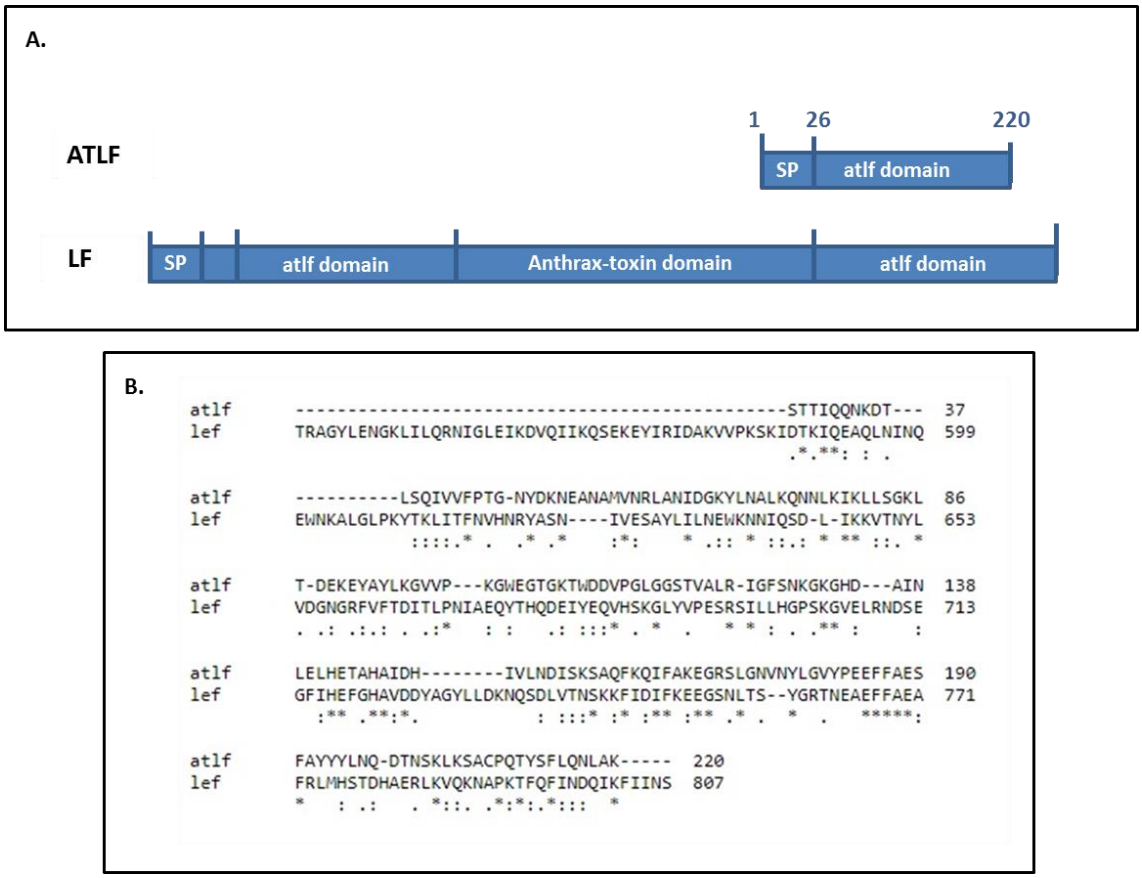


Figure 4-14 Amino acid homology comparison between ATLF and LF.

A. Schematic diagram showing similarity between the structure of ATLF from CD630 and the C- terminus of LF from *B. anthracis*. Figure is adapted from (Cafardi *et al.*, 2013). **B.** ATLF and LF amino acid sequences were aligned using Clustal Omega. The star (*); indicates identical amino acids, colon (:); indicates strongly similar amino acids (conservation), and dot (.); indicates weakly similar amino acids (semi-conservation). There was 22% homology between the two amino acid sequences of the proteins. SP; signal peptide, atlf; anthrax toxin lethal factor, lef; lethal factor.

4.5 Conclusions

CD630 has a six genes which, when cloned in *E. coli* lead to different levels of haemolysis ranging from slight, under the colonies, to large zones of haemolysis. Gene CD630_28300 encoding uncharacterised protein and containing ATLF domain that was homologous to the C-terminus of the lethal factor domain of the anthrax toxin produced by *B. anthracis* showed very large zones of haemolysis around the colonies when it was cloned in *E. coli*. Further investigation into this gene is required to understand its role in the haemolysis of CD630.

Chapter 5

Mutagenic analysis of ATLF-domain protein

Chapter 5 Mutagenic analysis of ATLF-domain protein

5.1 Introduction

Protein production and purification is widely used in biochemical, industrial, and pharmaceutical settings for various applications. In research laboratories it is mainly for studying protein structure and function. Many chromatography methods have been developed for the purification of proteins such as: gel filtration, ion exchange, affinity, and hydrophobic interaction chromatography (Asenjo & Andrews, 2009). In these chromatographic techniques proteins are usually purified according to their size, charge, binding affinity, or solubility. It is crucial to obtain protein samples with high yield, high purity, and high activity (HHH-grade) specifically if their crystal structure (3D image) is required (Vassylyeva *et al.*, 2017).

One simple, inexpensive method to purify proteins is affinity chromatography. Fusion tags used in this method are attached to recombinant proteins and have high affinities for a specific chemical or biological ligand usually immobilized on a chromatography column, hence permitting its purification (Saraswat *et al.*, 2013). Poly-histidine, or His, tags are frequently used for academic and industrial purposes where six or more histidine residues are fused to the N- or C- terminus of the target protein and sometimes both ends (Seidler, 1994). The His-tag can form stable complexes with metal ions such as nickel (Ni), zinc (Zn), manganese (Mn), or copper (Cu) (Hage *et al.*, 2017). These metals can be immobilized by chelation with nitrilotriacetic acid (NTA) or iminodiacetic acid (IDA) bound to a solid surface such as agarose gel or

silica particles (Figure 5-1) (el Rassi & Horváth, 1986; Hochuli *et al.*, 1987). His-tagged-protein-metal-ion complex will bind to the affinity ligand (metal-ion chelator) while other proteins in the mixture will pass through the chromatography column during the washing steps (Figure 5-2) (Spriestersbach *et al.*, 2015). The target protein can then be detached from the metal binding by the addition of imidazole or a chelating agent like EDTA (Ethylenediaminetetraacetic Acid), or by lowering the pH (pH 4.5- 5.3) (Bornhorst & Falke, 2000). This chromatography purification is known as immobilized metal affinity chromatography (IMAC) (Porath *et al.*, 1975; Smith *et al.*, 1988). IMAC is a very efficient purification method for His-tagged recombinant proteins with 90% yield and up to 90% purity in a single step (Hochuli *et al.*, 1988) .

As discussed previously in chapter 4, *atlf* (CD630_28300) encoded a protein that showed the largest zones of haemolysis on blood agar and hence considered a possible haemolysin in CD630. This chapter describes the mutagenic analysis of the *atlf* and the production and purification of recombinant ATLF protein (rATLF). The *atlf* was cloned into an expression plasmid and the recombinant protein with 6x His-tag was purified using Nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography.

Figure 5-1 Chelating agents used for immobilised metal affinity chromatography.

Nitrilotriacetic acid (NTA) or iminodiacetic acid (IDA) are fixed on a stationary surface and can bind to protein-metal-ion complexes. Figure is taken with permission from (Hage *et al.*, 2017).

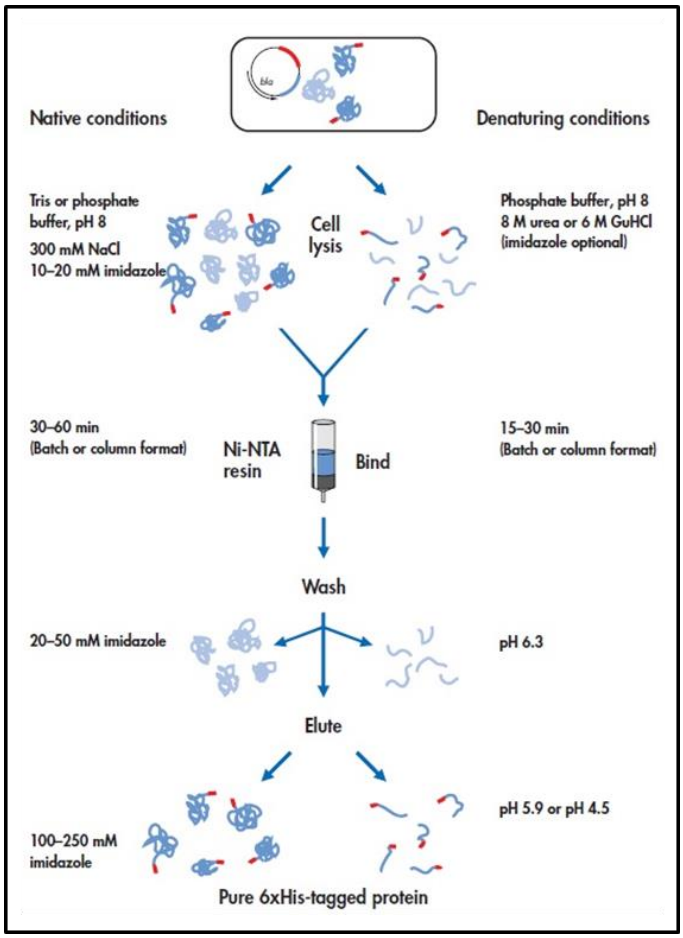


Figure 5-2 Ni-NTA affinity chromatography for His-tagged protein purification.

An illustration of the steps used in this type of chromatography in native and denaturing conditions. Figure is taken from the handbook for high-level expression and purification of 6xHis-tagged proteins by Qiagen.

5.2 Materials and methods

5.2.1 Knock out *E. coli* clone carrying *atlf* by SDM

The SDM was performed using Q5 Site-Directed Mutagenesis Kit (NEB, UK). The primers were designed (NEBaseChanger.neb.com) to change the ATG start codon to GCG. This mutation was created by designing forward primer with a mismatch (GCG instead of ATG) in the centre of it (ZK26-F and ZK27-F). The reverse primer was designed in the opposite direction on the complementary strand where its 5´ end began at the nucleotide next to the 5´ end of the forward primer (ZK26-R and ZK27-R). The steps following the primers design in the SDM protocol was described in detail in Chapter 2 section 2.5.

5.2.2 His-tag purification of ATLF

The ATLF was fused to 6x His-tags and purified using Ni-NTA affinity chromatography. The protocol used for cloning and expression of the *atlf* was based on the Novagen pET System Manual (11th Edition) (http://kirschner.med.harvard.edu/files/protocols/Novagen_petsystem. pdf). For the purification procedure, the handbook for high-level expression and purification of 6xHis-tagged proteins by Qiagen was followed (file:///C:/Users/user /Downloads/EN-The-QIAexpressionist.pdf).

5.2.2.1 Cloning of *atlf* in pET28a expression plasmid

The *atlf* was amplified by PCR utilising CD630 genomic DNA and primer pairs ZK28-F with ZK28-R and ZK29-F with ZK29-R to construct pZK-ATLF

and pZK-ATLFΔSP respectively. The purified amplicons and pET28a were digested using *EcoR*I with *Hind*III and *BamH*I with *Xho*I and ligated to produce plasmids pZK-ATLF and pZK-ATLFΔSP respectively. The digested inserts and plasmids were purified before ligation. The ligation mix was transformed into α-select silver *E. coli* competent cells and plated on LB agar supplemented with 30 µg/ml Km. Plasmid DNA was extracted from colonies and digested with the same restriction enzymes to check for the presence of the 662 bp and 497 bp inserts. The correct sized clones were further analysed by Sanger sequencing. The recombinant plasmid was then transformed into competent expression host BL21 (DE3) using a routine transformation protocol (section 2.2.11) to create Ec-ATLF and Ec-ATLFΔSP.

5.2.2.2 IPTG induction

After the recombinant plasmid was established in BL21 cells, a single colony was inoculated into 10 ml LB broth supplemented with 30 µg/ml Km and incubated at 37°C for 18 hours. Two controls were used in the induction protocol: BL21 cells and BL21 containing empty pET28a plasmid. Both were cultured the same way except that BL21 cells were cultured without kanamycin. Next, 2 ml of the overnight culture (18 hours) was used to inoculate 50 ml LB supplemented with 30 µg/ml Km in 250 ml Erlenmeyer flask. The culture was incubated at 37°C for 1-2 hours until the cells reached exponential phase (OD600 of 0.6-1). Expression was induced by adding 1 mM IPTG to the culture and incubating it at 37°C for up to 24 hours. The induction protocol was optimised by testing the best incubation time for the expression as 1 ml sample was collected at 1, 2, 3, 4, and 24 hour intervals

after IPTG addition. The cells collected in Eppendorf tubes were harvested by centrifugation at 2348 x *g* for 10 minutes (5000 rpm in Eppendorf centrifuge 5424). The pellets were stored at -20°C or used immediately in the purification step (section 5.2.2.3). For large scale expression, 500 ml of cell culture was prepared for protein purification (section 5.2.2.4).

5.2.2.3 Purification from test expression experiments

The purification of 1 ml test samples of the IPTG induction was done using protocol 14 of the Qiagen purification manual. All the purification procedures were performed on ice or in the cold room (4°C). The pallet was resuspended in the lysis buffer (B-PER) containing ammonium chloride then gently rotated for 30 minutes. The supernatant was then separated from cell debris by cold centrifugation at 18,000 x *g* for 5 minutes (13,000 rpm in Eppendorf centrifuge 5417). The supernatant was then added to a fresh tube containing Ni-NTA agarose and gently rotated for 60 minutes. The tubes were centrifuged as previously mentioned for 1 minute to discard the supernatant. The pallets were washed three times with the washing buffer, by centrifugation for 1 minute in between washes and carefully removing the supernatant. The protein was eluted by adding the elution buffer (twice) and collecting the supernatant in a fresh tube using centrifugation at 1 minute. The concentrations and volumes of the reagents used are listed in Table 5-1.

5.2.2.4 Purification of large scale protein expression

The purification of large scale protein expression was done by following protocol 12 of the Qiagen purification manual with one modification where the washing step was performed up to 9 times. All the purification procedures were performed on ice or in the cold room (4°C). The pallet of the 500 ml of bacterial culture was lysed (as in section 5.2.2.3) and the supernatant was collected after cold centrifugation at 10,000 x *g* for 20 minutes (15,000 rpm in Sorvall RC-5B Plus). The Ni-NTA agarose was added to the supernatant and gently rotated for 60 minutes. The mixture was loaded into a 5 ml polypropylene column (Qiagen) and the flow-through was collected in a fresh tube. The column was washed 5 to 9 times with the washing buffer and the wash fractions were collected in a fresh tube. The protein was eluted by adding the elution buffer (4 times) and collecting the eluate in a fresh tube. The concentrations and volumes of the reagents used are listed in Table 5-1. The purified protein was aliquoted and stored in 20% glycerol at -20°C.

Table 5-1 List of the reagents used in the purification of 6xHis-tagged proteins.

Reagent	Test expression	Large scale expression	Source
B-PER™ Bacterial Protein Extraction Reagent	100 µl	12 ml	ThermoFisher Scientific
Ammonium Chloride	750 mM	750 mM	Sigma
Ni-NTA agarose 50% slurry	20 µl	3 : 1 ratio supernatant : Ni-NTA	Qiagen
Wash buffer pH 8.0 100 mM NaH_2PO_4 20 mM Imidazole	100 µl	4 ml	BDH chemicals
Elution buffer pH 8.0 100 mM NaH_2PO_4 250 mM Imidazole	30 µl	0.5 ml	BDH chemicals

5.2.2.5 Calculating the molecular weight and concentration of the protein

The molecular weight of the protein was calculated using online protein molecular weight calculator (www.bioinformatics.org/sms/prot_mw.html). The concentration was estimated according to Beer-Lambert Law by the following equation (Grimsley & Pace, 2003):

$$\textit{Protein concentration (mg/ml)} = \frac{protein\ Absorption\ A280}{Absorbance\ of\ extinction\ coefficient}$$

Protein extinction coefficient and 0.1% absorbance was calculated using an online tool called ProtParam (https://web.expasy.org/protparam/). Absorption at a wavelength of 280nm was measured using ND-1000 spectrophotometer (Labtech UK). The final protein concentration was converted to per litre of the culture rather than per millilitre as described in literature.

5.2.3 SDS-PAGE

The purified proteins were analysed by discontinuous Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE). The gel was prepared using protocol described previously (Laemmli, 1970). Samples from each stage of the purification and eluted proteins (15 µl) were mixed with Pierce™ lane marker non-reducing sample buffer (5 µl) (ThermoFisher Scientific, UK). They were visualized on a 12% SDS gel running in 1x TRIS/GLYCINE/SDS buffer for 2-3 hours at 100V. The bands were compared to 11-190 kDa blue prestained protein standard or 10-200 kDa unstained protein ladder (NEB, UK). The gel was stained with Coomassie Brilliant Blue R-250 dye (Bio-Rad, UK) then treated with a de-staining solution (40% (v/v) Methanol and 10% (v/v) Acetic Acid) (VWR, UK). The reagents used are listed in Table 5-2.

5.2.4 Buffer exchange of the protein using dialysis

The purified protein was dialyzed using Slide-A-Lyzer™ Dialysis Cassette (10K MWCO, 3 mL) (ThermoFisher Scientific, UK) according to manufacturer's instructions. The dialysis was performed using 1.5 litres of PBS (137 mM NaCL, 3 mM KCL, 8mM Na_2HPO_4, and 15 mM KH_2PO_4, pH 7.3) in 4°C cold room using lowest stirring speed setting. The buffer was changed three times a day every hour for three successive days. The protein was then collected and visualised using SDS-PAGE.

Table 5-2 List of the reagents used in the SDS-PAGE.

Reagent	Concentration	Source
Ammonium Persulfate (APS)	10%	Sigma, UK
SDS solution	0.1%	ThermoFisher Scientific
<u>Resolving Gel</u> Protogel Protogel Buffer Deionized water APS TEMED	 4 ml 2.5 ml 3.4 ml 100 µl 10 µl	 National Diagnostics National Diagnostics - Sigma, UK National Diagnostics
<u>Stacking Gel</u> Protogel Protogel Stacking Buffer Deionized water APS TEMED	 1.3 ml 2.5 ml 6.1 ml 50 µl 10 µl	 National Diagnostics National Diagnostics - Sigma, UK National Diagnostics
SDS-PAGE running buffer	10x TRIS/GLYCINE/SDS 0.25 M TRIS 1.92 M Glycine 1% SDS	National Diagnostics

5.2.5 Haemolysis assay with the purified protein

The activity of the protein was tested quantitatively using an assay similar to the one developed for *C. difficile* in section (3.2.6) with the following modifications. The assay plate was incubated aerobically at 37°C for 16 hours. Saponin solution (0.05%) (Sigma-Aldrich) was used as a positive control. Also 100 µl of the purified proteins were mixed with several dilutions of defibrinated whole horse blood (E &O Laboratories).

5.3 Results

5.3.1 Mutagenic analysis of Ec-CD630-28300

To confirm that the expression of *atlf* in *E. coli* was the reason this clone showed high levels of haemolysis, it was required to mutate it. The mutation of the clone carrying *atlf* (Ec-CD630-28300) was created by SDM to change the start codon (ATG) into GCG, encoding an alanine. Upon analysing the sequence of the *atlf*, another ATG in the same reading frame was found 166 bp downstream of the start codon preceded by a sequence that resembles a ribosome binding site (RBS) (Figure 5-3A). The second ATG was upstream of the region encoding the catalytic domain, which makes it possible that knocking out the first start codon will not affect the expression (Figure 5-3B). To test this theory three mutants were created: one with a mutation in the first putative start codon, the second mutation in both putative start codons, and the third with a mutation in the second putative start codon.

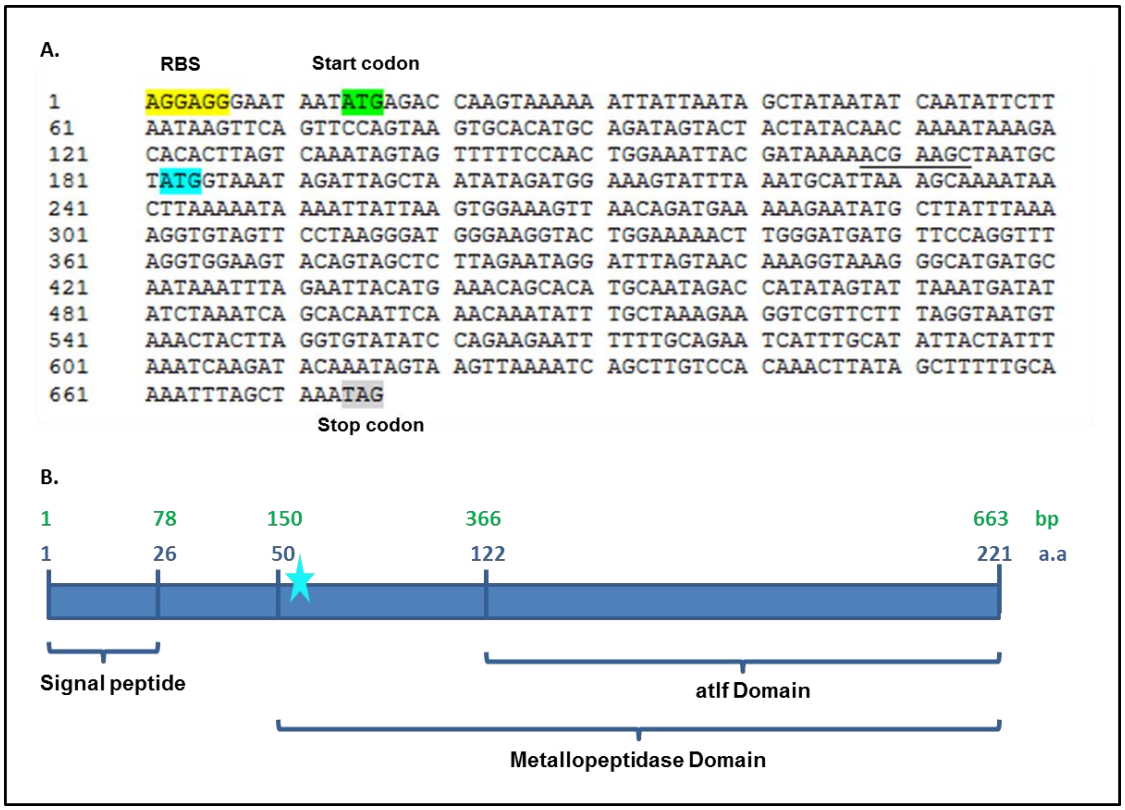


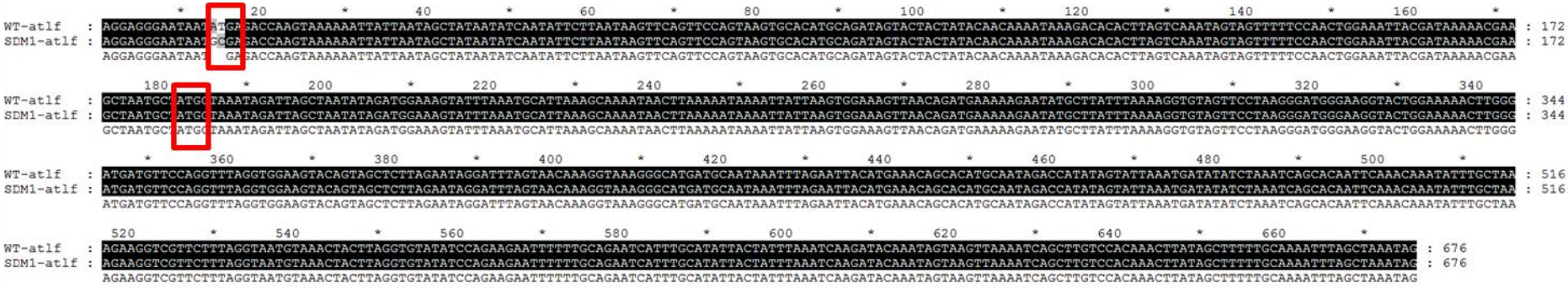
Figure 5-3 Analysis of *atlf* (CD630_28300).

A. Nucleotide sequence of *atlf*. The gene is 663 bp in size with the first putative ATG start codon and TAG stop codon highlighted in green and grey respectively. The first ribosome binding site is highlighted in yellow. Within the same reading frame of this gene, there was a second putative start codon (highlighted in blue) preceded by a putative ribosome binding site (ACGAAGC underlined). **B. Schematic representation of the ATLF.** The protein product of *atlf* and the locations of the functional domains in addition to the signal peptide are shown. The position of the second ATG is shown as a star sign.

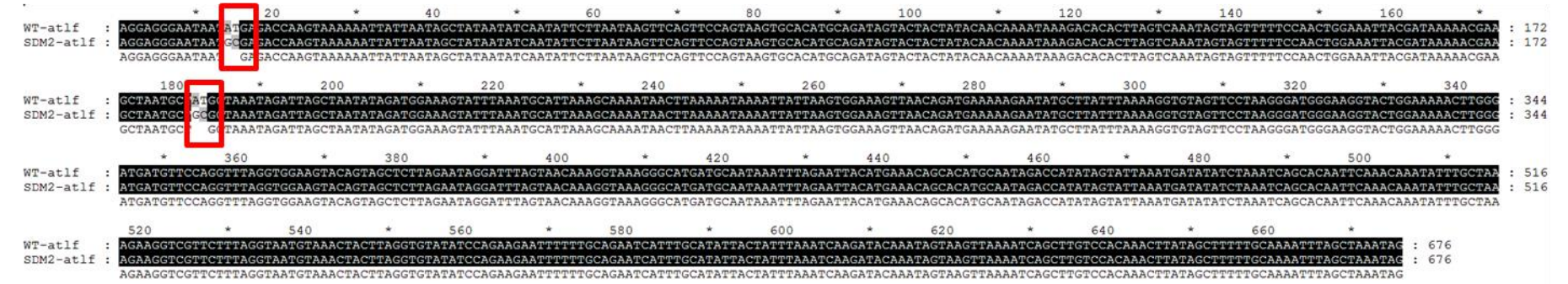
5.3.1.1 Constructing SDM mutants with substitution in ATG

Site-directed mutagenesis was used to change the first start codon (ATG) to GCG (encoding alanine) in Ec-CD630-28300 to create *E. coli* mutant Ec-CD630-28300-S1 (Figure 5-4A). Then a mutant with mutation in both putative start codons Ec-CD630-28300-S1-S166 was constructed (Figure 5-4B). A third mutant Ec-CD630-28300-S166 was created with the second putative ATG (at nucleotide 166) changed to GCG (Figure 5-4C). A fourth mutant was created by spontaneous mutation with a 58 bp deletion around the second ATG (Figure 5-4D).

A.



B.



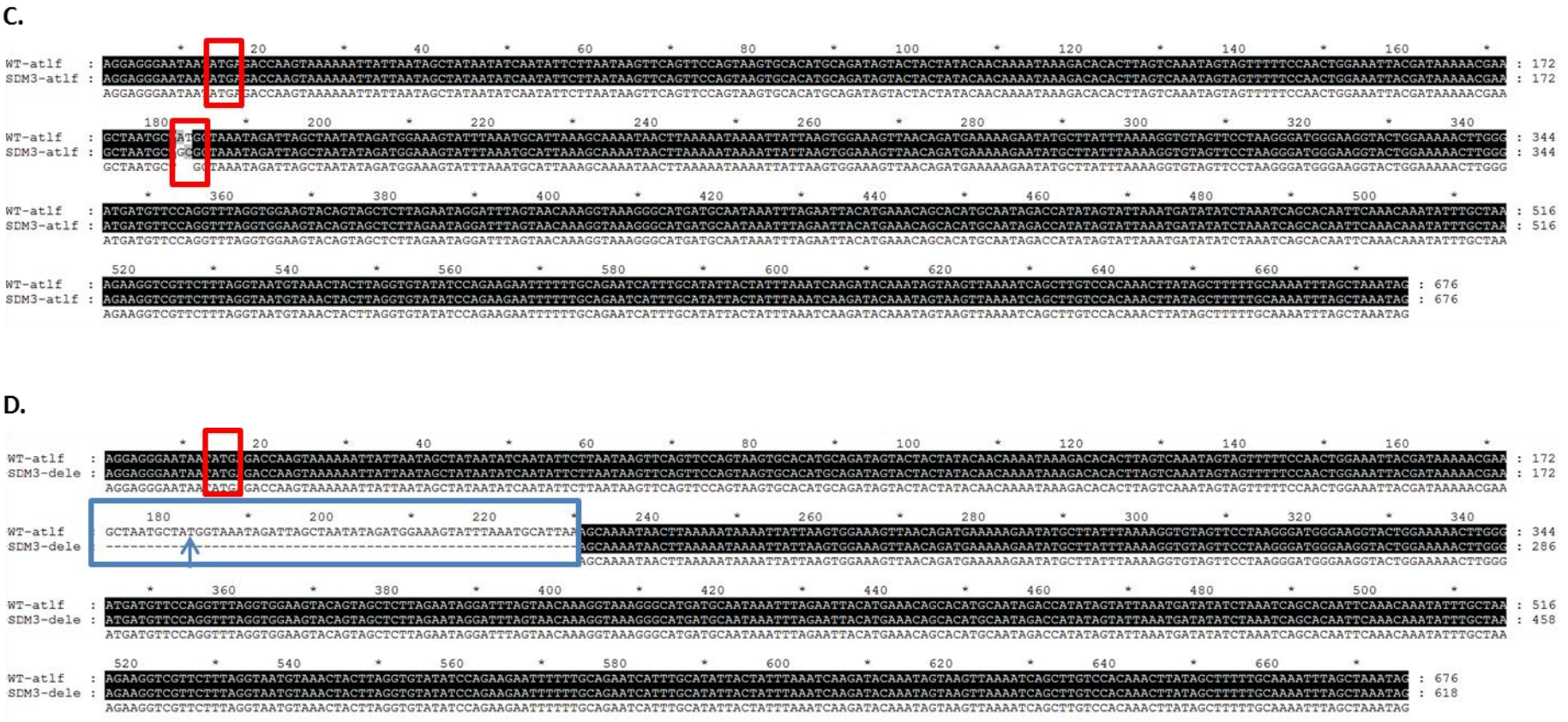


Figure 5-4 Alignment of the sequence of two SDM mutants with the wild type.

The sequence of mutants Ec-CD630-28300-S1 (A.) and Ec-CD630-28300-S1-S166 (B.) Ec-CD630-28300-S166 (C.) and Ec-CD630-28300Δ58 (D.) were aligned with the wild type clone Ec-CD630-28300. The position of start codon (ATG) is represented by the red boxes. The 58 bp deletion is represented by the blue box and the position of the second ATG is indicated by the blue arrow. The alignment was performed using GeneDoc.

5.3.1.2 Testing the haemolysis of the SDM mutants

Ec-CD630-28300 clone and Ec-CD630-28300-S1 mutant were haemolytic whereas Ec-CD630-28300-S1-S166 (double mutant), Ec-CD630-28300-S166 and Ec-CD630-28300Δ58 were non-haemolytic (Table 5-3) (Figure 5-5). This demonstrated that knocking out the first start codon did not affect haemolysis. But rather showed that the second ATG was responsible for expressing *atlf*, or at least the part responsible for haemolysis, demonstrated by loss of haemolysis when the two ATGs were knocked out. It was further confirmed as Ec-CD630-28300-S166 and Ec-CD630-28300Δ58 were non-haemolytic. This provides evidence that the phenotype appears dependant on the expression from the second start codon (without the signal peptide present at 1-78 bp).

Table 5-3 Summary of the haemolytic activity of the SDM mutants created in this Chapter.

Bacteria	Properties/mutations	Haemolytic activity
Ec-CD630-28300	Wild type	Haemolytic
Ec-CD630-28300-S1	First ATG	Haemolytic
Ec-CD630-28300-S1-S166	Both ATGs	Non-haemolytic
Ec-CD630-28300-S166	Second ATG	Non-haemolytic
Ec-CD630-28300Δ58	58 bp deletion including second ATG	Non-haemolytic

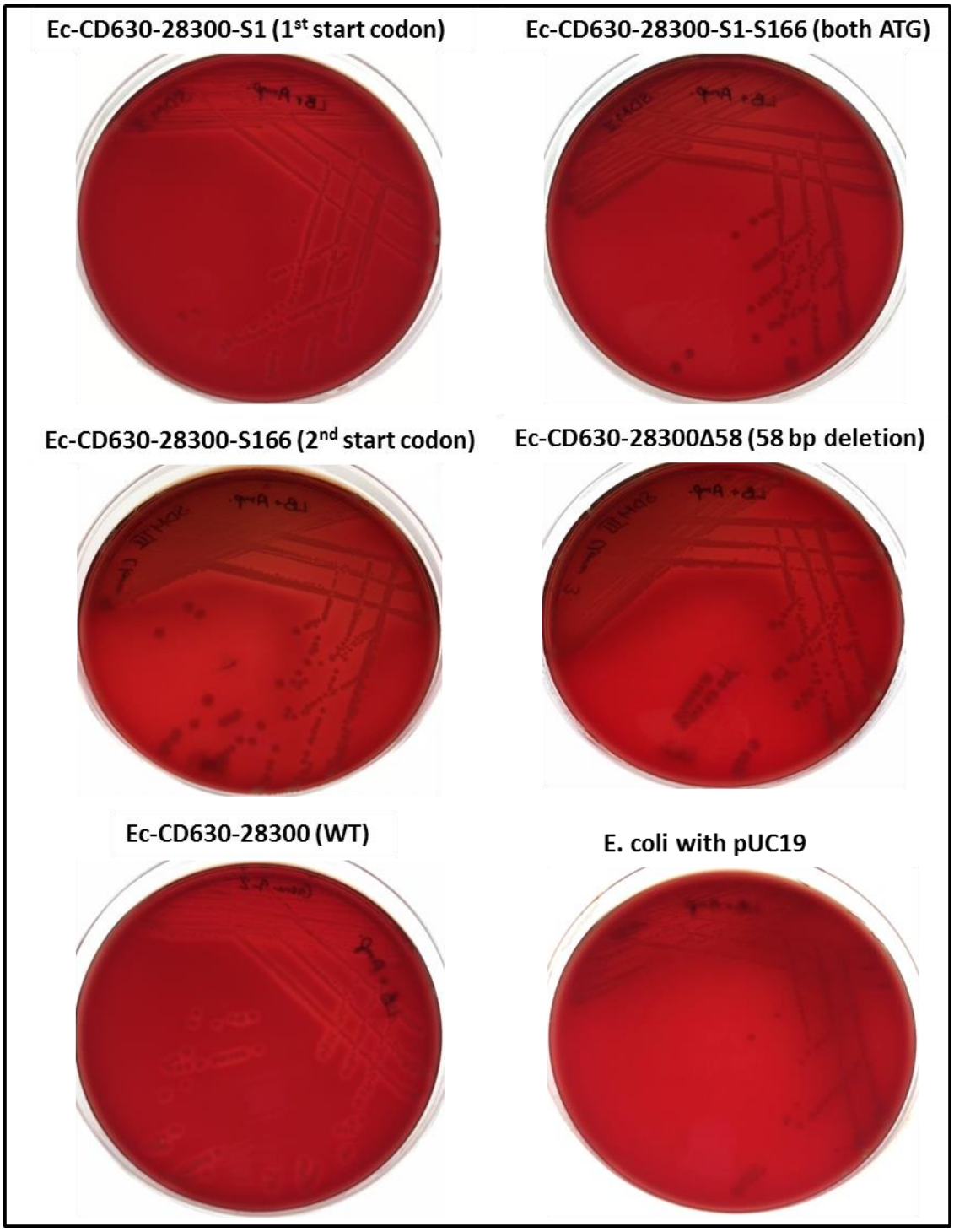


Figure 5-5 Haemolysis of the *E. coli* mutants in which *atlf* was knocked out by SDM.

Mutant Ec-CD630-28300-S1 with substitution in the first start codon showed haemolysis similar to the wild type clone Ec-CD630-28300. Mutant Ec-CD630-28300-S1-S166 with substitution in both ATGs showed non-haemolytic phenotype similar to the control strain, *E. coli* carrying the empty plasmid. Mutant Ec-CD630-28300-S166 with substitution in the ATG in the second start codon was non-haemolytic. Also mutant Ec-CD630-28300Δ58 was non-haemolytic similar to the control strain.

5.3.2 Purification of the proteins expressed by *atlf*

Two recombinant proteins were intended to be purified: rATLF that is expressed from the first start codon of *atlf* including the predicted signal peptide (1-78 bp), and rATLFΔSP that is expressed from the second ATG at position 166 bp of *atlf* which mean that the purified protein does not have a signal peptide. The proteins were purified using Ni-NTA affinity chromatography.

5.3.2.1 Purification of recombinant ATLF

The predicted size of the rATLF was determined as 24.32 kDa. The different purified proteins from each induction interval were visualised by SDS-PAGE (Figure 5-6). The induction time was set at 2 hours. The purification yielded 190 mg of rATLF per litre of culture. The purity of the recombinant protein was evaluated by SDS-PAGE (Figure 5-7). Impurities were detected on the gel although the protein was washed up to 9 times before being eluted.

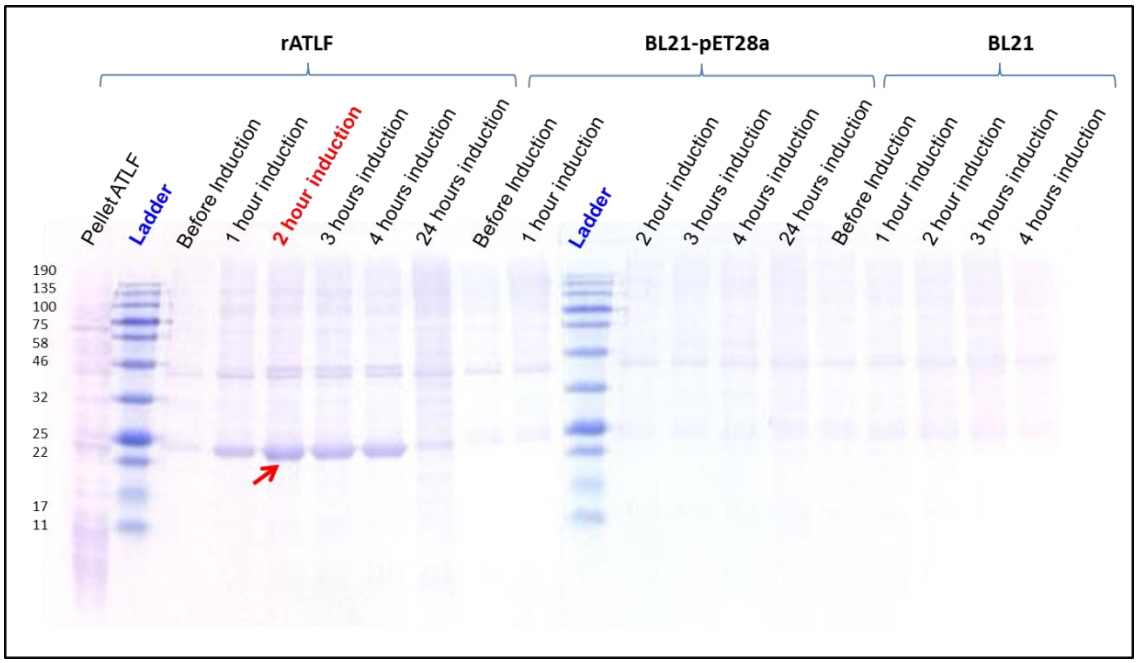


Figure 5-6 SDS-PAGE image of ATLF induction.

The induction of rATLF was tested at 1, 2, 3, 4, and 24 hours of addition of 1mM IPTG to a growing culture. Also two controls were tested the same way: BL21 and BL21-pET28a. The red arrow points at the purified protein that is 24.32 kDa in size. A 15 µl of the sample was loaded into the gel.

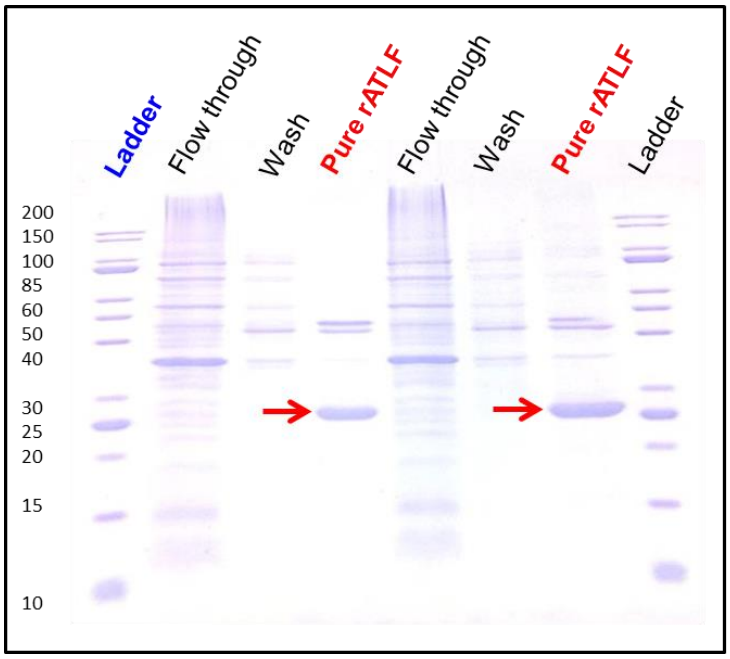


Figure 5-7 SDS-PAGE image of recombinant ATLF.

The rATLF was purified using Ni-NTA affinity chromatography of 6xHis-tagged proteins. The red arrows point at the purified protein that is 24.32 kDa in size. A 15 µl of the sample was loaded into the gel.

5.3.2.2 Purification of recombinant ATLFΔSP

The predicted size of the rATLFΔSP protein was determined as 18.25 kDa. The purification yielded 60 mg of rATLFΔSP protein per litre of culture. The purity of the purified rATLFΔSP protein was visualized by SDS-PAGE (Figure 5-8). The rATLFΔSP was not expressed as well as rATLF as demonstrated by the faint bands on the gel. Also impurities were seen on the SDS-PAGE.

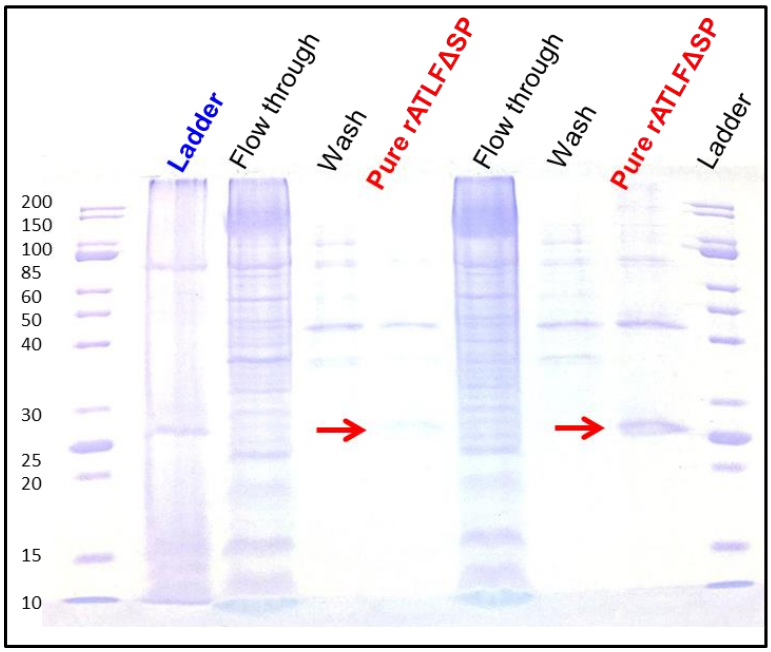


Figure 5-8 SDS-PAGE image of recombinant ATLFΔSP protein purification.

The rATLFΔSP protein was purified using Ni-NTA affinity chromatography of 6xHis-tagged proteins. The red arrows point at the purified protein that should be 18.25 kDa in size. A 15 µl of the sample was loaded into the gel.

5.3.3 Haemolysis assay of purified proteins in elution buffer

The purified rATLF and rATLFΔSP proteins were used in the haemolysis assay to directly test their activity against RBCs. The assay included a negative control (PBS with 2% glucose), positive control (1% RBCs in water), 1% saponin as an indication of 50% haemolysis, and elution buffer (same buffer the proteins are eluted in). It was observed that the elution buffer caused haemolysis (Figure 5-9A). Therefore the proteins cannot be assayed in this buffer. The NaH_2PO_4 buffer (100 mM at pH 8.0) was tested to determine if it can be used as an exchange buffer for the proteins. It was found to lyse RBCs but at a level less than the elution buffer (Figure 5-9B) whereas PBS did not show any haemolysis. Therefore it was concluded to change the buffer of the proteins to PBS (pH 7.3).

5.3.4 Haemolysis assay for proteins eluted in PBS

The purified recombinant proteins were subjected to dialysis to exchange the elution buffer with PBS. The activity of the proteins was tested in the haemolysis assay but no haemolysis was detected (Figure 5-10). Also there was no activity detected when whole blood was used instead of 1% RBCs suspension.

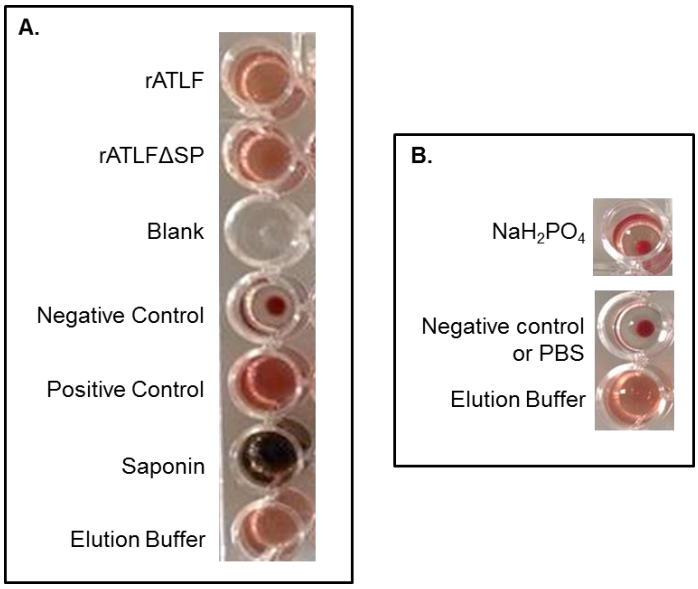


Figure 5-9 Haemolysis assay of rATLF and rATLFΔSP purified proteins in elution buffer.

A. 100 µl of rATLF and rATLFΔSP proteins were added to equal amount of 1% RBCs suspension to test their activity. Blank; PBS with 2% glucose, Negative control; 1% RBCs suspended in PBS supplemented with 2% glucose and 0.5% BSA, Positive control; 1% lysed RBCs in sterile distilled water, Saponin; 1% used as positive control for 50% haemolysis, Elution buffer; 100 mM NaH_2PO_4 and 250 mM Imidazole. The elution buffer caused haemolysis which meant that it could not be used as a buffer for the proteins. **B.** The NaH_2PO_4 buffer was tested on 1% RBCs suspension and was found to cause slight haemolysis in comparison to PBS.

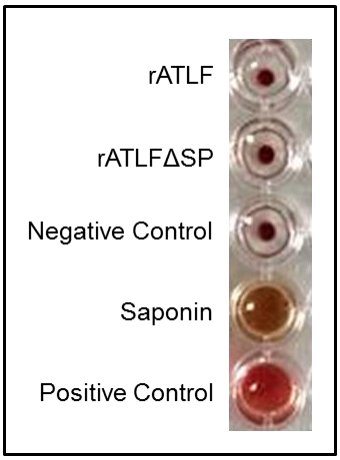


Figure 5-10 Haemolysis assay of rATLF and rATLFΔSP purified proteins in PBS.

The proteins showed no haemolytic activity when 1% RBCs suspension was added to 100 µl of each. rATLF; protein encoded by *atlf* (CD630_28300), rATLFΔSP; protein encoded by *atlf* without the signal peptide, Negative control; 1% RBCs suspended in PBS supplemented with 2% glucose and 0.5% BSA, Saponin; 1% used as positive control for 50% haemolysis, Positive control; 1% lysed RBCs in sterile distilled water.

5.4 Discussion

Blast analysis revealed that the *atlf* is conserved among sequenced *C. difficile* strains with an amino acid sequence identity of 98-100 % (Cafardi *et al.*, 2013). The presence of *atlf* was also confirmed experimentally by PCR in 30 *C. difficile* strains belonging to the main six lineages (Hensbergen *et al.*, 2014). This suggests that it is an important component of the genome of *C. difficile.* In addition, *atlf* encodes large zones of haemolysis when cloned in *E. coli* which make it worth investigating.

Recombinant ATLF and ATLFΔSP proteins were expressed in *E. coli* and then purified by the His-tag IMAC method. The ATLF was purified by others without the signal peptide region with N-terminus six or ten His-tag using the same method (Cafardi *et al.*, 2013; Hensbergen *et al.*, 2014; Schacherl *et al.*, 2015). It was noticeable that the full length rATLF was expressed at a higher level than the rATLFΔSP without the signal peptide (Figure 5-7 and Figure 5-8). Upon testing the rATLF and rATLFΔSP in the haemolysis assay, there was no haemolytic activity detected.

Both proteins had 6x His-tag at the N- and C-terminus which were not removed. The N-terminus His-tag can be cleaved by using a protease that hydrolyses thrombin (Waugh, 2011). Generally affinity tags do not affect structure or interfere with the activity of proteins due to their small size and charge (Carson *et al.*, 2007; Terpe, 2003). Although rare interference of His-tag in protein function has been reported (Wu & Filutowicz, 1999). However,

the N-terminal 10x His-tag used for the previous purification of ATLF did not affect its activity (Hensbergen *et al.*, 2014).

The concentration of the rATLF (190 mg/l) was much higher compared to the LF (5 mg/l) and previously purified ATLF (50 mg/l) (Liu *et al.*, 2013; Schacherl *et al.*, 2015). This variation in protein concentration between this work and others might be explained as impurities. There were two bands (45-50 kDa) visible on the SDS gel of the purified rATLF even after washing the chromatography column 9 times (Figure 5-7). Also the SDS gel of rATLFΔSP protein showed multiple bands and a faint band for the protein itself (Figure 5-8). Proteins which are rich in histidine and cysteine or naturally metal binding are able to compete, for nickel binding in the column, with the tagged protein and eventually co-elute and contaminate the final sample (Saraswat *et al.*, 2013). Contamination might also occur due to disulfide bond formation or non-specific hydrophobic interactions between the tagged protein and other proteins (Bornhorst & Falke, 2000). Some applications can help remove impurities and enhance the yield of proteins like the use of low concentrations of EDTA in the washing buffer but none were used here (Arnau *et al.*, 2006; Westra *et al.*, 2001). The purification of *B. anthracis* recombinant LF protein in *E. coli* was contaminated by a molecular chaperon (only binds to unfolded protein to stabilize it) which was removed by the addition of a dissociation buffer (Kim *et al.*, 2003). The purity of any protein is crucial when testing its activity as impurities might give misleading results (Raynal *et al.*, 2014). For example they act as a competitive inhibitor by blocking the active site of the purified protein. When the contaminants are

structurally similar to the substrate of the target protein, they might irreversibly bind to it preventing it from functioning (He *et al.*, 2005; Robinson, 2015).

Also the purified proteins might not have folded correctly during expression and purification which will alter the structure and function of the proteins. Denaturing conditions like low pH, temperature, high osmolarity, and oxidative stress during purification protocols might change the native conformation of the protein (Gasser *et al.*, 2008). Many proteins can refold to their native functional state when the denaturing factor is eliminated (Lodish *et al.*, 2000).

Another reason for failure to detect haemolysis might be the concentration of the proteins added in the haemolysis assay. The proteins were not concentrated but rather used immediately after purification. It was reported that long incubation and high concentration of the protein is needed for the proteolytic activity (Hensbergen *et al.*, 2014). Moreover, the rATLF and rATLFΔSP proteins were tested against whole blood but there was no haemolysis observed. The LF protein from *B. anthracis* can haemolyse whole blood only in the presence of the PA protein which is responsible for translocating the LF through the target cell membrane (Wu *et al.*, 2003) .

The ATLF has proteolytic activity on human plasma fibronectin and fibrinogen in the presence of Zinc (Zn^{2+}). *E. coli* mutants with one amino acid substitution in the catalytic domain created by SDM showed loss of the activity of the protein (Cafardi *et al.*, 2013). The activity of the ATLF and

ATLFΔSP proteins in this work was not tested by a protease activity assay to determine whether they are active or not before testing them in the haemolysis assay (Cupp-Enyard, 2008). Also Zn^{2+} was not added in the haemolysis assay when the proteins were tested. Future work should focus on producing proteins without impurities and to test their proteolytic activity before using them in the haemolysis assay. Also optimizing of the haemolysis assay for proteins is very crucial.

The specific cleavage site for the ATLF was found to be between two proline residues in a proline-rich motif (Schacherl *et al.*, 2015). The protein hydrolyses heat shock protein (HSP90β), immunoglobulin A (IgA2), and two putative adhesins encoded by CD630_28310 and CD630_32460 (Cafardi *et al.*, 2013; Hensbergen *et al.*, 2014). While ATLF demonstrates proteolytic activity for human plasma and native *C. difficile* proteins, it is possible that it has additional substrates that might be on the surface of erythrocytes. This prompted the construction of a *C. difficile* mutant with inactive *atlf* to be tested in the haemolysis assay which is discussed in the next chapter.

5.5 Conclusions

The *atlf* can be expressed in *E. coli* without the sequence encoding the putative signal peptide. The purified recombinant ATLF and ATLFΔSP proteins did not show any haemolytic activity towards erythrocytes in the conditions tested.

Chapter 6

Mutagenesis of CD630 using CRISPR-Cas9 system

Chapter 6 Mutagenesis of CD630 using CRISPR-Cas9 system

6.1 Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated sequences (Cas) systems confer archaea and bacteria with an adaptive defence mechanism to evade invasion by foreign genomes like viruses, integrative conjugative elements, and plasmids (Barrangou *et al.*, 2007; Makarova *et al.*, 2006). A CRISPR array is composed of a leader sequence followed by short direct repeats (25-50 bp) separated by a succession of unique sequences, known as spacers, which have homology to foreign DNA (Bolotin *et al.*, 2005; Jansen *et al.*, 2002b; Mojica *et al.*, 2005; Pourcel *et al.*, 2005). Cas genes are usually found close to the CRISPR array and encode proteins required for cleaving the invader's DNA (Jansen *et al.*, 2002a). There are three distinct stages in CRISPR-Cas systems: adaptation, expression, and interference (Figure 6-1). In the adaptation stage, acquisition of a novel spacer takes place when Cas proteins (Cas1 and Cas2) recognise the foreign genome and incorporates segment of it into the leader-end of the CRISPR array (Barrangou *et al.*, 2007; Horvath *et al.*, 2008; Yosef *et al.*, 2012). This is followed by the expression phase where the CRISPR array is transcribed by a promoter within the leader sequence into pre-CRISPR RNA (pre-crRNA) (Tang *et al.*, 2002). These transcripts mature into individual crRNA each containing a repeat and a targeting sequence (Brouns *et al.*, 2008). In the interference stage, crRNA act as a guide (gRNA) and forms a ribonucleoprotein complex with other Cas proteins (Karvelis *et al.*, 2013). The

complex then targets invading nucleic acid that is complementary to the spacer sequence in the gRNA with the help of a conserved region named proto-spacer adjacent motif (PAM), which flank the targeted foreign sequence (Mojica *et al.*, 2009; Semenova *et al.*, 2011; Wiedenheft *et al.*, 2011). Cas proteins then act as nucleases leading to nucleic acid cleavage and hence destruction of the invading genetic element (Makarova *et al.*, 2006).

Figure 6-1 A simple illustration of the mechanism of the CRISPR-Cas system.

A. Adaptation: CRISPR-Cas proteins will recognise the phage DNA and incorporate a fragment of it into the CRISPR array immediately downstream from the leader sequence. **B. Expression:** The CRISPR array is transcribed into pre-crRNA then mature into crRNA which act as the target sequence (guide or gRNA) for future encounter with the phage. **C. Interference:** the crRNA-Cas protein complex will anneal to the phage DNA sequence that is complimentary to it. The Cas protein cause a double strand break within the target sequence destroying the phage. Figure is taken with permission from (Fineran & Charpentier, 2012).

There are numerous and diverse CRISPR-Cas systems in nature, but all are DNA encoded, RNA mediated, and nucleic acid targeting. But they differ in their gRNA types, PAM sequences, and cleavage outcomes. They can target foreign DNA or RNA (Hale *et al.*, 2009; Jiang *et al.*, 2016). CRISPR-Cas systems are classified into two classes, six types, and over 19 subtypes (Koonin *et al.*, 2017). The classification is based on Cas proteins, their function, and the CRISPR-Cas mechanism (Figure 6-2).

Figure 6-2 Classification of CRISPR according to Cas proteins.

The diagram lists the two classes and six types of CRISPR-Cas systems along with the Cas proteins and their functions. Figure is taken with permission from (Makarova *et al.*, 2015).

Class 2 type II-A CRISPR-Cas9 system of *S. pyogenes* was adapted for genome editing in human, mice, and bacteria (Cong *et al.*, 2013; Jiang *et al.*, 2013; Mali *et al.*, 2013). The CRISPR-Cas9 system can be programmed to target any DNA sequence adjacent to a PAM region by changing the 20 nucleotide sequence in the gRNA (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012). This system was utilized in clostridia for genetic manipulation (Bruder *et al.*,

2016; Huang *et al.*, 2016a; McAllister *et al.*, 2017; Pyne *et al.*, 2016). The principle of the method is to customize the gRNA-Cas9 complex to cause a blunt-ended double strand break (DSB) within the target sequence then provide the cells with a DNA repair template through homology-directed repair (HDR) for specific genome editing (Doudna & Charpentier, 2014). In eukaryotes and some bacteria like mycobacteria, random repair can occur through non-homologous end joining (NHEJ) (Shuman & Glickman, 2007) (Figure 6-3).

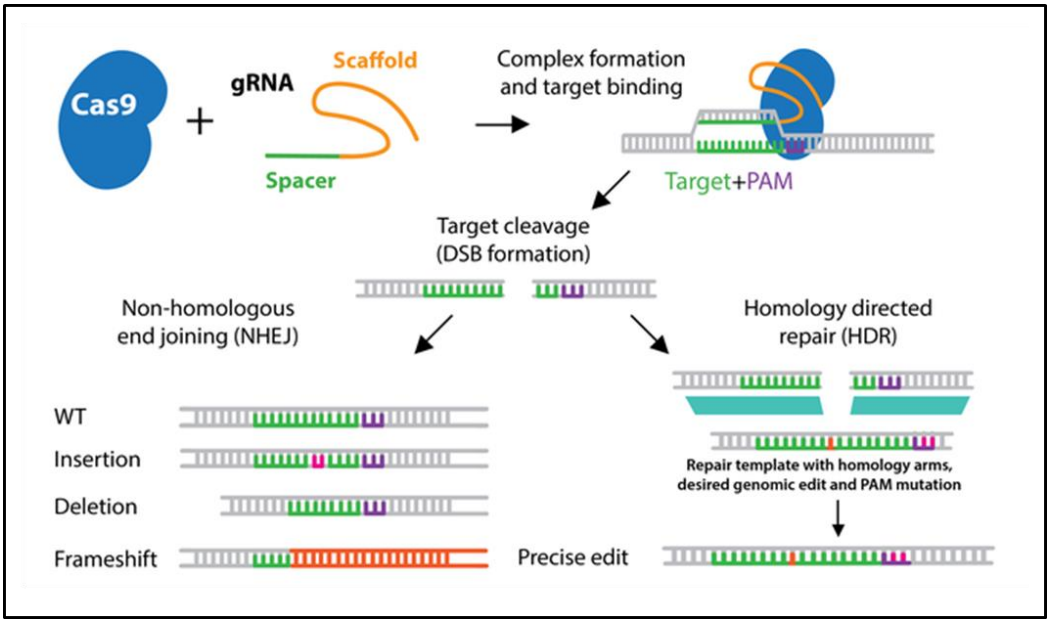


Figure 6-3 Mechanism of CRISPR-Cas9 system for genome editing.

CRISPR-Cas9 can be programmed to target any DNA sequence upstream of a PAM region by changing the 20 nucleotides in the gRNA (spacer in green). Cas 9 has two nucleases: RuvC and HNH that will cause a blunt end DSB within the target sequence leading to repair either by random mutation through NHEJ or precise editing by a DNA template through HDR. Figure is taken from (https://www.addgene.org/crispr/cut/).

CRISPR-Cas9 can be also used as an efficient tool for selecting deletion, insertion, and substitution mutants constructed by homologous recombination without the need to screen all transconjugants. This is done by allowing the bacterial cells to perform allelic exchange first then activating the *cas9*, by putting it under the control of an inducible promoter, leading to a DSB within the target sequence in the wild type. Therefore mutants with target sequence deletion can be selected (Wang *et al.*, 2018; Wang *et al.*, 2015; Wang *et al.*, 2016b).

In this chapter, a CRISPR-Cas9 based plasmid was constructed for the selection of a mutant with a deletion in the *atlf*. Also a general plasmid was created for any genetic manipulation in *C. difficile* that can be used by simply performing two cloning steps.

6.2 Materials and methods

6.2.1 Construction of ClosTron mutant

The pMTL007C-atlf ClosTron plasmid designed to knock out *atlf* was ordered from ATUM (formerly known as DNA 2.0) and recovered from the GFC filter according to the manufacturer's instructions. Conjugation was performed as described previously (Heap *et al.*, 2009) (Heap *et al.*, 2010). To screen transconjugants that contain group II insertions, PCR utilising Clostron-atlf-F and Clostron-atlf-R primer pair was performed. And to confirm that the group II intron was inserted at the correct position within the gene, Clostron-atlf-F and EBS primer pair was used in a second PCR reaction.

6.2.2 CRISPR-Cas9 plasmid construction

The plasmid (pMTL83151-CRISPR-Cas9-atlfΔd) constructed in this work to delete the *atlf* (CD630_28300) was derived from pMTL83151. The inducible promotor $P_{xyl/tetO}$ and the *cas9* gene were amplified using pRPF185 and pCas9 as templates respectively utilising Pxyl/tetO-F with Pxyl/tetO-R and Cas9-F with Cas9-R primer pair respectively. The two fragments were fused together using splicing by overlap extension (SOE) PCR utilizing Pxyl/tetO-F and Cas9-R primer pair. The Pxyl/tetO-Cas9 fragment was cloned into pMTL83151 upstream of Fdx terminator between *Xma*I and *Sal*I restriction enzyme sites resulting in plasmid pMTL83151-CRISPR-Cas9 with as shown in Figure 6-4A. The sgRNA fragment was synthesised by ThermoFisher Scientific and consisted of a strong *C. difficile* promoter (P_{tetM}), 20 bp target sequence within the *atlf*, and gRNA. The 20 bp target sequence was selected

using an algorithm for scoring and ranking potential target sites from the chopchop.rc.fas.harvard.edu website. The sgRNA fragment was amplified with gRNA-F & gRNA-R then cloned into pMTL83151-CRISPR-Cas9 upstream of CD0164 terminator between *Xma*I and *Not*I sites resulting in plasmid pMTL83151-CRISPR-Cas9-atlf as shown in Figure 6-4B. The homology regions were amplified by PCR using two pairs of primers (ZK32-F with ZK32-R and ZK33-F with ZK33-R) to produce 491 bp and 386 bp PCR products upstream and downstream from *atlf* (CD630_28300) respectively. The homology regions were then fused together by SOE PCR using ZK32F & ZK33R primer pair to produce a final 893 bp DNA region that was inserted into pMTL83151-CRISPR-Cas9-atlf at *Asc*I site resulting in plasmid pMTL83151-CRISPR-Cas9-atlfΔd as shown in Figure 6-4C.

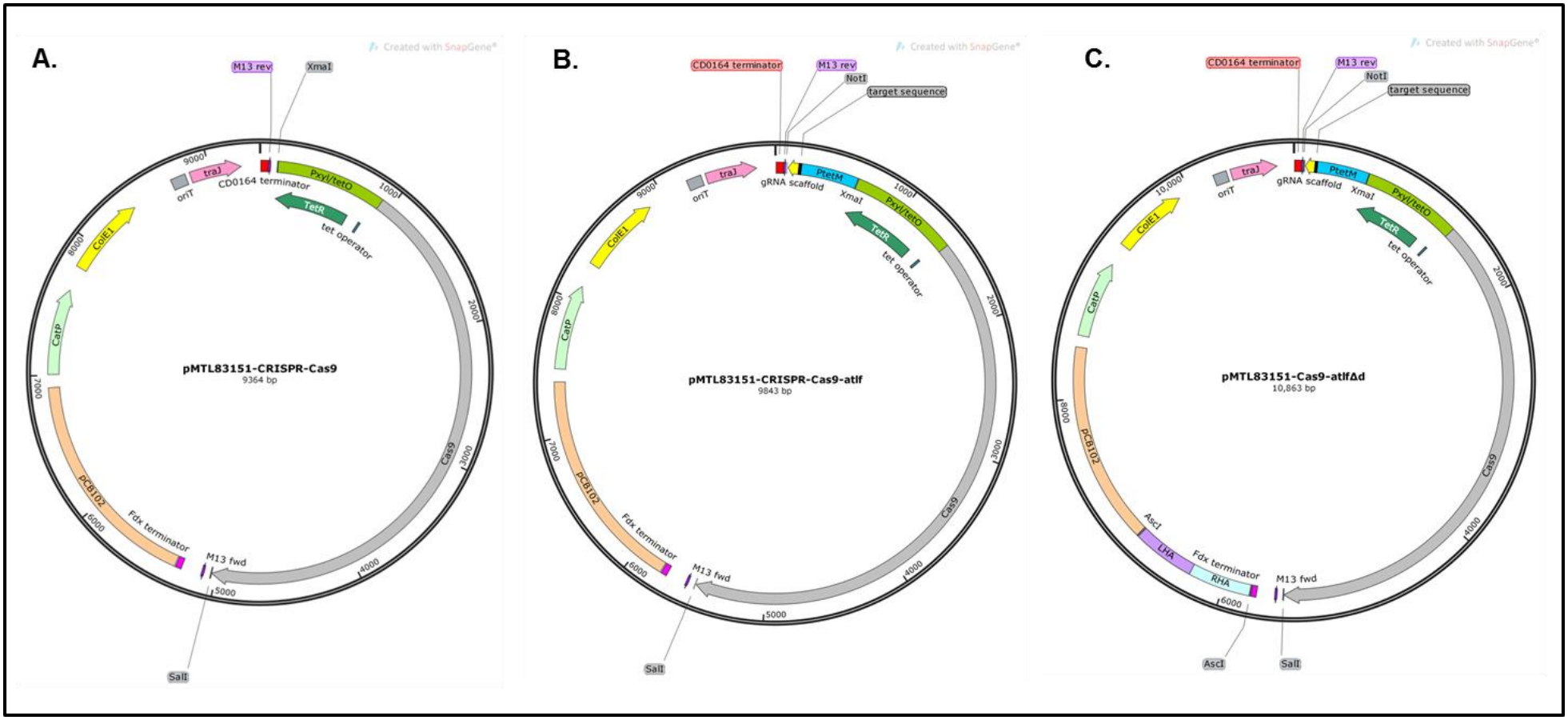


Figure 6-4 CRISPR-Cas9 plasmid maps.

A- Plasmid map of pMTL83151-CRISPR-Cas9 that was derived from pMTL83151 by insertion of *cas9* gene and $P_{xyl/tetO}$ inducible promoter between *Xma*I and *Sal*I sites. **B-** Plasmid map of pMTL83151-CRISPR-Cas9-atlf derived from plasmid pMTL83151-CRISPR-Cas9 with insertion of P_{tetM} promoter and sgRNA between *Xma*I and *Not*I sites. **C-** Plasmid map of pMTL83151-CRISPR-Cas9-atlfΔd derived from pMTL83151-CRISPR-Cas9-atlf with insertion of right and left homology regions (RHR; 491 bp and LHR; 386 bp upstream and downstream from the *atlf* (CD630_28300) respectively) at *Asc*I site. The maps were created by SnapGene.

6.2.3 Utilization of CRISPR-Cas9 system for the selection of deletion mutants

Plasmid pMTL83151-CRISPR-Cas9-atlfΔd was transferred to CD630Δerm by conjugation. Transconjugants grown on BHI plates supplemented with 15 µg/ml Tm and *C. difficile* selective supplement were picked and restreaked onto fresh plates of the same media for purification. After 48-72 hours, transconjugants were screened by colony PCR using Cas9-F and Cas9-R primer pair. Once the presence of the pMTL83151-CRISPR-Cas9-atlfΔd plasmid was confirmed, BHI broth containing Tm (15 µg/ml) and anhydrous tetracycline (ATc, 0.5 µg/ml from Sigma) was inoculated with single colony and incubated overnight. A serial dilution was made and 100 µl of the 10^{-3} to 10^{-6} dilutions were spread on two BHI plates supplemented with Tm and ATc. Mutants were analysed by PCR using three primer pairs (Figure 6-10, Figure 6-11, and Figure 6-12): i) Homology region primers (ZK34-F and ZK34-R); which anneal to regions 688 bp and 777 bp upstream and downstream from the deleted *atlf* in the chromosomal sequence respectively, ii) Plasmid primers (ZK35-F and ZK35-R): which anneal to regions 252 bp and 267 bp upstream and downstream from the homology regions within the plasmid sequence respectively, iii) Gene primers (ZK24-F and ZK24-R): that amplify the target gene. The amplicon from the homology region primers was sequenced to confirm the deletion.

6.2.4 Serial subculture to lose the plasmid

A single colony of the mutant with the desired deletion was then subcultured on BHI agar without any selection to lose the plasmid. If the plasmid was still

detected, a single colony of the mutant with the desired deletion was subcultured in BHI liquid medium without Tm. Next day 100 µl of the 18 hours culture was used to inoculate 10 ml of BHI. This subculture was repeated for five days. After each subculture, a serial dilution was made and 50 µl of the 10^{-5} dilution was spread on two plates of BHI without antibiotic. Replica plating was performed on BHI and BHI supplemented with Tm to select the colonies that can no longer grow on BHI supplemented with Tm. Loss of the plasmid was further confirmed by PCR using plasmid primers (ZK35-F and ZK35-R).

6.2.5 Haemolysis assay

The haemolysis assay for *C. difficile* (section 3.2.6) was used with the following modifications. Bacterial suspensions were all adjusted to an OD600 of 1 or to the lowest absorption measured. In four wells of a 96-well round-bottom microtiter plate, 100 µl of the bacterial suspensions were added. The plate was not centrifuged and the incubation time was increased to 48-72 hours or until haemolysis was visible. One unit of Streptolysin O solution (Sigma) was used as a second positive control and an indication of 50% haemolysis. Any reading which is ≥ 50% haemolysis is considered haemolytic. The defibrinated horse blood used to test the mutants was from TCS Biosciences Ltd (UK). The blood from E&O laboratories was used to confirm the assay's results.

6.2.6 Complementation of the mutants

Plasmid pRPF185-atlf was constructed by replacing *gusA* with *atlf* in between *Sac*I and *BamH*I in pRPF185 (Figure 6-5). The *atlf* was amplified by PCR utilizing C-atlf-F and C-atlf-R primer pair. Transfer of the pRPF185 and pRPF185-atlf by conjugation into the wild type and mutants respectively was performed as described in section 2.2.11 with one modification. Cultures were heated in a 50°C water bath for 5 minutes prior to conjugation with *E. coli* donor CA434 to increase the conjugation efficiency (Kirk & Fagan, 2016). The presence of the pRPF185-atlf plasmid was confirmed by PCR using pRPF-fdx-F and C-atlf-R primer pair and sequencing. The complemented mutants were maintained with Tm (15 µg/ml), and ATc (0.5 µg/ml) was added when haemolysis was tested.

6.2.7 Whole genome sequence of the CRISPR mutants and variation analysis

The extracted DNA from the mutants selected by CRISPR-Cas plasmid was sequenced using Illumina next-generation DNA sequencing by MicrobesNG, Birmingham, UK. And the DNA sequence was analysed for variations by alignment to the reference CD630 genome (272563.8) using PATRIC 3.5.26 (Wattam *et al.*, 2017). The variation analysis reports SNPs and indels with their specific locations and whether they are intergenic, synonymous, or nonsynonymous mutations.

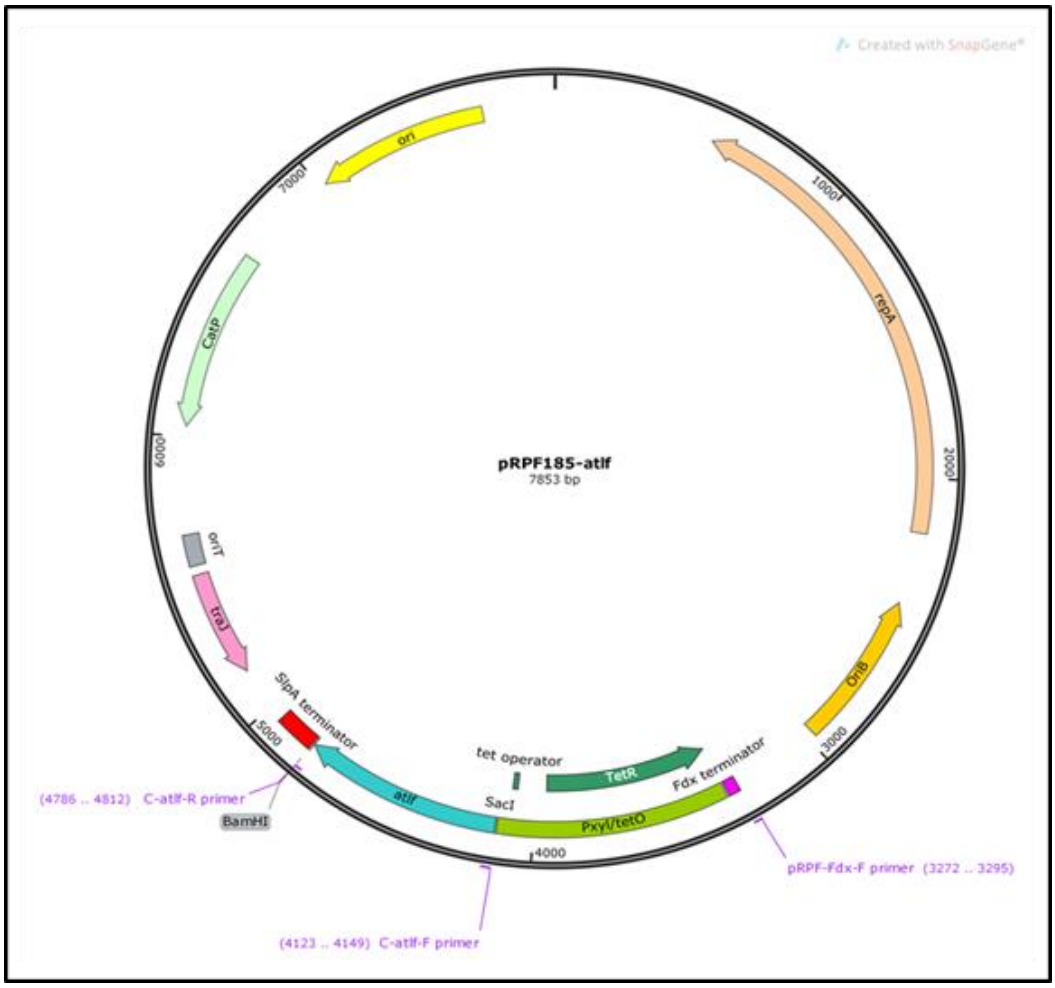


Figure 6-5 The complementation plasmid.

Plasmid pRPF185-atlf was derived from pRPF185 by insertion of *atlf* between *Sac*I and *BamH*I sites. The primers used for construction of the plasmid and for detection of the complimented mutant are indicated in purple. The map was created with SnapGene.

6.3 Results

6.3.1 Constructing a ClosTron mutant of *atlf*

The ClosTron mutagenesis system was utilised to construct an insertion mutant in the *atlf* (CD630_28300). The group II intron was inserted in the antisense orientation immediately after the 464th nucleotide in the coding sequence of *atlf*.

When genomic DNA isolated from the ClosTron mutant (CD630Δatlf-ClosTron) was used as the template for PCR with the Clostron-atlf-F & Clostron-atlf-R primer pair, a product of 2273 bp was amplified (Figure 6-6). This product is 1800 bp larger than that amplified from the wild type CD630Δerm (473 bp) with the same primer pair and represents group II intron insertion. To verify that the group II intron had inserted into the correct target gene, PCR was carried out using EBS and Clostron-atlf-F primers. A PCR product should only be amplified from the ClosTron mutant (475 bp) and not from the wild type. Sequence results confirmed that group II intron was inserted into the *C. difficile* chromosome at the correct site and in the correct orientation (Figure 6-7).

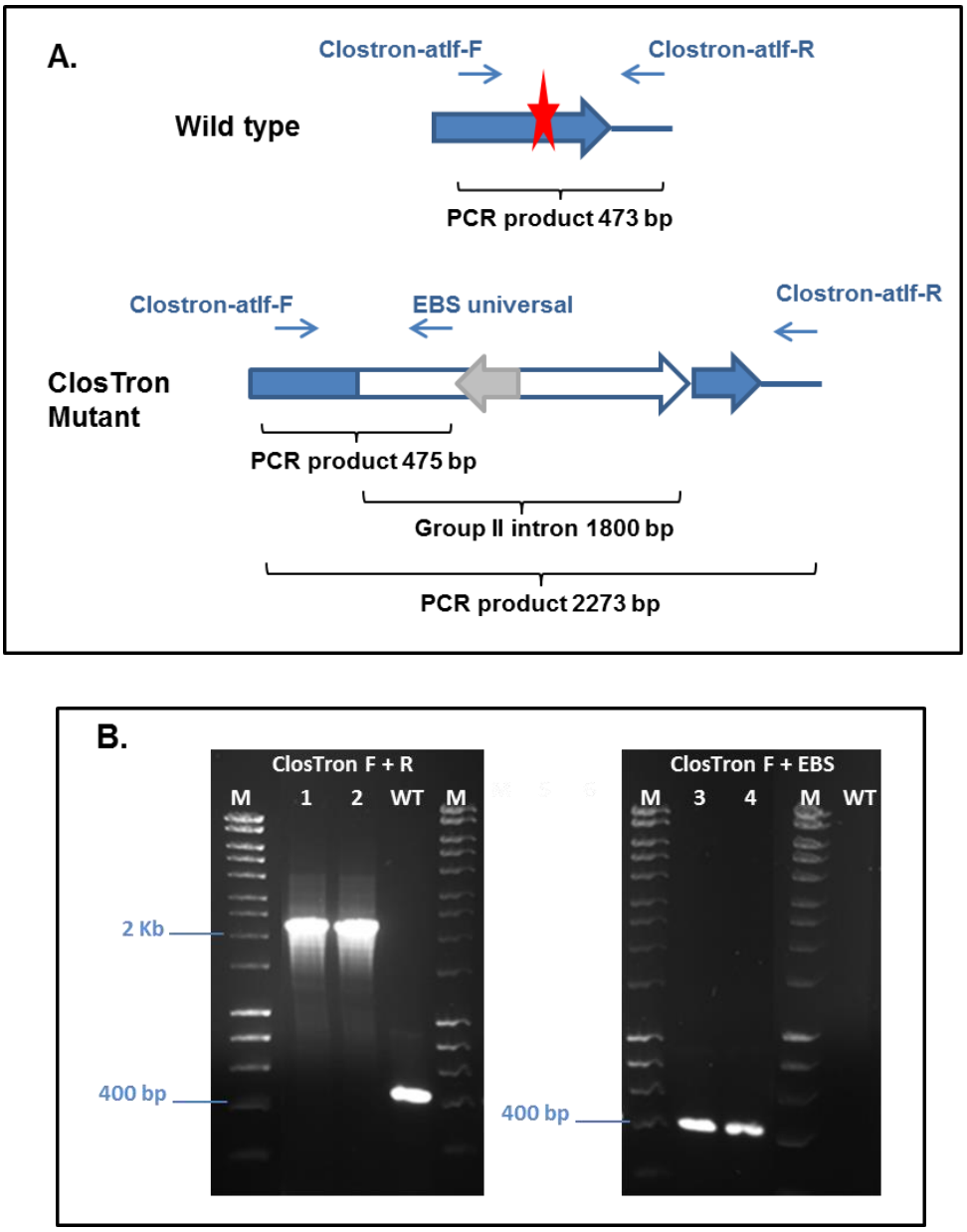


Figure 6-6 PCR screens of ClosTron mutagenesis to confirm insertion of group II intron in *atlf*.

A. Schematic representation of *atlf* (CD630_28300 in blue coloured arrow) where the group II intron (white arrow encoding a region of ermB gene in grey arrow) was inserted. The small blue arrows represent the primers used to screen for the insertion and the red star represents the position of the insertion within the targeted gene. **B.** 1% Agarose gel showing the PCR products amplified with Clostron-atlf-F/Clostron-atlf-R primers should give 2273 bp and 473 bp band sizes in the mutant and the wild type respectively. PCR product utilizing Clostron-atlf-F/EBS should give 475 bp band size in the mutant and no product with the wild type. M, 1 Kb DNA ladder; 1&2, CD630Δatlf-ClosTron mutant genomic DNA with Clostron-atlf-F/Clostron-atlf-R primers; WT, wild type (CD630Δerm) genomic DNA with Clostron-atlf-F/Clostron-atlf-R primers; 3&4, CD630Δatlf-ClosTron mutant genomic DNA with Clostron-atlf-F/EBS primers; WT, wild type (CD630Δerm) genomic DNA with Clostron-atlf-F/EBS primers.

Figure 6-7 DNA sequence of the *atlf* in CD630Δatlf-ClosTron mutant.

The full sequence of *atlf* is presented with the exact insertion site of the group II intron. The start and stop codons are highlighted in green. The sequence highlighted in grey is the group II intron and within the EBS primer is highlighted in yellow. Other primers designed to retarget the group II intron insertion are in blue. The target sequence within *atlf* chosen for group II intron insertion is in red.

6.3.2 Haemolysis of the ClosTron *atlf* mutant

To determine if group II intron insertion in the *atlf* (CD630_28300) had any effect on haemolysis, CD630Δatlf-ClosTron mutant was grown on blood agar with and without 2% glucose (Figure 6-8). Both the wild type and CD630Δatlf-ClosTron mutant were haemolytic in the presence of 2% glucose. This could be because i) *atlf* causes haemolysis in *E. coli* and not in *C. difficile*, ii) insertion of group II intron did not completely abolish the activity, or iii) more than one gene is causing haemolysis and multiple mutations are needed to see a lack of phenotype. To address the second theory, a new mutagenesis method was required to delete the complete *atlf* in *C. difficile*.

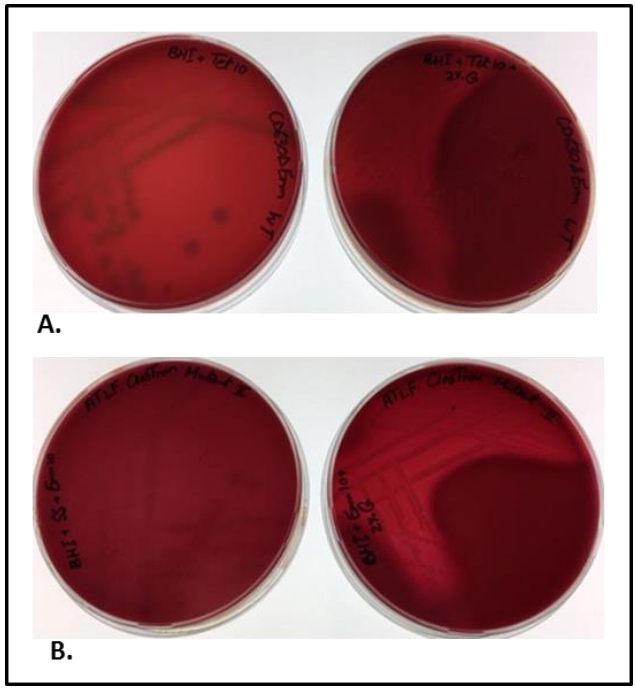


Figure 6-8 Haemolysis of CD630Δatlf-ClosTron mutant.

The ClosTron mutant with an insertion in *atlf*, CD630Δatlf-ClosTron, did not lose the haemolytic activity. **A.** Haemolysis of wild type CD630Δerm with glucose (right plate) and no haemolysis without glucose (left plate). **B.** CD630Δatlf-ClosTron mutant is haemolytic when grown in the presence of glucose (right plate) and shows no signs of haemolysis without (left plate). This demonstrates that the knock-out did not affect haemolysis.

6.3.3 Selecting deletion mutants using CRISPR-Cas9

Plasmid pMTL83151-CRISPR-Cas9-atlfΔd (Figure 6-4C) was constructed to select mutants with an *atlf* deletion in CD630Δerm after allelic exchange. The plasmid contained sgRNA, whose expression was controlled by a known strong promoter in *C. difficile* P_{tetM}, and *cas9* gene which was under the control of an inducible promoter to give *C. difficile* cells more time to undergo double-crossover. Two homology regions were included in the plasmid, which are specific for the *atlf*. The mechanism of mutant selection using CRISPR-Cas9 is illustrated in Figure 6-9.

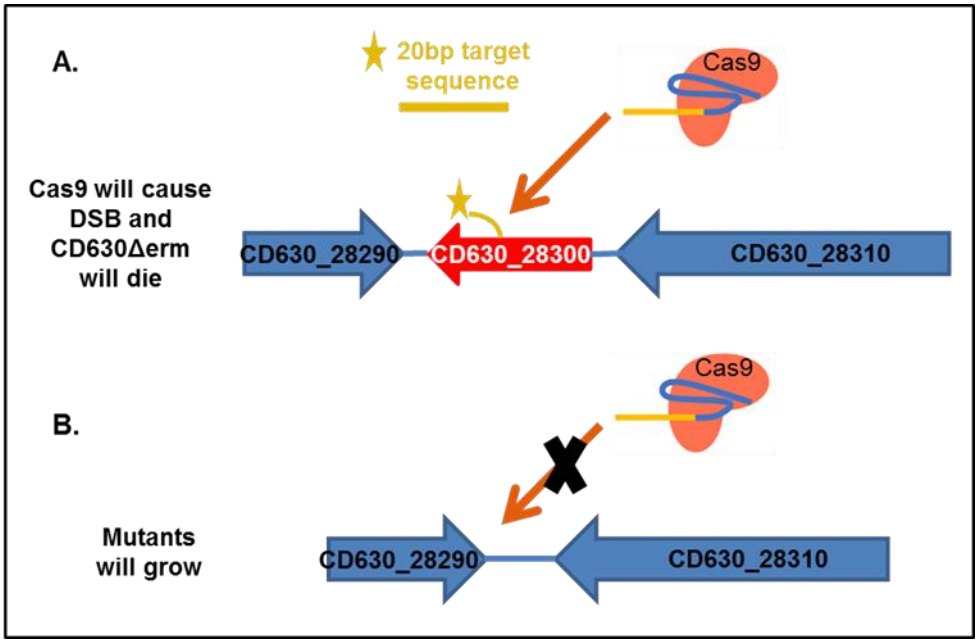


Figure 6-9 Utilizing CRISPR-Cas9 system as a selection rather than a mutation tool for genome editing in *C. difficile*.

A. The gRNA-Cas9 complex will recognise the target sequence and cause a DSB within it leading to death of the wild type strain. **B.** The gRNA-Cas9 complex will not find the target sequence in the mutant because it was deleted by allelic exchange hence will not cause a DSB leading to growth of the mutant only.

Four transconjugants were recovered from the 10^{-3} dilution plate, they were designated CD630Δatlf-CRISPR-MI, CD630Δatlf-CRISPR-MII, CD630Δatlf-CRISPR-MIII, and CD630Δatlf-CRISPR-MIV and all had the expected deletion. They were screened by PCR using three primer pairs as illustrated in Figure 6-10, Figure 6-11, and Figure 6-12. Amplicons from primers flanking the homology region within the CD630 genome were 2128 bp and 1468 bp in size in CD630Δerm (WT) and *atlf* deletion mutants respectively (Figure 6-10). PCR utilizing primers flanking the homology region within the plasmid sequence showed 1493 bp band for the *atlf* deletion mutants and no band for the wild type (CD630Δerm) DNA (Figure 6-11). When primers used to amplify the target gene were used, as expected no PCR product was obtained with the mutant DNA and 600 bp band was seen when CD630 genomic DNA was utilised as template (Figure 6-12). The PCR product of the first primer set was sequenced and the results confirmed the deletion of *atlf* (Figure 6-13).

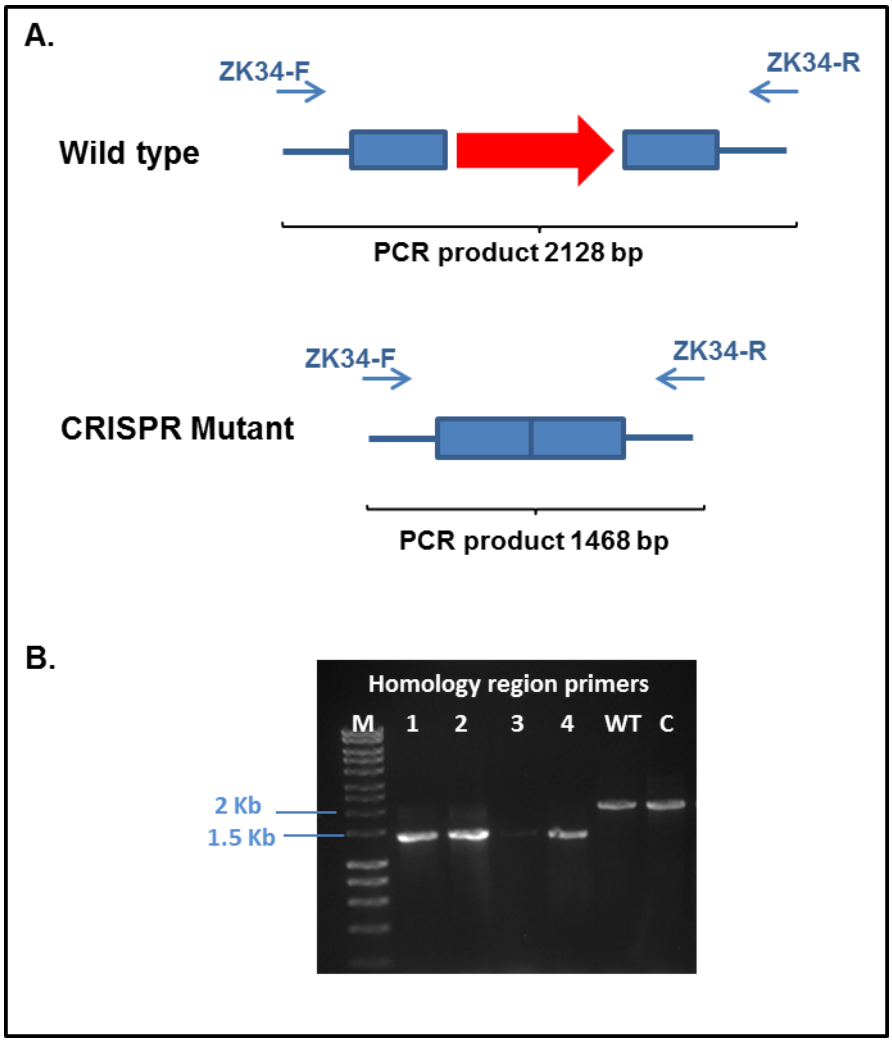


Figure 6-10 Screening of the CRISPR mutants with PCR utilizing primers flanking the region used for homologous recombination within CD630 genome.

A. An illustration showing the position of the primer pair and the expected PCR product size in CD630Δerm wild type and mutants. **B.** The agarose gel image shows the PCR products which were 2128 bp and 1468 bp in CD630Δerm (WT) and *atlf* deletion mutants respectively. M, 1 Kb DNA ladder; 1, CD630Δatlf-CRISPR-MI mutant; 2, CD630Δatlf-CRISPR-MII mutant; 3, CD630Δatlf-CRISPR-MIII mutant; 4, CD630Δatlf-CRISPR-MIV mutant; WT, wild type (CD630Δerm); C, CD630Δerm carrying pMTL83151-CRISPR-Cas9-atlfΔd plasmid as control.

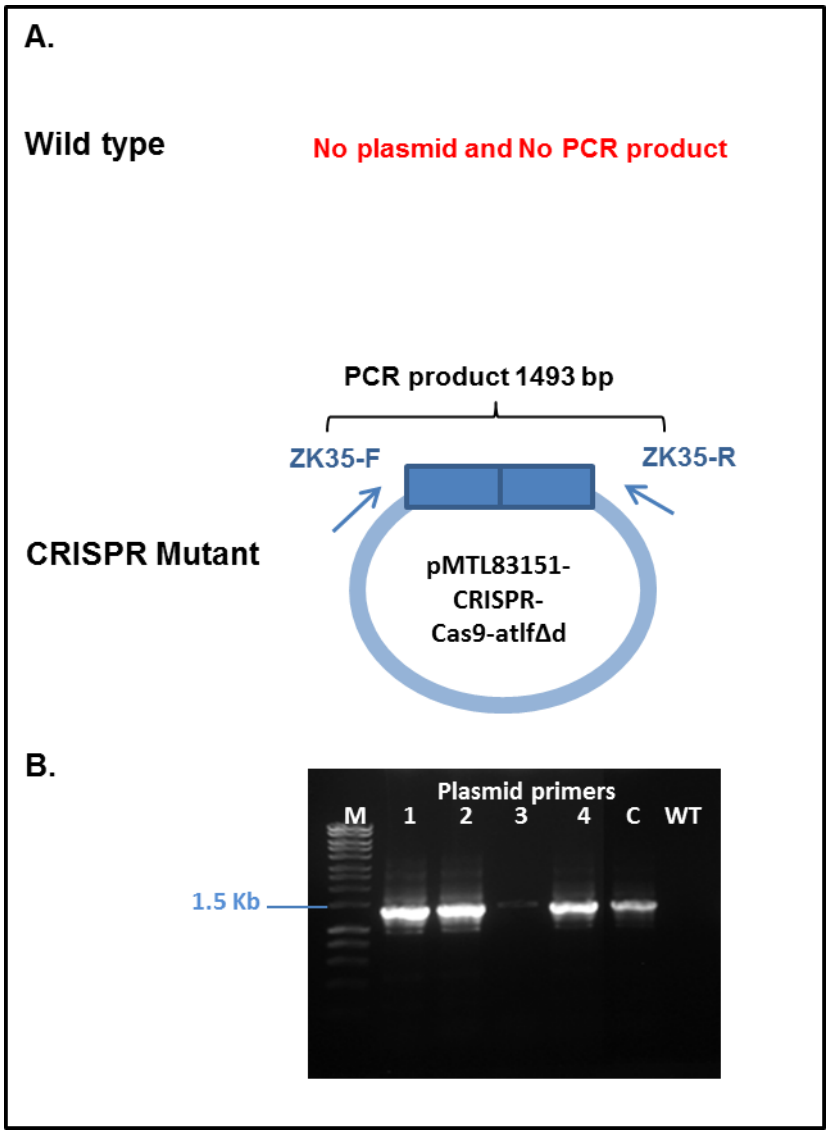


Figure 6-11 Screening of the CRISPR mutants with PCR utilizing primers flanking the homology region within the plasmid backbone.

A. An illustration showing the position of ZK35-F and ZK35-R primer pair and the expected PCR product size in CRISPR mutants and CD630Δerm. **B.** The agarose gel image shows 1493 bp band for *atlf* deletion mutants DNA and no PCR product for CD630Δerm (WT). M, 1 Kb DNA ladder; 1, CD630Δatlf-CRISPR-MI mutant; 2, CD630Δatlf-CRISPR-MII mutant; 3, CD630Δatlf-CRISPR-MIII mutant; 4, CD630Δatlf-CRISPR-MIV mutant; C, CD630Δerm carrying pMTL83151-CRISPR-Cas9-atlfΔd plasmid as control; WT, wild type (CD630Δerm).

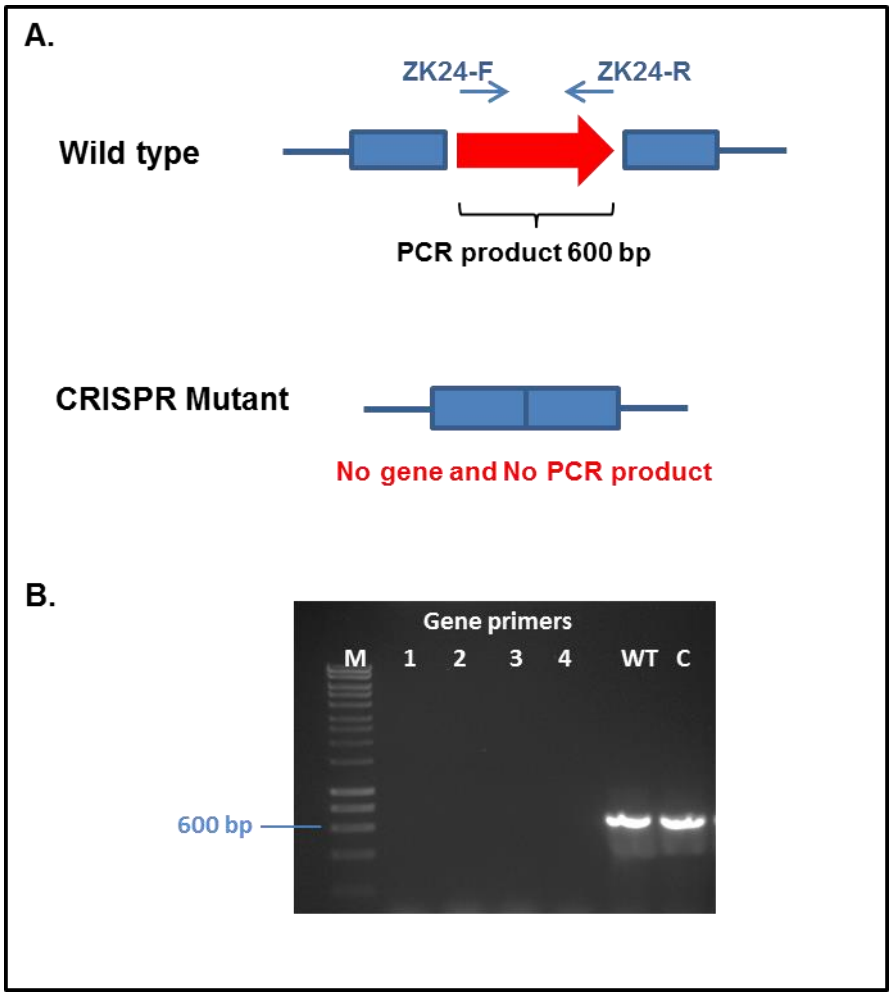


Figure 6-12 Screening of the CRISPR mutants with PCR utilizing primers amplifying *atlf*.

A. An illustration showing the position of ZK24-F and ZK24-R primer pair and the expected PCR product size for CD630Δerm and CRISPR mutants. **B.** The agarose gel image shows 600 bp band size for CD630Δerm genomic DNA and no PCR product for the *atlf* deletion mutants. M, 1 Kb DNA ladder; 1, CD630Δatlf-CRISPR-MI mutant; 2, CD630Δatlf-CRISPR-MII mutant; 3, CD630Δatlf-CRISPR-MIII mutant; 4, CD630Δatlf-CRISPR-MIV mutant; WT, wild type (CD630Δerm); C, CD630Δerm carrying pMTL83151-CRISPR-Cas9-atlfΔd plasmid as control.

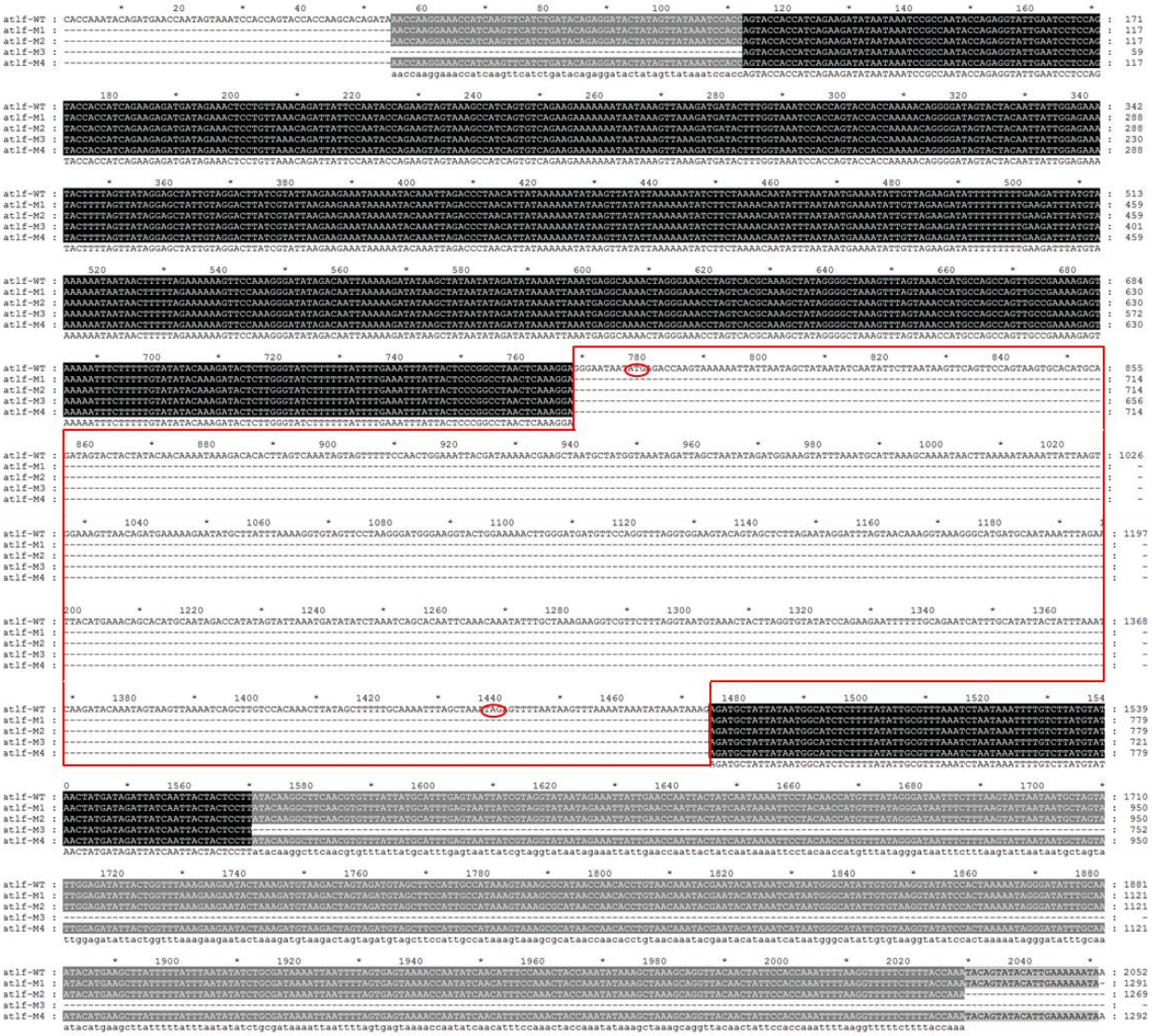


Figure 6-13 Sequence alignment of the four CRISPR mutants with CD630Δerm.

The mutants CD630Δatlf-CRISPR-MI, CD630Δatlf-CRISPR-MII, CD630Δatlf-CRISPR-MIII, and CD630Δatlf-CRISPR-MIV have a 713 bp deletion including the 660 bp *atlf*. The highlighted sequence represents the identical alignment between the mutant and wild-type. The sequence in the red box represents the deleted region in the mutant. The start and stop codons are shown in red circles.

6.3.4 Removal of the CRISPR plasmid from mutants

To eliminate any external effect on haemolysis, pMTL83151-CRISPR-Cas9-atlfΔd plasmid in the mutants should be removed. The mutants were grown for five successive days without selection. PCR confirmed that the plasmid was lost during the first overnight culture (Figure 6-14).

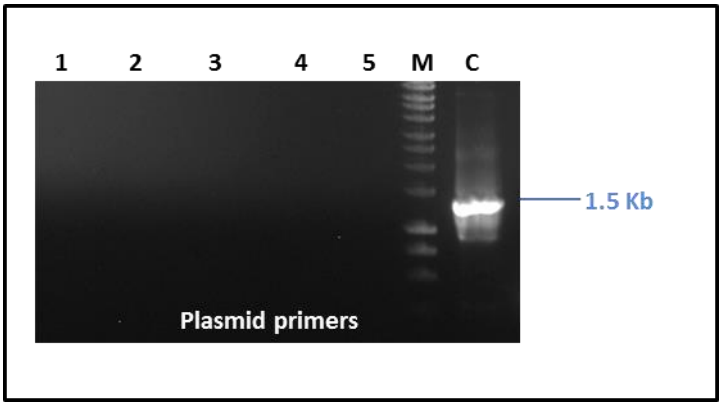


Figure 6-14 Detection of pMTL83151-CRISPR-Cas9-atlfΔd plasmid in the CRISPR mutants by PCR.

The PCR was done utilizing ZK35- F and ZK35-R primer pair illustrated in Figure 6-11. The presence of the plasmid was not detected in mutants growing without Tm each day for 5 days. There was no PCR product indicating loss of the plasmid. 1, day one; 2, day two; 3, day three; 4, day four; 5, day five; M, 1 Kb DNA ladder; C, pMTL83151-CRISPR-Cas9-atlfΔd plasmid DNA.

6.3.5 Haemolysis of CRISPR mutants

CD630Δatlf-CRISPR-MI and CD630Δatlf-CRISPR-MII mutants were grown on BHI with and without glucose. Haemolysis was observed only in the presence of glucose resembling the wild type CD630Δerm (Figure 6-15).

The mutants were also tested in the haemolysis assay along with the wild type to quantitatively record their haemolysis (Figure 6-16A). When mutants were tested at mid-exponential phase, CD630Δatlf-CRISPR-MI had no haemolysis when compared to the 50% haemolysis exhibited by the Streptolysin O control. On the other hand, CD630Δatlf-CRISPR-MII had haemolytic activity but less than the wild type. CD630Δatlf-CRISPR-MI demonstrated significant reduction in haemolysis when compared to the wild type and CD630Δatlf-CRISPR-MII (Figure 6-16B). In the haemolysis assay of mutants at the stationary phase, CD630Δatlf-CRISPR-MI showed very significant decrease in haemolysis compared to the wild type while CD630Δatlf-CRISPR-MII had similar activity as seen in Figure 6-16C. In conclusion, both mutants had an identical deletion but their phenotype was different.

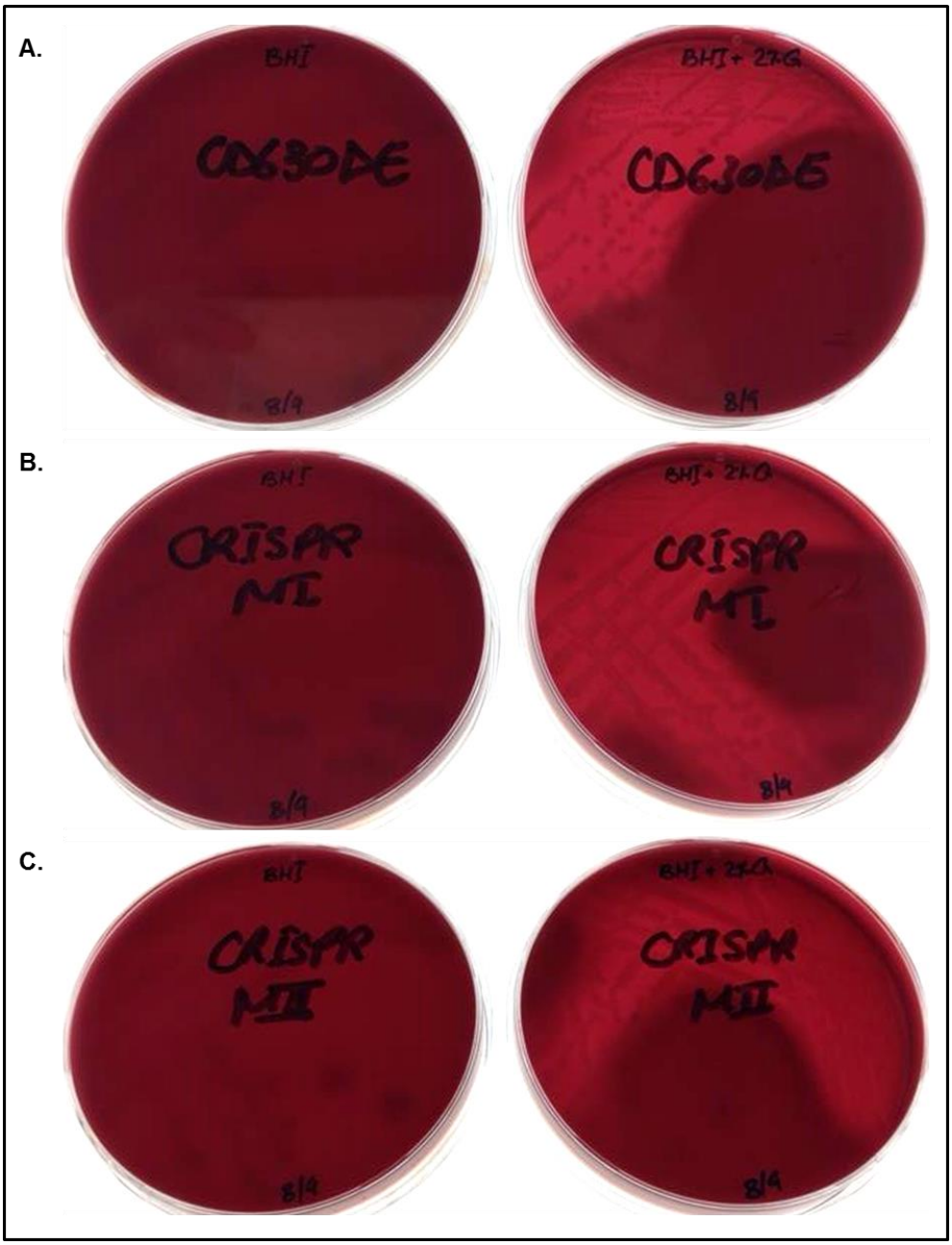


Figure 6-15 Haemolysis of CRISPR mutants on blood agar plates.

A. Wild type CD630Δerm **B.** CD630Δatlf-CRISPR-MI **C.** CD630Δatlf-CRISPR-MII. The plates on the right-hand side contain 2% glucose while the plates on the left are without. Both mutants showed haemolysis that is identical to the wild type.

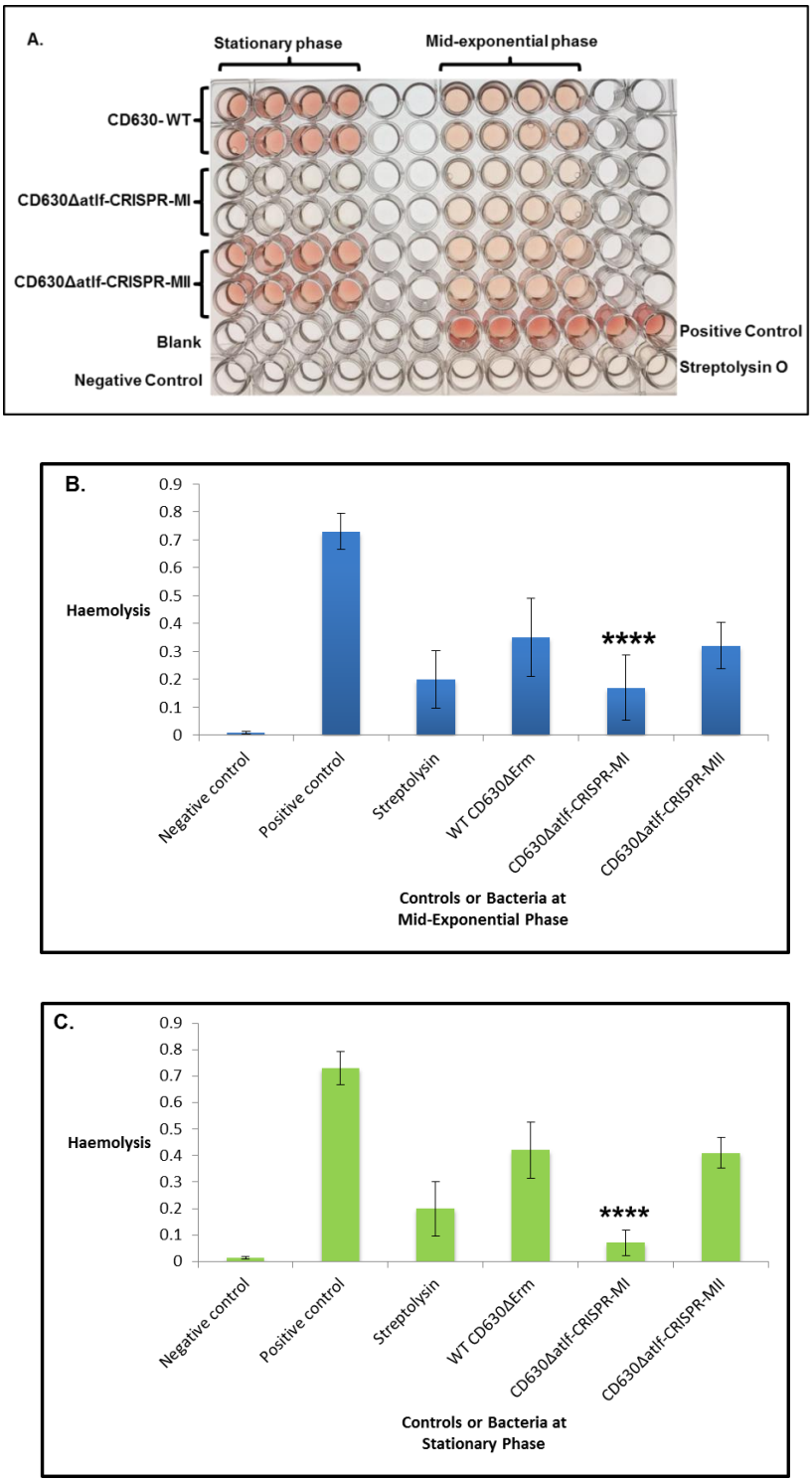


Figure 6-16 Haemolysis assay of the CRISPR mutants.

A. 96-well round-bottom microtiter plate used for the haemolysis assay demonstrating visible difference in haemolytic activity even without the measurement of OD540. **B.** Haemolysis at mid-exponential phase **C.** Haemolysis at stationary phase. Mutant CD630Δatlf-CRISPR-MI showed significant loss of haemolysis whereas CD630Δatlf-CRISPR-MII showed haemolytic activity that resembles the wild type. The standard deviation was calculated and presented as error bars. Statistical significance is indicated by four asterisks ****, $p < 0.0001$ (unpaired two-tailed t-test measured by GraphPad Prism v8.1.2).

6.3.6 Complementation of the CRISPR mutants

Mutants CD630Δatlf-CRISPR-MI and CD630Δatlf-CRISPR-MII were complemented by expressing *atlf* from the pRPF185-atlf plasmid using ATc. The complemented mutants were called CD630Δatlf-CRISPR-MI::atlf and CD630Δatlf-CRISPR-MII::atlf respectively. To make sure that the pRPF185 plasmid present in the mutants would not affect haemolysis, it was conjugated into CD630Δatlf-CRISPR-MI to create CD630Δatlf-CRISPR-MI-pRPF185. The presence of the pRPF-atlf plasmid in the mutants was confirmed by PCR (Figure 6-17A).

6.3.7 Haemolysis assay for the complemented CRISPR mutants

The mutants along with their complements and the wild type were tested in the haemolysis assay at stationary phase only (Figure 6-18). Mutant CD630Δatlf-CRISPR-MI showed haemolytic phenotype similar to the haemolysis exhibited by CD630Δatlf-CRISPR-MII and the wild type. The complemented strains CD630Δatlf-CRISPR-MI::atlf and CD630Δatlf-CRISPR-MII::atlf were sometimes haemolytic and sometimes non-haemolytic in the assay. The assay was performed four times and to confirm the results, it was repeated twice using defibrinated horse blood from another company (E & O Laboratories). The results are summarised in Table 6-1. To confirm that *atlf* is present in pRPF185-atlf plasmid in the complements, PCR utilizing the same primers used for the detection of complementation was done from colonies growing on agar and broth (Figure 6-17B). The *atlf* was present in the plasmid under the conditions tested.

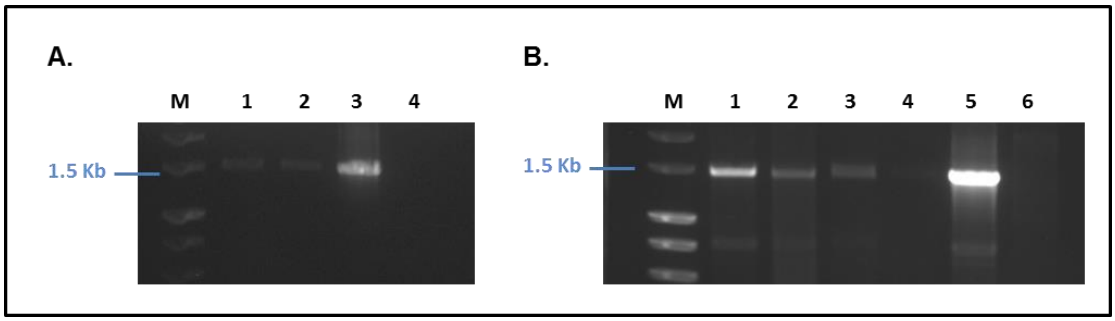


Figure 6-17 Detection of *atlf* in the complements.

A. PCR was performed to confirm the presence of pRPF-atlf plasmid in CD630Δatlf-CRISPR-MI and CD630Δatlf-CRISPR-MII mutants. The expected size of the PCR product was 1541 bp. M, 1 Kb DNA ladder; 1, CD630Δatlf-CRISPR-MI; 2, CD630Δatlf-CRISPR-MII; 3, pRPF185-atlf as positive control; 4, pRPF185 as negative control **B.** PCR to confirm the presence of the complementation plasmid in the mutants when grown in agar and liquid media. M, 1 Kb DNA ladder; 1, CD630Δatlf-CRISPR-MI from agar; 2, CD630Δatlf-CRISPR-MII from agar; 3, CD630Δatlf-CRISPR-MI from overnight growth; 4, CD630Δatlf-CRISPR-MII from overnight growth; 5, pRPF185-atlf as positive control; 6, pRPF185 as negative control.

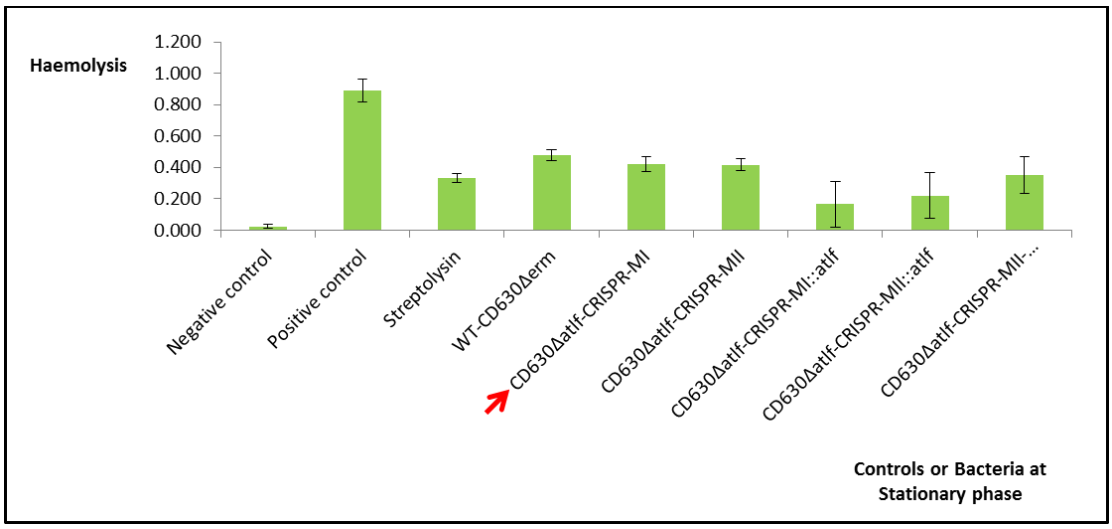


Figure 6-18 Haemolysis assay of the CRISPR mutants and their complements.

The bar graph represents the haemolysis at stationary phase where CD630Δatlf-CRISPR-MI mutant (red arrow) was surprisingly showing haemolytic activity (compared to Figure 6-16) which was similar to the haemolysis exerted by CD630Δatlf-CRISPR-MII. The complemented mutants CD630Δatlf-CRISPR-MI::atlf and CD630Δatlf-CRISPR-MII::atlf showed variability in haemolysis based on the error bars. The results were the mean of four replicates using horse blood from TCS Biosciences Ltd and the standard deviation was calculated and presented as error bars.

Table 6-1 Summary of the haemolytic activity of the CRISPR mutants and their complements.

Bacteria	Haemolysis experiment 1*	Haemolysis experiment 2*
WT	haemolytic	haemolytic
CD630Δatlf-CRISPR-MI	Non-haemolytic	haemolytic
CD630Δatlf-CRISPR-MII	haemolytic	haemolytic
CD630Δatlf-CRISPR-MI::atlf	-	haemolytic/ Non-haemolytic
CD630Δatlf-CRISPR-MII::atlf	-	haemolytic/ Non-haemolytic

*Haemolysis experiment 1; assay after creating mutants, Haemolysis experiment 2; assay after complementation.

6.3.8 Variation analysis of the CRISPR mutants

To understand the variability in the phenotypes observed, the DNA of mutants CD630Δatlf-CRISPR-MI and CD630Δatlf-CRISPR-MII was sequenced using Illumina next-generation DNA sequencing. And the sequence was analysed to see if there is a mutation elsewhere in the genomes.

The variation analysis of the whole genome sequence of the CRISPR-based mutants CD630Δatlf-CRISPR-MI and CD630Δatlf-CRISPR-MII are listed in Table 6-2 and Table 6-3. There were 28 SNPs and indels detected in both mutants in comparison to the wild type reference strain CD630 [272563.8].

There were three variations detected in CD630Δatlf-CRISPR-MI and not in CD630Δatlf-CRISPR-MII (highlighted in the Table 6-2) including two intergenic SNPs and one frameshift. A nucleotide substitution downstream of large subunit ribosomal RNA and upstream of tRNA-Leu and a nucleotide deletion downstream of predicted ATPase and upstream of Seryl-tRNA synthetase were detected. Also one nucleotide insertion that caused a frameshift within *flhB* gene (CD630_02620) that encodes a bifunctional flagellar biosynthesis protein FliR/FlhB (export apparatus membrane proteins).

There were six variations detected in CD630Δatlf-CRISPR-MII and not in CD630Δatlf-CRISPR-MI (highlighted in the Table 6-3) including five intergenic SNPs and one missense variation. A nucleotide substitution

downstream of large subunit ribosomal RNA and upstream of Maltose/maltodextrin transport ATP-binding protein MalK and four substitutions downstream of cell division-associated, ATP-dependent zinc metalloprotease FtsH and upstream of a translation initiation factor 3 were detected. In addition, a nonsynonymous substitution in the putative sigma-54-dependent transcriptional regulator (CD630_30940) was detected.

Table 6-2 Variation analysis of CD630Δatlf-CRISPR-MI mutant.

no	Pos	Ref	Var	Type	Ref_nt	Var_nt	Ref_nt_pos_change	Ref_aa_pos_change	Frameshift	Locus tag	Gene name	Function	Upstream feature	Downstream feature	SNP Eff type	SNP Eff impact
1	1000995	A	G	Nonsyn	act	Gct	121 A>G	Thr 41 Ala		CD0826		Transcriptional regulator, Fur family	Rubrerythrin	Superoxide reductase (EC 1.15.1.2)	missense_variant	MODERATE
2	103225	G	T	Synon	ggg	ggT	342 G>T	Gly 114 Gly		CD0073	rplC	LSU ribosomal protein L3p (L3e)	SSU ribosomal protein S10p (S20e)	LSU ribosomal protein L4p (L1e)	synonymous_variant	LOW
3	11002	A	G									Small Subunit Ribosomal RNA	RNA polymerase sigma factor SigB	tRNA-Ala-TGC	intergenic_region	MODIFIER
4	1391850	T	C	Nonsyn	ttt	Ctt	397 T>C	Phe 133 Leu		CD1190		Acetyltransferase TM0577	Uncharacterized amino acid permease, GabP family	Fructose-1,6-bisphosphatase, Bacillus type (EC 3.1.3.11)	missense_variant	MODERATE
5	1413056	AAGAATGTAGGAAATATAGA	AAGAATGTAGGAAATATAGAGAATGTAGGAAATATAGA	Insertion	aagaatgtaggaaatatagagacagaaatcacaaataata	AAGAATGTAGGAAATATAGAGAATGTAGGAAATATAGAcagaaatcacaaataata	465_466ins GAATGTAGGAAATATAGA	Glu 155_Thr 156 ins GluAsnValGlyAsnIleGlu		CD1214	spo0A	Stage 0 sporulation two-component response regulator (Spo0A)	Stage IV sporulation protein B	FIG005773: conserved membrane protein ML1361	inframe_insertion	MODERATE
6	146638	A	G										Large Subunit Ribosomal RNA	tRNA-Leu-TAA	intergenic_region	MODIFIER
7	1607452	GTTTTTTC	GTTTTTTTC	Insertion									Transcriptional regulator, MerR family	L-alanine-DL-glutamate epimerase (EC 5.1.1.n1)	intergenic_region	MODIFIER
8	2044514	C	G	Nonsyn	ccc	Gcc	97 C>G	Pro 33 Ala		CD1767	gapA	NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)	hypothetical protein	hypothetical protein	missense_variant	MODERATE
9	2203033	A	T										hypothetical protein	Transcriptional regulator, Xre family	intergenic_region	MODIFIER
10	2832892	G	T										hypothetical protein	Maltodextrin ABC transporter, ATP-binding protein MsmX	intergenic_region	MODIFIER
11	29024	C	T							CDr005	23s_rRNA	Large Subunit Ribosomal RNA	Small Subunit Ribosomal RNA	tRNA-Asn-GTT	intergenic_region	MODIFIER
12	2924655	C	T	Synon	gag	gaA	912 G>A	Glu 304 Glu		CD2532		Serine--glyoxylate aminotransferase (EC 2.6.1.45)	Substrate-specific component PdxU2 of predicted pyridoxin-related ECF transporter	ABC transporter-coupled two-component system, fused permease protein	synonymous_variant	LOW
13	2937176	C	A										Pyridine nucleotide-disulfide oxidoreductase	Putative transmembrane symporter	intergenic_region	MODIFIER

no	Pos	Ref	Var	Type	Ref_nt	Var_nt	Ref_nt_pos_change	Ref_aa_pos_change	Frameshift	Locus tag	Gene name	Function	Upstream feature	Downstream feature	SNP Eff type	SNP Eff impact
14	3005866	T	G										Lipoprotein signal peptidase (EC 3.4.23.36)	Peptidoglycan N-acetylglucosamine deacetylase (EC 3.5.1.-)	intergenic_region	MODIFIER
15	3034953	C	A	Nonsyn	ggt	Tgt	202 G>T	Gly 68 Cys		CD2627		ACT domain protein CAC_0478	FIG00519347: Ribonucleotide reductase-like protein	hypothetical protein	missense_variant	MODERATE
16	3080703	C	T	Nonsyn	gtt	Att	682 G>A	Val 228 Ile		CD2667	ptsG	PTS system, glucose-specific IIA/ IIB/ IIC components (EC 2.7.1.199)	PTS system, glucose-specific IIA/ IIB/ IIC components (EC 2.7.1.199)	Beta-glucoside bgl operon antiterminator, BglG family	missense_variant	MODERATE
17	322458	ATT	ATTT	Insertion	atttta	ATTTtta	559_560 ins T	Leu 187_Gly 188 fs	yes	CD0262	flhB	Flagellar biosynthesis protein FliR / FlhB	Flagellar biosynthesis protein FliQ	Flagellar biosynthesis protein FlhA	frameshift_variant	HIGH
18	3319972	ATT	AT	Deletion									predicted ATPase	Seryl-tRNA synthetase (EC 6.1.1.11)	intergenic_region	MODIFIER
19	3590223	C	A	Nonsyn	ggt	Tgt	802 G>T	Gly 268 Cys		CD3089		PTS system, alpha-D-glucoside-specific IIB / IIC component	Cellobiose phosphotransferase system YdjC-like protein	Trehalose operon transcriptional repressor	missense_variant	MODERATE
20	3591103	G	A										PTS system, alpha-D-glucoside-specific IIB / IIC component	Trehalose operon transcriptional repressor	intergenic_region	MODIFIER
21	3686534	TC	TAC	Insertion	ggattg	gGTAttg	308_309 ins T	Gly 103_Leu 104 fs	yes			hypothetical protein	hypothetical protein	hypothetical protein	frameshift_variant	HIGH
22	4004797	T	C							CDr028	23s_rRNA	Large Subunit Ribosomal RNA	Maltose/maltodextrin transport ATP-binding protein MalK (EC 3.6.3.19)	Small Subunit Ribosomal RNA	intergenic_region	MODIFIER
23	4007462	TA	TCA	Insertion									Large Subunit Ribosomal RNA	Small Subunit Ribosomal RNA	intergenic_region	MODIFIER
24	4007603	A	G										Large Subunit Ribosomal RNA	Small Subunit Ribosomal RNA	intergenic_region	MODIFIER
25	4007746	ACC	ACCC	Insertion								Small Subunit Ribosomal RNA	Large Subunit Ribosomal RNA	Precorrin-2 oxidase (EC 1.3.1.76) @ Sirohydrochlorin ferrochelatase activity of CysG (EC 4.99.1.4)	intergenic_region	MODIFIER
26	4166495	G	A	Nonsyn	gca	gTa	296 C>T	Ala 99 Val		CD3565		Predicted regulator PutR for proline utilization, GntR family	Stage II sporulation protein required for processing of pro-sigma-E (SpoIIR)	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148)	missense_variant	MODERATE

no	Pos	Ref	Var	Type	Ref_nt	Var_nt	Ref_nt_pos_change	Ref_aa_pos_change	Frameshift	Locus tag	Gene name	Function	Upstream feature	Downstream feature	SNP Eff type	SNP Eff impact
27	682619	C	T										hypothetical protein	Threonyl-tRNA synthetase (EC 6.1.1.3)	intergenic_region	MODIFIER
28	682653	G	A										hypothetical protein	Threonyl-tRNA synthetase (EC 6.1.1.3)	intergenic_region	MODIFIER
29	690658	A	T										HDIG domain protein	putative efflux protein	intergenic_region	MODIFIER
30	84143	C	T	Synon	gac	gaT	420 C>T	Asp 140 Asp		CD0058	tufA	Translation elongation factor Tu	RNA polymerase sporulation specific sigma factor SigH	LSU ribosomal protein L33p @ LSU ribosomal protein L33p, zinc-dependent	synonymous_variant	LOW
31	84227	C	T	Synon	gac	gaT	504 C>T	Asp 168 Asp		CD0058	tufA	Translation elongation factor Tu	RNA polymerase sporulation specific sigma factor SigH	LSU ribosomal protein L33p @ LSU ribosomal protein L33p, zinc-dependent	synonymous_variant	LOW

The highlighted rows are the variations detected in CD630Δatlf-CRISPR-MI in comparison to CD630 and not in CD630Δatlf-CRISPR-MII. Pos; position, Ref; reference strain CD630 [272563.8], Var; variant strain CD630Δatlf-CRISPR-MI, Ref_nt; reference strain nucleotide, Var_nuc; variant strain nucleotide, Ref_nt_pos; reference strain nucleotide position, Ref_aa_pos; reference strain amino acid position, SNP Eff; single nucleotide polymorphism effect, Synon; synonymous, Nonsyn; non- synonymous. The variation analysis was done by Patric 3.5.26.

Table 6-3 Variation analysis of CD630Δatlf-CRISPR-MII mutant.

no	Pos	Ref	Var	Type	Ref_nt	Var_nt	Ref_nt_pos_change	Ref_aa_pos_change	Frameshift	Locus tag	Gene name	Function	Upstream feature	Downstream feature	SNP Eff type	SNP Eff impact
1	1000995	A	G	Nonsyn	act	Gct	121 A>G	Thr 41 Ala		CD0826		Transcriptional regulator, Fur family	Rubrerythrin	Superoxide reductase (EC 1.15.1.2)	missense_variant	MODERATE
2	103225	G	T	Synon	ggg	ggT	342 G>T	Gly 114 Gly		CD0073	rplC	LSU ribosomal protein L3p (L3e)	SSU ribosomal protein S10p (S20e)	LSU ribosomal protein L4p (L1e)	synonymous_variant	LOW
3	11002	A	G									Small Subunit Ribosomal RNA	RNA polymerase sigma factor SigB	tRNA-Ala-TGC	intergenic_region	MODIFIER
4	1391850	T	C	Nonsyn	ttt	Ctt	397 T>C	Phe 133 Leu		CD1190		Acetyltransferase TM0577	Uncharacterized amino acid permease, GabP family	Fructose-1,6-bisphosphatase, Bacillus type (EC 3.1.3.11)	missense_variant	MODERATE
5	1413056	AAGAATGTAGGAAATATAGA	AAGAATGTAGGAAATATAGAGAATGTAGGAAATATAGA	Insertion	aagaatgtaggaaatatagagacagaaatcacaaatata	AAGAATGTAGGAAATATAGAGAATGTAGGAAATATAGAcagaaatcacaaatata	465_466 insGAATGTAGGAAATATAGA	Glu 155_Thr 156 insGluAsnValGlyAsnIleGlu		CD1214	spo0A	Stage 0 sporulation two-component response regulator (Spo0A)	Stage IV sporulation protein B	FIG005773: conserved membrane protein ML1361	inframe_insertion	MODERATE
6	1607452	GTTTTTTC	GTTTTTTTC	Insertion									Transcriptional regulator, MerR family	L-alanine-DL-glutamate epimerase (EC 5.1.1.n1)	intergenic_region	MODIFIER
7	2044514	C	G	Nonsyn	ccc	Gcc	97 C>G	Pro 33 Ala		CD1767	gapA	NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)	hypothetical protein	hypothetical protein	missense_variant	MODERATE
8	2203033	A	T										hypothetical protein	Transcriptional regulator, Xre family	intergenic_region	MODIFIER
9	2832892	G	T										hypothetical protein	Maltodextrin ABC transporter, ATP-binding protein MsmX	intergenic_region	MODIFIER
10	29024	C	T							CDr005	23s_rRNA	Large Subunit Ribosomal RNA	Small Subunit Ribosomal RNA	tRNA-Asn-GTT	intergenic_region	MODIFIER
11	2924655	C	T	Synon	gag	gaA	912G>A	Glu304Glu		CD2532		Serine--glyoxylate aminotransferase (EC 2.6.1.45)	Substrate-specific component PdxU2 of predicted pyridoxin-related ECF transporter	ABC transporter-coupled two-component system, fused permease protein	synonymous_variant	LOW
12	2937176	C	A										Pyridine nucleotide-disulfide oxidoreductase	Putative transmembrane symporter	intergenic_region	MODIFIER

no	Pos	Ref	Var	Type	Ref_nt	Var_nt	Ref_nt_pos_change	Ref_aa_pos_change	Frameshift	Locus tag	Gene name	Function	Upstream feature	Downstream feature	SNP Eff type	SNP Eff impact
13	3005866	T	G										Lipoprotein signal peptidase (EC 3.4.23.36)	Peptidoglycan N-acetylglucosamine deacetylase (EC 3.5.1.-)	intergenic_region	MODIFIER
14	3034953	C	A	Nonsyn	ggt	Tgt	202 G>T	Gly 68 Cys		CD2627		ACT domain protein CAC_0478	FIG00519347: Ribonucleotide reductase-like protein	hypothetical protein	missense_variant	MODERATE
15	3080703	C	T	Nonsyn	gtt	Att	682 G>A	Val 228 Ile		CD2667	ptsG	PTS system, glucose-specific IIA/ IIB/ IIC components (EC 2.7.1.199)	PTS system, glucose-specific IIA/ IIB/ IIC components (EC 2.7.1.199)	Beta-glucoside bgl operon antiterminator, BglG family	missense_variant	MODERATE
16	3590223	C	A	Nonsyn	ggt	Tgt	802 G>T	Gly 268 Cys		CD3089		PTS system, alpha-D-glucoside-specific IIB / IIC component	Cellobiose phosphotransferase system YdjC-like protein	Trehalose operon transcriptional repressor	missense_variant	MODERATE
17	3591103	G	A										PTS system, alpha-D-glucoside-specific IIB / IIC component	Trehalose operon transcriptional repressor	intergenic_region	MODIFIER
18	3597946	T	C	Nonsyn	agt	Ggt	490A>G	Ser164Gly		CD3094		putative sigma-54-dependent transcriptional regulator	Glutamine amidotransferase, class I	6-phospho-beta-glucosidase (EC 3.2.1.86)	missense_variant	MODERATE
19	3686534	TC	TAC	Insertion	ggattg	gGTAttg	308_309 ins T	Gly 103_Leu 104 fs	yes			hypothetical protein	hypothetical protein	hypothetical protein	frameshift_variant	HIGH
20	4004797	T	C							CDr028	23s_rRNA	Large Subunit Ribosomal RNA	Maltose/maltodextrin transport ATP-binding protein MalK (EC 3.6.3.19)	Small Subunit Ribosomal RNA	intergenic_region	MODIFIER
21	4004828	C	T							CDr028	23s_rRNA	Large Subunit Ribosomal RNA	Maltose/maltodextrin transport ATP-binding protein MalK (EC 3.6.3.19)	Small Subunit Ribosomal RNA	intergenic_region	MODIFIER
22	4007462	TA	TCA	Insertion									Large Subunit Ribosomal RNA	Small Subunit Ribosomal RNA	intergenic_region	MODIFIER
23	4007603	A	G										Large Subunit Ribosomal RNA	Small Subunit Ribosomal RNA	intergenic_region	MODIFIER
24	4007746	ACC	ACCC	Insertion								Small Subunit Ribosomal RNA	Large Subunit Ribosomal RNA	Precorrin-2 oxidase (EC 1.3.1.76) @ Sirohydrochlorin ferrochelatase activity of CysG (EC 4.99.1.4)	intergenic_region	MODIFIER
25	4166495	G	A	Nonsyn	gca	gTa	296 C>T	Ala 99 Val		CD3565		Predicted regulator PutR for proline utilization, GntR family	Stage II sporulation protein required for processing of pro-sigma-E (SpoIIR)	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148)	missense_variant	MODERATE

no	Pos	Ref	Var	Type	Ref_nt	Var_nt	Ref_nt_pos_change	Ref_aa_pos_change	Frameshift	Locus tag	Gene name	Function	Upstream feature	Downstream feature	SNP Eff type	SNP Eff impact
26	682619	C	T										hypothetical protein	Threonyl-tRNA synthetase (EC 6.1.1.3)	intergenic_region	MODIFIER
27	682653	G	A										hypothetical protein	Threonyl-tRNA synthetase (EC 6.1.1.3)	intergenic_region	MODIFIER
28	690658	A	T										HDIG domain protein	putative efflux protein	intergenic_region	MODIFIER
29	830614	A	G										Cell division-associated, ATP-dependent zinc metalloprotease FtsH	Translation initiation factor 3	intergenic_region	MODIFIER
30	830620	T	C										Cell division-associated, ATP-dependent zinc metalloprotease FtsH	Translation initiation factor 3	intergenic_region	MODIFIER
31	830639	G	A										Cell division-associated, ATP-dependent zinc metalloprotease FtsH	Translation initiation factor 3	intergenic_region	MODIFIER
32	830656	A	G										Cell division-associated, ATP-dependent zinc metalloprotease FtsH	Translation initiation factor 3	intergenic_region	MODIFIER
33	84143	C	T	Synon	gac	gaT	420 C>T	Asp 140 Asp		CD0058	tufA	Translation elongation factor Tu	RNA polymerase sporulation specific sigma factor SigH	LSU ribosomal protein L33p @ LSU ribosomal protein L33p, zinc-dependent	synonymous_variant	LOW
34	84227	C	T	Synon	gac	gaT	504 C>T	Asp 168 Asp		CD0058	tufA	Translation elongation factor Tu	RNA polymerase sporulation specific sigma factor SigH	LSU ribosomal protein L33p @ LSU ribosomal protein L33p, zinc-dependent	synonymous_variant	LOW

The highlighted rows are the variations detected in CD630Δatlf-CRISPR-MII in comparison to CD630 and not in CD630Δatlf-CRISPR-MI. Pos; position, Ref; reference strain CD630 [272563.8], Var; variant strain CD630Δatlf-CRISPR-MI, Ref_nt; reference strain nucleotide, Var_nuc; variant strain nucleotide, Ref_nt_pos; reference strain nucleotide position, Ref_aa_pos; reference strain amino acid position, SNP Eff; single nucleotide polymorphism effect, Synon; synonymous, Nonsyn; non- synonymous. The variation analysis was done by Patric 3.5.26.

6.4 Discussion

The CD630Δatlf-ClosTron mutant with interrupted *atlf* that was constructed by ClosTron technology was haemolytic on blood agar. As discussed in chapter 1 (section 1.8), the ClosTron system has the limitation that in an insertion mutant one cannot always be sure that gene function is completely abolished. Therefore it was decided to completely delete *atlf.*

The CRISPR-Cas9 system has been used for different mutational purposes in human, animal, plant, fungi, and bacteria (Adli, 2018; Doudna & Charpentier, 2014). It can be used to generate deletions, insertions, and single nucleotide integration, making it currently the best tool available. In this work CRISPR-Cas9 was utilized as a selection tool for low-frequency homologous recombination events. Plasmid pMTL83151-CRISPR-Cas9-atlfΔd was constructed for the selection of deletion mutant of *atlf*. A CRISPR-Cas9 based method was used to select mutants in industrially significant clostridia, *C. beijerinckii*, and recently in *C. difficile* (Wang *et al.*, 2018; Wang *et al.*, 2015; Wang *et al.*, 2016b). The work here was done during the same time and independently from the work done by Wang *et al.* as explained in the next section (Wang *et al.*, 2018).

Deletion of 50 bp, 262 bp, and 1500 bp from the *C. beijerinckii* genome was achieved using a CRISPR-Cas9 based plasmid with high efficiency. The plasmid contained a lactose inducible promoter *spoIIE* to express *cas9* and *C. beijerinckii* small RNA gene promoter for gRNA expression. In addition thiolase gene and *S. pyogenes* terminators were cloned downstream of *cas9*

and gRNA respectively (Wang *et al.*, 2015; Wang *et al.*, 2016b). In *C. difficile*, the same components were cloned in pMTL82151 (Wang *et al.*, 2018). The plasmid constructed in this work has different components including an ATc inducible promoter ($P_{xyl/tetO}$) and P_{tetM} promoter for the expression of *cas9* and gRNA respectively. The *cas9* and gRNA with their promoters were cloned upstream of terminators Fdx and CD0146 respectively present in pMTL83151 backbone.

Cpf1 is characterised as a sgRNA endonuclease, similar to Cas9, but cleaves DNA by a staggered DSB creating a -5' overhang (Zetsche *et al.*, 2015). The CRISPR-Cpf1 system is described as more favourable for the genetic manipulation because it is less toxic to the cells and has higher conjugation efficiency compared to CRISPR-Cas9. Recently, it was used to delete up to 49.2 Kb *phiCD630-2* phage from CD630 in addition to *ermB1/2* and *tetM* (Hong *et al.*, 2018). The toxicity of CRISPR-Cas9 system has been determined in this work but was found to be only slightly toxic to bacteria. This was done when the pMTL83151-CRISPR-Cas9 plasmid was conjugated into CD630Δerm and the $P_{xyl/tetO}$ promoter was induced, the number of viable cells when *cas9* was expressed was lower than the control (7.9×10^{-7} compared to 8.4×10^{-7} cfu/ml). Also only one experiment was needed to produce transconjugants. The CRISPR-Cas9 system used in this work was very effective in generating the required mutation.

The CRISPR-Cas9 system was used to create CD630Δatlf-CRISPR-MI and CD630Δatlf-CRISPR-MII mutants which had a deletion of the 663 bp *atlf*.

Both mutants had the same deletion confirmed by PCR and sequencing but they had different phenotypes. Mutant CD630Δatlf-CRISPR-MI showed a significant decrease in haemolysis while CD630Δatlf-CRISPR-MII showed haemolysis level comparable to the wild type CD630Δerm at the stationary phase. Complementation of the two mutants and testing their haemolysis at the stationary growth phase showed variations in the haemolytic activity. It was thought that the *atlf* was lost from the complementation plasmid during conjugation into the mutants but PCR confirmed its presence.

Surprisingly, mutant CD630Δatlf-CRISPR-MI when tested the second time with the complement in the haemolysis assay showed haemolytic phenotype unlike the first time tested (Figure 6-16 and Figure 6-18). After sequencing the two mutants, few differences were detected between them. The impact of the frameshift detected in *flhB-fliR* gene in mutant CD630Δatlf-CRISPR-MI and its implication in haemolysis is unknown. A mutant with insertional inactivation of *flhB-fliR* (by ClosTron) showed loss of motility, reduced toxin production, reduced toxin transcript levels, and lower virulence in hamster model (Aubry *et al.*, 2012).

The change in phenotype in CD630Δatlf-CRISPR-MI may be due to the blood used for the haemolysis assay. Although it was from the same company, it may have different cell membrane protein composition leading to change in the target site for ATLF.

Another hypothesis is that there is more than one haemolysin gene in *C. difficile* and they are controlled by phase variation. This means that some

genes have an ON and OFF orientation leading to switching on and off the expression and hence the phenotype. The reason why only a variable phenotype in the mutant is seen is that there is more than one haemolysin and deleting *atlf* allows the expression of the variable haemolysin to be observed. A good example of phase variation in *C. difficile* is the flagellar switch. When the promoter of the *flgB* is oriented as shown in the NCBI database, the flagella is expressed and CD630 is motile and produces toxins (phase ON). Whereas in CD630E, the *flg* promoter is in the inverse orientation, therefor leading to loss of motility and reduced toxin production (phase OFF) (Collery *et al.*, 2017). The flagellar switch in ribotypes 027 and 017 is different than CD630, as it was found to be a 154 bp invertible sequence flanked by 21 bp inverted repeats upstream of the *flgB* operon. During growth *in vitro*, the orientation of the flagellar switch is reversible (Anjuwon-Foster & Tamayo, 2017). The conditions that control the switch are not yet determined but was postulated that ON phase promotes bacterial colonization and infection while OFF phase allows host immune evasion (Anjuwon-Foster & Tamayo, 2018). Another genetic switch responsible for phase variation in *C. difficile* is the *cwp*V switch. The 195 bp invertible sequence flanked by imperfect 21 bp inverted repeats was found between *cwp* and its promoter. When the invertible sequence is in one orientation, the CwpV is expressed (phase ON). However, when the sequence is in the opposite orientation, it forms a stem loop terminator that prevents the expression of CwpV (phase OFF) (Emerson *et al.*, 2009). CwpV has a direct aggregation-promoting function and provides phage resistance in *C. difficile* (Reynolds *et al.*, 2011; Sekulovic *et al.*, 2015). Therefore, future work should

focus on the haemolysin genes discussed in Chapter 4 or to identify new genes.

The construction of the pMTL83151-CRISPR-Cas9-atlfΔd plasmid was laborious and time consuming but a general plasmid derived from pMTL83151-CRISPR-Cas9-atlf can be used as the basis for any *C. difficile* mutagenesis in the future. The general plasmid pMTL83151-CRISPR-Cas9-Cd was created by replacing the 20 bp target sequence in the gRNA with LacZ fragment (β-galactosidase encoded by *lacZ*) flanked by two *Kas*I sites. This plasmid can be used by two step cloning: first the 20 bp target sequence (designed as forward and reversed primers that are annealed together), then the homology regions specific for any target gene (Figure 6-19).

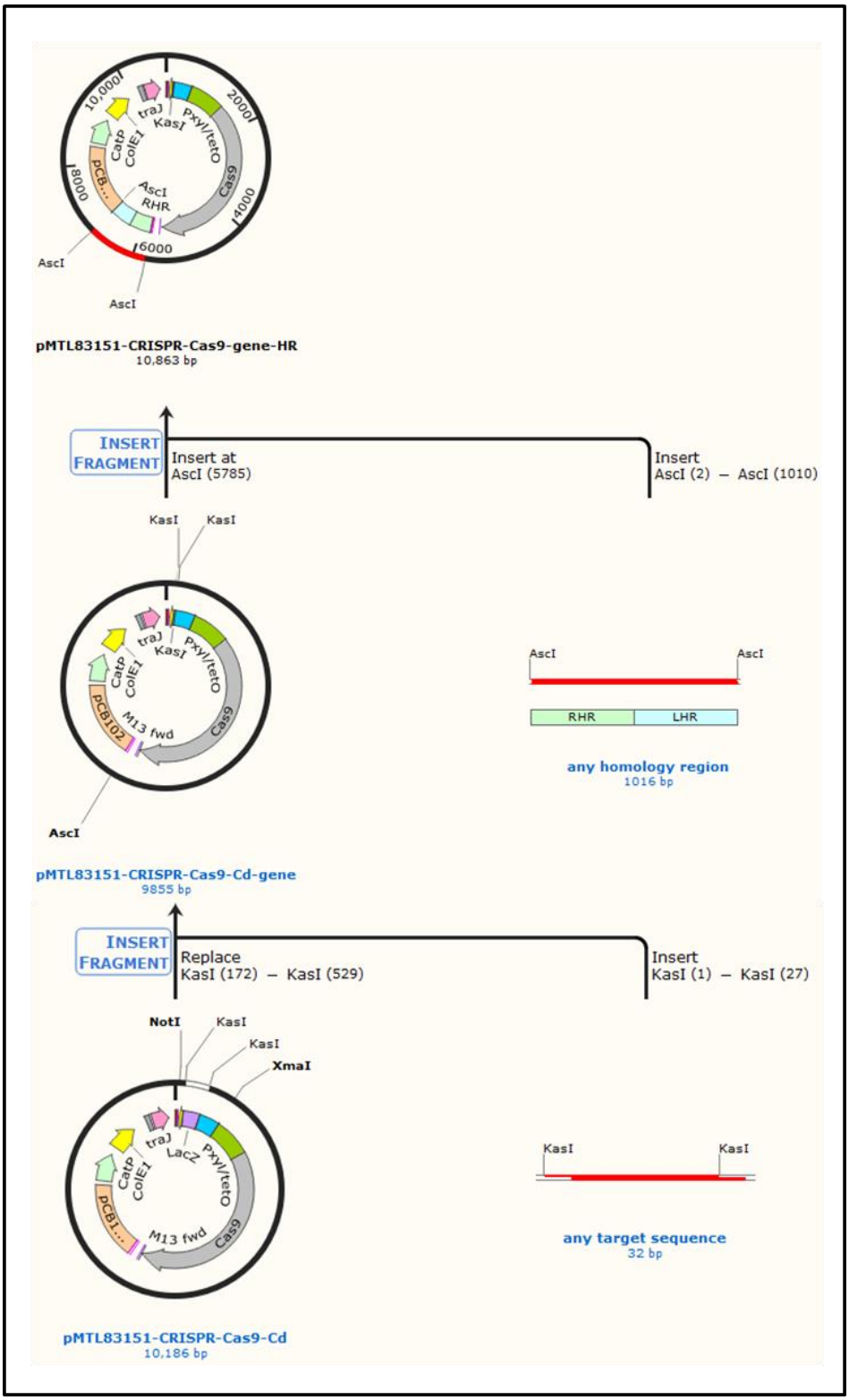


Figure 6-19 How to use pMTL83151-CRISPR-Cas9-Cd for any genetic manipulation in *C. difficile*.

To construct the desired plasmid, two cloning steps are required. Any target sequence can be designed as 20 bp primer with *Kas*I restriction site (5'-GCGCC-20bp-G-3') then cloned into the plasmid. Any homology regions can be amplified by SOE-PCR then cloned in between the *Asc*I site. Created by SnapGene.

6.5 Conclusions

The construction of a CRISPR-Cas9 based plasmid for the selection of a low-frequency homologous recombination event in *C. difficile* is very efficient. Deletion of the *atlf* showed variability in the haemolytic phenotype. Haemolysis is a multifactorial phenomenon that may involve several genes that encode haemolysins which suggest more work is needed to understand haemolysis in *C. difficile*.

Chapter 7

General conclusions and scope for future work

Chapter 7 General conclusions and scope for future work

This work describes the haemolytic activity of CD630 when 2% glucose is added to BHI agar containing horse blood. Previously, a CD630 genomic library has revealed that the *E. coli* clone carrying genes CD630_10330 and CD630_10340 encode a haemolysin. Cloning CD630_10340 gene individually confirmed its ability to confer haemolysis in *E. coli*. However, knocking out this gene in *C. difficile* did not have any effect on haemolysis. This work suggests a correlation between glucose and haemolysis in *C. difficile*. The precise correlation was never investigated and requires further work as discussed below.

This work concentrated on finding genes that encodes haemolysins in CD630 through the construction of a genomic library and selection and analysis of haemolytic clones. The library revealed six possible genes responsible for haemolysis in *E. coli*. CD630_02921 predicted to encode a protein which contains phospholipase D-nuclease N-terminal domain downstream of a gene encoding an HTH transcriptional regulator. Three putative genes encoding dephospho CoA kinase, putative lytic transglycosylase-like protein, and ABC transporter substrate-binding protein (CD630_11290, CD630_11300, and CD630_11310 respectively) are thought to be within one operon. The operon also contains two consecutive genes, one encoding metalloprotease (CD630_28300) and the other a putative adhesin (CD630_28310). Continued work with these genes by constructing mutants is essential to understand their role in haemolysis in *C. difficile*. Also performing

a quantitative reverse transcription PCR (qRT-PCR) to measure the expression of these putative haemolysin genes in the presence and absence of glucose might contribute to better understanding their regulation (Antunes *et al.*, 2012). In addition, further analysis of the putative operon to experimentally identify the promoter and terminator and prove that it is an operon could be performed.

The main focus of this work was on the CD630_28300 gene (designated *atlf*) because the clone carrying this gene showed larger zones of haemolysis compared to the others. The recombinant ATLF was expressed and purified in *E. coli* as it could provide potential insight into the activity of this protein and effect on erythrocytes. However no activity under the conditions tested was found. Pull-down assays can be used to determine the protein's specific target on the membrane of erythrocytes. Although the haemolysis assay is generally the best quantitative method to determine haemolysis by bacteria and their toxins, this work suggest that the parameters of the assay were not optimal for proteins and therefore requires further work.

The exact role of the *atlf* in the haemolysis of *C. difficile* has been investigated by constructing a mutant. A new genetic method using a CRISPR-Cas9 based plasmid for the selection of a rare homologous recombination event was developed and employed to create a deletion mutant. The initial testing of the haemolytic activity of two mutants having the same deletion revealed different phenotypes. However, after complementation, the haemolysis assay showed that the non-haemolytic

mutant reverted to a haemolytic phenotype. A hypothesis was proposed that haemolysis in *C. difficile* is encoded by more than one gene and the mechanism may be controlled by phase variation. Future work should aim on discovering inverted repeats and chimeric sequences to localise inversions near haemolysin genes using advanced sequence methods for example high-throughput paired-end sequencing followed by *in silico* analysis, orientation-specific PCR, qRT-PCR, and use of a fluorescent gene-reporter system (Goldberg *et al.*, 2014; Kuwahara *et al.*, 2004; Sekulovic *et al.*, 2018). And to determine whether the haemolysin genes are under the control of the flagellar switch by measuring their expression in *flg* ON and OFF isolates using qRT-PCR.

The CRISPR-Cas9 plasmid constructed in this work can be further employed for any mutagenesis in *C. difficile* to gain insights into potential virulence factors that will aid in understanding the pathogenesis and dissemination of infection in this pathogen.

Chapter 8

References

Chapter 8 References

Abou Chakra, C. N., McGeer, A., Labbé, A.-C., Simor, A. E., Gold, W. L., Muller, M. P., Powis, J., Katz, K., Garneau, J. R., Fortier, L.-C., Pépin, J., Cadarette, S. M., & Valiquette, L. (2015). Factors associated with complications of *Clostridium difficile* infection in a multicenter prospective cohort. *Clinical Infectious Diseases, 61*(12), 1781-1788.

Abt, M. C., McKenney, P. T., & Pamer, E. G. (2016). *Clostridium difficile* colitis: pathogenesis and host defence. *Nature Reviews Microbiology, 14*, 609.

Actelion Ltd. (2017). Update on the phase III IMPACT program with cadazolid in CDAD. Retrieved from https://www.actelion.com/media/media-releases?newsId=2111437

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications, 9*(1), 1911.

Ajdic, D., McShan, W. M., McLaughlin, R. E., Savic, G., Chang, J., Carson, M. B., Primeaux, C., Tian, R., Kenton, S., Jia, H., Lin, S., Qian, Y., Li, S., Zhu, H., Najar, F., Lai, H., White, J., Roe, B. A., & Ferretti, J. J. (2002). Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *Proceedings of the National Academy of Sciences of the United States of America, 99*(22), 14434-14439.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Robert s, K., & Walter, P. (2014). Chapter 3: The Shape and Structure of Proteins *Molecular Biology of the Cell* (6th ed.): New York: Garland Science.

Albesa-Jové, D., Bertrand, T., Carpenter, E. P., Swain, G. V., Lim, J., Zhang, J., Haire, L. F., Vasisht, N., Braun, V., Lange, A., von Eichel-Streiber, C., Svergun, D. I., Fairweather, N. F., & Brown, K. A. (2010). Four distinct structural domains in Clostridium difficile toxin B visualized using SAXS. *Journal of Molecular Biology, 396*(5), 1260-1270.

Alouf, J. (2003). Molecular features of the cytolytic pore-forming bacterial protein toxins. *Folia Microbiologica, 48*(1), 5-16.

Alouf, J. E. (2005). *26- Paradigms and classification of bacterial membrane-damaging toxins. The comprehensive sourcebook of bacterial protein toxins* (3rd ed.): Academic Press.

Alouf, J. E., Ladant, D., & Popoff, M. R. (2015). *The Comprehensive Sourcebook of Bacterial Protein Toxins* (4 ed.): Elsevier Science.

Andreeva, A., Howorth, D., Chothia, C., Kulesha, E., & Murzin, A. G. (2014). SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Research, 42*(Database issue), D310-314.

Anjuwon-Foster, B. R., & Tamayo, R. (2017). A genetic switch controls the production of flagella and toxins in *Clostridium difficile*. *PLoS genetics, 13*(3), e1006701.

Anjuwon-Foster, B. R., & Tamayo, R. (2018). Phase variation of *Clostridium difficile* virulence factors. *Gut Microbes, 9*(1), 76-83.

Antunes, A., Camiade, E., Monot, M., Courtois, E., Barbut, F., Sernova, N. V., Rodionov, D. A., Martin-Verstraete, I., & Dupuy, B. (2012). Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. *Nucleic Acids Research, 40*(21), 10701-10718.

Antunes, A., Martin-Verstraete, I., & Dupuy, B. (2011). CcpA-mediated repression of *Clostridium difficile* toxin gene expression. *Molecular Microbiology, 79*(4), 882-899.

Archibald, L. K., Banerjee, S. N., & Jarvis, W. R. (2004). Secular trends in hospital-acquired *Clostridium difficile* disease in the United States, 1987-2001. *Journal of Infectious Diseases, 189*(9), 1585-1589.

Arnau, J., Lauritzen, C., Petersen, G. E., & Pedersen, J. (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expression and Purification, 48*(1), 1-13.

Arvinte, T., Cudd, A., Schulz, B., & Nicolau, C. (1989). Low-pH association of proteins with the membranes of intact red blood cells. II. Studies of the mechanism *Biochimica et Biophysica Acta (BBA) - Biomembranes, 981*(1), 61-68.

Asenjo, J. A., & Andrews, B. A. (2009). Protein purification using chromatography: selection of type, modelling and optimization of operating conditions. *Journal of Molecular Recognition, 22*(2), 65-76.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics, 25*(1), 25-29.

Aubry, A., Hussack, G., Chen, W., KuoLee, R., Twine, S. M., Fulton, K. M., Foote, S., Carrillo, C. D., Tanha, J., & Logan, S. M. (2012). Modulation of toxin production by the flagellar regulon in *Clostridium difficile*. *Infection and Immunity, 80*(10), 3521-3532.

Ausiello, C. M., Cerquetti, M., Fedele, G., Spensieri, F., Palazzo, R., Nasso, M., Frezza, S., & Mastrantonio, P. (2006). Surface layer proteins from *Clostridium difficile* induce inflammatory and regulatory cytokines in human monocytes and dendritic cells. *Microbes and Infection, 8*(11), 2640-2646.

Austin, M., Mellow, M., & Tierney, W. M. (2014). Fecal microbiota transplantation in the treatment of *Clostridium difficile* infections. *The American Journal of Medicine, 127*(6), 479-483.

Awad, M. M., Bryant, A. E., Stevens, D. L., & Rood, J. I. (1995). Virulence studies on chromosomal alpha-toxin and theta-toxin mutants constructed by allelic exchange provide genetic evidence for the essential role of alpha-toxin in *Clostridium perfringens*-mediated gas gangrene. *Molecular Microbiology, 15*(2), 191-202.

Ba-Thein, W., Lyristis, M., Ohtani, K., Nisbet, I. T., Hayashi, H., Rood, J. I., & Shimizu, T. (1996). The virR/virS locus regulates the transcription of genes encoding extracellular toxin production in *Clostridium perfringens*. *Journal of Bacteriology, 178*(9), 2514-2520.

Baban, S. T., Kuehne, S. A., Barketi-Klai, A., Cartman, S. T., Kelly, M. L., Hardie, K. R., Kansau, I., Collignon, A., & Minton, N. P. (2013). The role of flagella in *Clostridium difficile* pathogenesis: comparison between a non-epidemic and an epidemic strain. *PLoS One, 8*(9), e73026.

Bakker, D., Buckley, A. M., de Jong, A., van Winden, V. J., Verhoeks, J. P., Kuipers, O. P., Douce, G. R., Kuijper, E. J., Smits, W. K., & Corver, J. (2014). The HtrA-like protease CD3284 modulates virulence of *Clostridium difficile*. *Infection and Immunity, 82*(10), 4222-4232.

Bakker, D., Smits, W. K., Kuijper, E. J., & Corver, J. (2012). TcdC does not significantly repress toxin expression in *Clostridium difficile* 630DeltaErm. *PLoS One, 7*(8), e43247.

Bakri, M. M., Brown, D. J., Butcher, J. P., & Sutherland, A. D. (2009). *Clostridium difficile* in ready-to-eat salads, Scotland. *Emerging Infectious Diseases, 15*(5), 817-818.

Barketi-Klai, A., Hoys, S., Lambert-Bordes, S., Collignon, A., & Kansau, I. (2011). Role of fibronectin-binding protein A in *Clostridium difficile* intestinal colonization. *Journal of Medical Microbiology, 60*(Pt 8), 1155-1161.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science, 315*(5819), 1709-1712.

Barth, H., Aktories, K., Popoff, M. R., & Stiles, B. G. (2004). Binary bacterial toxins: biochemistry, biology, and applications of common *Clostridium* and *Bacillus* proteins. *Microbiology and Molecular Biology Reviews, 68*(3), 373-402, table of contents.

Barth, H., Pfeifer, G., Hofmann, F., Maier, E., Benz, R., & Aktories, K. (2001). Low pH-induced formation of ion channels by *Clostridium difficile* toxin B in target cells. *The Journal of Biological Chemistry, 276*(14), 10670-10676.

Bartlett, J. G. (2017). Bezlotoxumab - A new agent for *Clostridium difficile* infection. *The New England Journal of Medicine, 376*(4), 381-382.

Bartlett, J. G., Moon, N., Chang, T. W., Taylor, N., & Onderdonk, A. B. (1978). Role of *Clostridium difficile* in antibiotic associated pseudomembranous colitis. *Gastroenterology, 75*(5), 778-782.

Batah, J., Kobeissy, H., Bui Pham, P. T., Denève-Larrazet, C., Kuehne, S., Collignon, A., Janoir-Jouveshomme, C., Marvaud, J.-C., & Kansau, I. (2017). *Clostridium difficile* flagella induce a pro-inflammatory response in intestinal epithelium of mice in cooperation with toxins. *Scientific Reports, 7*(1), 3256.

Bates, C. S., Montanez, G. E., Woods, C. R., Vincent, R. M., & Eichenbaum, Z. (2003). Identification and characterization of a *Streptococcus pyogene*s operon involved in binding of hemoproteins and acquisition of iron. *Infection and Immunity, 71*(3), 1042-1055.

Baverud, V. (2002). *Clostridium difficile* infections in animals with special reference to the horse. A review. *Veterinary Quarterly, 24*(4), 203-219.

Benz, R. (2015). 21 - Basic mechanism of pore-forming toxins A2 - Alouf, Joseph. In D. Ladant & M. R. Popoff (Eds.), *The Comprehensive Sourcebook of Bacterial Protein Toxins (Fourth Edition)* (pp. 605-626). Boston: Academic Press.

Berger, S. (2018). *GIDEON Guide to Medically Important Bacteria. E-Book Series*. LA, California, USA: Gideon Informatics, Inc.

Bilverstone, T. W., Kinsmore, N. L., Minton, N. P., & Kuehne, S. A. (2017). Development of *Clostridium difficile* R20291DeltaPaLoc model strains and in vitro methodologies reveals CdtR is required for the production of CDT to cytotoxic levels. *Anaerobe, 44*, 51-54.

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology, 151*(Pt 8), 2551-2561.

Bornhorst, J. A., & Falke, J. J. (2000). Purification of proteins using polyhistidine affinity tags. *Methods in Enzymology, 326*, 245-254.

Borriello, S. P., Davies, H. A., & Barclay, F. E. (1988). Detection of fimbriae amongst strains of *Clostridium difficile*. *FEMS Microbiology Letters, 49*(1), 65-67.

Bradshaw, W. J., Kirby, J. M., Thiyagarajan, N., Chambers, C. J., Davies, A. H., Roberts, A. K., Shone, C. C., & Acharya, K. R. (2014). The structure of the cysteine protease and lectin-like domains of Cwp84, a surface layer-associated protein from *Clostridium difficile*. *Acta Crystallographica D, Biological Crystallography, 70*(7), 1983-1993.

Bragg, T. S., & Robertson, D. L. (1989). Nucleotide sequence and analysis of the lethal factor gene (lef) from *Bacillus anthracis*. *Gene, 81*(1), 45-54.

Braun, V., Hundsberger, T., Leukel, P., Sauerborn, M., & von Eichel-Streiber, C. (1996). Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. *Gene, 181*(1-2), 29-38.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V., & van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science, 321*(5891), 960-964.

Brouwer, M. S., Allan, E., Mullany, P., & Roberts, A. P. (2012). Draft genome sequence of the nontoxigenic *Clostridium difficile* strain CD37. *Journal of Bacteriology, 194*(8), 2125-2126.

Brouwer, M. S. M., Roberts, A. P., Hussain, H., Williams, R. J., Allan, E., & Mullany, P. (2013). Horizontal gene transfer converts non-toxigenic *Clostridium difficile* strains into toxin producers. *Nature Communications, 4*, 2601.

Brouwer, M. S. M., Warburton, P. J., Roberts, A. P., Mullany, P., & Allan, E. (2011). Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of Clostridium difficile. *PLoS One, 6*(8), e23014.

Bruder, M. R., Pyne, M. E., Moo-Young, M., Chung, D. A., & Chou, C. P. (2016). Extending CRISPR-Cas9 Technology from Genome Editing to

Transcriptional Engineering in the Genus Clostridium. *Applied and Environmental Microbiology, 82*(20), 6109-6119.

Burns, D. A., Heeg, D., Cartman, S. T., & Minton, N. P. (2011). Reconsidering the sporulation characteristics of hypervirulent *Clostridium difficile* BI/NAP1/027. *PLoS One, 6*(9), e24894.

Burrowes, B., Harper, D., Anderson, J., McConville, M., & Enright, M. (2011). Bacteriophage therapy: potential uses in the control of antibiotic-resistant pathogens (Vol. 9, pp. 775-785). London: Informa Healthcare.

Buxton, R. (2005). Image 2881. Blood agar plates and hemolysis: *Streptococcus* and other catalase negative Gram-positive cocci. ASM Microbe Library: http://www.asmscience.org/content/education/imagegallery/image.2881.

Cafardi, V., Biagini, M., Martinelli, M., Leuzzi, R., Rubino, J. T., Cantini, F., Norais, N., Scarselli, M., Serruto, D., & Unnikrishnan, M. (2013). Identification of a novel zinc metalloprotease through a global analysis of *Clostridium difficile* extracellular proteins. *PLoS One, 8*(11), e81306.

Calabi, E., Calabi, F., Phillips, A. D., & Fairweather, N. F. (2002). Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues. *Infection and Immunity, 70*(10), 5770-5778.

Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G., & Fairweather, N. (2001). Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Molecular Microbiology, 40*(5), 1187-1199.

Carquin, M., Pollet, H., Veiga-da-Cunha, M., Cominelli, A., Van Der Smissen, P., N'kuli, F., Emonard, H., Henriet, P., Mizuno, H., Courtoy, P. J., & Tyteca, D. (2014). Endogenous sphingomyelin segregates into submicrometric domains in the living erythrocyte membrane. *Journal of Lipid Research, 55*(7), 1331–1342.

Carson, M., Johnson, D. H., McDonald, H., Brouillette, C., & Delucas, L. J. (2007). His-tag impact on structure. *Biological Crystallography, 63*(Pt 3), 295-301.

Carter, G. P., Chakravorty, A., Pham Nguyen, T. A., Mileto, S., Schreiber, F., Li, L., Howarth, P., Clare, S., Cunningham, B., Sambol, S. P., Cheknis, A., Figueroa, I., Johnson, S., Gerding, D., Rood, J. I., Dougan, G., Lawley, T. D., & Lyras, D. (2015). Defining the Roles of TcdA and TcdB in Localized Gastrointestinal Disease, Systemic Organ Damage, and the Host Response during *Clostridium difficile* Infections. *mBio, 6*(3), e00551.

Carter, G. P., Lyras, D., Allen, D. L., Mackin, K. E., Howarth, P. M., O'Connor, J. R., & Rood, J. I. (2007). Binary toxin production in *Clostridium difficile* is regulated by CdtR, a LytTR family response regulator. *Journal of Bacteriology, 189*(20), 7290-7301.

Carter, G. P., Purdy, D., Williams, P., & Minton, N. P. (2005). Quorum sensing in *Clostridium difficile:* analysis of a luxS-type signalling system. *Journal of Medical Microbiology, 54*(Pt 2), 119-127.

Cartman, S. T., Heap, J. T., Kuehne, S. A., Cockayne, A., & Minton, N. P. (2010). The emergence of 'hypervirulence' in *Clostridium difficile*. *International Journal of Medical Microbiology, 300*(6), 387-395.

Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T., & Minton, N. P. (2012). Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. *Applied and Environmental Microbiology, 78*(13), 4683-4690.

Cartman, S. T., & Minton, N. P. (2010). A mariner-based transposon system for in vivo random mutagenesis of *Clostridium difficile*. *Applied and Environmental Microbiology, 76*(4), 1103-1109.

Cassidy, S. K., & O'Riordan, M. X. (2013). More than a pore: the cellular response to cholesterol-dependent cytolysins. *Toxins (Basel), 5*(4), 618-636.

CDC. (2013). Antibiotic resistance threats in the United States, 2013. Retrieved from http://www.cdc.gov/drugresistance/threat-report-2013/pdf/ar-threats-2013-508.pdf

CDC. (2015). Annual report for the emerging infections program for *Clostridium difficile* infection. Retrieved from https://www.cdc.gov/hai/eip/Annual-CDI-Report-2015.html

Cerquetti, M., Molinari, A., Sebastianelli, A., Diociaiuti, M., Petruzzelli, R., Capo, C., & Mastrantonio, P. (2000). Characterization of surface layer proteins from different *Clostridium difficile* clinical isolates. *Microbial Pathogenesis, 28*(6), 363-372.

Chandrasekaran, R., Kenworthy, A. K., & Lacy, D. B. (2016). *Clostridium difficile* toxin A undergoes clathrin-independent, PACSIN2-dependent endocytosis. *PLoS Pathogens, 12*(12), e1006070.

Chen, P., Tao, L., Wang, T., Zhang, J., He, A., Lam, K.-h., Liu, Z., He, X., Perry, K., Dong, M., & Jin, R. (2018). Structural basis for recognition of frizzled proteins by *Clostridium difficile* toxin B. *Science, 360*(6389), 664-669.

Chen, S.-C., Huang, C.-H., Lai, S.-J., Yang, C. S., Hsiao, T.-H., Lin, C.-H., Fu, P.-K., Ko, T.-P., & Chen, Y. (2016). Mechanism and inhibition of human UDP-GlcNAc 2-epimerase, the key enzyme in sialic acid biosynthesis. *Scientific Reports, 6*, 23274.

Christie, P. J., Korman, R. Z., Zahler, S. A., Adsit, J. C., & Dunny, G. M. (1987). Two conjugation systems associated with *Streptococcus faecalis* plasmid pCF10: identification of a conjugative transposon that transfers between *S. faecalis* and *Bacillus subtilis*. *Journal of Bacteriology, 169*(6), 2529-2536.

Ciric, L., Mullany, P., & Roberts, A. P. (2011). Antibiotic and antiseptic resistance genes are linked on a novel mobile genetic element: Tn6087. *Journal of Antimicrobial Chemotherapy, 66*(10), 2235-2239.

Claro, T., Daniels, S., & Humphreys, H. (2014). Detecting Clostridium difficile spores from inanimate surfaces of the hospital environment: which method is best? *Journal of Clinical Microbiology, 52*(9), 3426-3428.

Cohen, S. H., Tang, Y. J., & Silva, J. (2000). Analysis of the pathogenicity locus in Clostridium difficile strains. *Journal of Infectious Diseases, 181*(2), 659-663.

Collery, M. M., Kuehne, S. A., McBride, S. M., Kelly, M. L., Monot, M., Cockayne, A., Dupuy, B., & Minton, N. P. (2017). What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence and phenotypes of *Clostridium difficile* strain 630 and derivatives. *Virulence, 8*(6), 767-781.

Collins, L. E., Lynch, M., Marszalowska, I., Kristek, M., Rochfort, K., O'Connell, M., Windle, H., Kelleher, D., & Loscher, C. E. (2014). Surface layer proteins isolated from *Clostridium difficile* induce clearance responses in macrophages. *Microbes and Infection, 16*(5), 391-400.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science, 339*(6121), 819-823.

Cooper, G. (2000). Cells As Experimental Models *The Cell: A Molecular Approach.* (2 ed.). Sunderland (MA): Sinauer Associates. available at: https://www.ncbi.nlm.nih.gov/books/NBK9917/.

Cossart, P., Vicente, M. F., Mengaud, J., Baquero, F., Perez-Diaz, J. C., & Berche, P. (1989). Listeriolysin O is essential for virulence of *Listeria monocytogenes*: Direct evidence obtained by gene complementation. *Infection and Immunity, 57*(11), 3629-3636.

Cotter, P. D., Draper, L. A., Lawton, E. M., Daly, K. M., Groeger, D. S., Casey, P. G., Ross, R. P., & Hill, C. (2008). Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I *Listeria monocytogenes*. *PLoS Pathogens, 4*(9), e1000144.

Cowardin, C. A., Buonomo, E. L., Saleh, M. M., Wilson, M. G., Burgess, S. L., Kuehne, S. A., Schwan, C., Eichhoff, A. M., Koch-Nolte, F., Lyras, D., Aktories, K., Minton, N. P., & Petri Jr, W. A. (2016a). The binary toxin CDT enhances *Clostridium difficile* virulence by suppressing protective colonic eosinophilia. *Nature Microbiology, 1*, 16108.

Cowardin, C. A., Jackman, B. M., Noor, Z., Burgess, S. L., Feig, A. L., & Petri, W. A., Jr. (2016b). Glucosylation Drives the Innate Inflammatory Response to *Clostridium difficile* Toxin A. *Infection and Immunity, 84*(8), 2317-2323.

Coykendall, A. L. (1977). Proposal to elevate the subspecies of *Streptococcus mutans* to species status, based on their molecular composition. *International Journal of Systematic Bacteriology, 27*, 26-30.

Cupp-Enyard, C. (2008). Sigma's non-specific protease activity assay - casein as a substrate. *Journal of visualized experiments : JoVE*(19), 899.

D'Herelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysent´eriques. *Comptes Rendus de l'Académie des Sciences, 165*, 373–375.

Dal Peraro, M., & van der Goot, F. G. (2016). Pore-forming toxins: ancient, but never really out of fashion. *Nature Reviews Microbiology, 14*(2), 77-92.

Daley, P., Louie, T., Lutz, J. E., Khanna, S., Stoutenburgh, U., Jin, M., Adedoyin, A., Chesnel, L., Guris, D., Larson, K. B., & Murata, Y. (2017). Surotomycin versus vancomycin in adults with *Clostridium difficile* infection: primary clinical outcomes from the second pivotal, randomized, double-blind, Phase 3 trial. *Journal of Antimicrobial Chemotherapy, 72*(12), 3462-3470.

Dannheim, H., Riedel, T., Neumann-Schaal, M., Bunk, B., Schober, I., Sproer, C., Chibani, C. M., Gronow, S., Liesegang, H., Overmann, J., & Schomburg, D. (2017). Manual curation and reannotation of the genomes of *Clostridium difficile* 630Deltaerm and *C. difficile* 630. *Journal of Medical Microbiology, 66*(3), 286-293.

Darkoh, C., DuPont, H. L., Norris, S. J., & Kaplan, H. B. (2015). Toxin synthesis by *Clostridium difficile* is regulated through quorum signaling. *mBio, 6*(2), e02569.

Datta, A. R., & Kothary, M. H. (1993). Effects of glucose, growth temperature, and pH on listeriolysin O production in *Listeria monocytogenes*. *Applied and Environmental Microbiology, 59*(10), 3495-3497.

Dembek, M., Barquist, L., Boinett, C. J., Cain, A. K., Mayho, M., Lawley, T. D., Fairweather, N. F., & Fagan, R. P. (2015). High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *mBio, 6*(2).

Deshpande, R. G., Khan, M. B., Bhat, D. A., & Navalkar, R. G. (1997). Isolation of a contact-dependent haemolysin from *Mycobacterium tuberculosis*. *Bacterial Pathogenicity, 46*, 233-238.

Dickson, I. (2018). Oral capsule FMT effective for *C. difficile* infection. *Nature Reviews Gastroenterology and Hepatology, 15*, 68.

Dineen, S. S., McBride, S. M., & Sonenshein, A. L. (2010). Integration of metabolism and virulence by *Clostridium difficile* CodY. *Journal of Bacteriology, 192*(20), 5350-5362.

Dineen, S. S., Villapakkam, A. C., Nordman, J. T., & Sonenshein, A. L. (2007). Repression of *Clostridium difficile* toxin gene expression by CodY. *Molecular Microbiology, 66*(1), 206-219.

Dinges, M. M., Orwin, P. M., & Schlievert, P. M. (2000). Exotoxins of *Staphylococcus aureus*. *Clinical Microbiology Reviews, 13*(1), 16.

Dingle, T. C., Mulvey, G. L., & Armstrong, G. D. (2011). Mutagenic analysis of the *Clostridium difficile* flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters. *Infection and Immunity, 79*(10), 4061-4067.

Dobson, G., Hickey, C., & Trinder, J. (2003). *Clostridium difficile* colitis causing toxic megacolon, severe sepsis and multiple organ dysfunction syndrome. *Intensive Care Medicine, 29*(6), 1030.

Doerks, T., Copley, R. R., Schultz, J., Ponting, C. P., & Bork, P. (2002). Systematic identification of novel protein domain families associated with nuclear functions. *Genome Research, 12*(1), 47-56.

Doudna, J. A., & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science, 346*(6213), 1258096.

Dumford, D. M., Nerandzic, M. M., Eckstein, B. C., & Donskey, C. J. (2009). What is on that keyboard? Detecting hidden environmental reservoirs of *Clostridium difficile* during an outbreak associated with North American

pulsed-field gel electrophoresis type 1 strains. *American Journal of Infection Control, 37*(1), 15-19.

Duncan, J. L. (1974). Characteristics of streptolysin O hemolysis: kinetics of hemoglobin and ^{86}rubidium release. *Infection and Immunity, 9*(6), 1022-1027.

Duncan, J. L., & Cho, G. J. (1972). Production of Staphylococcal Alpha Toxin II. Glucose Repression of Toxin Formation. *Infection and Immunity, 6*(5), 689-694.

Duncan, J. L., & Mason, L. (1976). Characteristics of streptolysin S hemolysis. *Infection and Immunity, 14*(1), 77-82.

Dupuy, B., & Sonenshein, A. L. (1998). Regulated transcription of *Clostridium difficile* toxin genes. *Molecular Microbiology, 27*(1), 107-120.

Eckert, C., Burghoffer, B., & Barbut, F. (2013). Contamination of ready-to-eat raw vegetables with *Clostridium difficile* in France. *Journal of Medical Microbiology, 62*(Pt 9), 1435-1438.

Eckert, C., Emirian, A., Le Monnier, A., Cathala, L., De Montclos, H., Goret, J., Berger, P., Petit, A., De Chevigny, A., Jean-Pierre, H., Nebbad, B., Camiade, S., Meckenstock, R., Lalande, V., Marchandin, H., & Barbut, F. (2015). Prevalence and pathogenicity of binary toxin-positive *Clostridium difficile* strains that do not produce toxins A and B. *New Microbes and New Infection, 3*, 12-17.

Edwards, A. N., Suarez, J. M., & McBride, S. M. (2013). Culturing and maintaining *Clostridium difficile* in an anaerobic environment. *Journal of visualized experiments : JoVE*(79), e50787-e50787.

Egerer, M., Giesemann, T., Jank, T., Satchell, K. J., & Aktories, K. (2007). Auto-catalytic cleavage of *Clostridium difficile* toxins A and B depends on cysteine protease activity. *Journal of Biological Chemistry, 282*(35), 25314-25321.

Eichenbaum, Z., Muller, E., Morse, S. A., & Scott, J. R. (1996). Acquisition of iron from host proteins by the group A *Streptococcus*. *Infection and Immunity, 64*(12), 5428-5429.

Eiffler, I., Behnke, J., Ziesemer, S., Muller, C., & Hildebrandt, J. P. (2016). *Staphylococcus aureus* alpha-toxin-mediated cation entry depolarizes membrane potential and activates p38 MAP kinase in airway epithelial cells. *American Journal of Physiology Lung Cell and Molecular Physiology, 311*(3), L676-685.

El Meouche, I., Peltier, J., Monot, M., Soutourina, O., Pestel-Caron, M., Dupuy, B., & Pons, J. L. (2013). Characterization of the SigD regulon of *C. difficile* and its positive control of toxin production through the regulation of tcdR. *PLoS One, 8*(12), e83748.

el Rassi, Z., & Horváth, C. (1986). Metal chelate-interaction chromatography of proteins with iminodiacetic acid-bonded stationary phases on silica support. *Journal of Chromatography, 359*, 241-253.

Elixhauser, A., & Jhung, M. (2006). *Clostridium difficile*-Associated Disease in U.S. Hospitals, 1993-2005: Statistical Brief #50 *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Rockville (MD): Agency for Healthcare Research and Quality (US).

Emerson, J. E., Reynolds, C. B., Fagan, R. P., Shaw, H. A., Goulding, D., & Fairweather, N. F. (2009). A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. *Molecular Microbiology, 74*(3), 541-556.

Eyre, D. W., Tracey, L., Elliott, B., Slimings, C., Huntington, P. G., Stuart, R. L., Korman, T. M., Kotsiou, G., McCann, R., Griffiths, D., Fawley, W. N., Armstrong, P., Dingle, K. E., Walker, A. S., Peto, T. E., Crook, D. W., Wilcox, M. H., & Riley, T. V. (2015). *Emergence and spread of predominantly community-onset Clostridium difficile PCR ribotype 244 infection in Australia, 2010-2012. Euro Surveillance. 20(10):pii=21059*

Ezzell, J. W., Ivins, B. E., & Leppla, S. H. (1984). Immunoelectrophoretic analysis, toxicity, and kinetics of *in vitro* production of the protective antigen and lethal factor components of *Bacillus anthracis* toxin. *Infection and Immunity, 45*(3), 761-767.

Fagan, R. P., Albesa-Jove, D., Qazi, O., Svergun, D. I., Brown, K. A., & Fairweather, N. F. (2009). Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. *Molecular Microbiology, 71*(5), 1308-1322.

Fagan, R. P., & Fairweather, N. F. (2011). *Clostridium difficile* has two parallel and essential Sec secretion systems. *Journal of Biological Chemistry, 286*(31), 27483-27493.

Fagan, R. P., Janoir, C., Collignon, A., Mastrantonio, P., Poxton, I. R., & Fairweather, N. F. (2011). A proposed nomenclature for cell wall proteins of *Clostridium difficile*. *Journal of Medical Microbiology, 60*(Pt 8), 1225-1228.

Farrand, A. J., Hotze, E. M., Sato, T. K., Wade, K. R., Wimley, W. C., Johnson, A. E., & Tweten, R. K. (2015). The Cholesterol-dependent Cytolysin

Membrane-binding Interface Discriminates Lipid Environments of Cholesterol to Support beta-Barrel Pore Insertion. *Journal of Biological Chemistry, 290*(29), 17733-17744.

Faulds-Pain, A., Twine, S. M., Vinogradov, E., Strong, P. C., Dell, A., Buckley, A. M., Douce, G. R., Valiente, E., Logan, S. M., & Wren, B. W. (2014). The post-translational modification of the *Clostridium difficile* flagellin affects motility, cell surface properties and virulence. *Molecular Microbiology, 94*(2), 272-289.

Feher, C., Soriano, A., & Mensa, J. (2017). A review of experimental and off-label therapies for *Clostridium difficile* infection. *Infectious Diseases and Therapy, 6*(1), 1-35.

Fineran, P. C., & Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology, 434*(2), 202-209.

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., Chang, H. Y., Dosztanyi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G. L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale, D. A., Necci, M., Nuka, G., Orengo, C. A., Park, Y., Pesseat, S., Piovesan, D., Potter, S. C., Rawlings, N. D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Sillitoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P. D., Tosatto, S. C., Wu, C. H., Xenarios, I., Yeh, L. S., Young, S. Y., & Mitchell, A. L. (2017). InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Research, 45*(D1), D190-D199.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., & Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research, 44*(D1), D279-285.

Flores-Díaz, M., Monturiol-Gross, L., & Alape-Girón, A. (2015). 22 - Membrane-damaging and cytotoxic sphingomyelinases and phospholipases A2 - Alouf, Joseph. In D. Ladant & M. R. Popoff (Eds.), *The Comprehensive Sourcebook of Bacterial Protein Toxins (Fourth Edition)* (pp. 627-676). Boston: Academic Press.

Flores-Diaz, M., Monturiol-Gross, L., Naylor, C., Alape-Giron, A., & Flieger, A. (2016). Bacterial sphingomyelinases and phospholipases as virulence factors. *Microbiology and Molecular Biology Reviews, 80*(3), 597-628.

Frazier, C. L., San Filippo, J., Lambowitz, A. M., & Mills, D. A. (2003). Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. *Applied and Environmental Microbiology, 69*(2), 1121-1128.

Freeman, J., Bauer, M. P., Baines, S. D., Corver, J., Fawley, W. N., Goorhuis, B., Kuijper, E. J., & Wilcox, M. H. (2010). The changing epidemiology of *Clostridium difficile* infections. *Clinical Microbiology Reviews, 23*(3), 529-549.

Garcia-Arribas, M. L., & Kramer, J. M. (1990). The effect of glucose, starch, and pH on growth, enterotoxin and haemolysin production by strains of *Bacillus cereus* associated with food poisoning and non-gastrointestinal infection *International Journal of Food Microbiology, 11*, 21-34.

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America, 109*(39), E2579-2586.

Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodríguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M. L., Mattanovich, D., & Villaverde, A. (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microbial Cell Factories, 7*, 11-11.

Gaulton, T., Misra, R., Rose, G., Baybayan, P., Hall, R., Freeman, J., Turton, J., Picton, S., Korlach, J., Gharbia, S., & Shah, H. (2015). Complete genome sequence of the hypervirulent bacterium *Clostridium difficile* strain G46, ribotype 027. *Microbiology Resource Announcements, 3*(2), 2.

Genth, H., & Just, I. (2015). 16- Large clostridial glucosylating toxins modifying small GTPases: cellular aspects. In J. E. Alouf, D. Ladant, & M. R. Popoff (Eds.), *The Comprehensive Sourcebook of Bacterial Protein Toxins* (4 ed., pp. 441-462): Elsevier Science.

Geoffroy, C., Mengaud, J., Alouf, J. E., & Cossart, P. (1990). Alveolysin, the thiol-activated toxin of *Bacillus alvei*, is homologous to listeriolysin O, perfringolysin O, pneumolysin, and streptolysin O and contains a single cysteine. *Journal of Bacteriology, 172*(12), 7301-7305.

Gerding, D. N., Johnson, S., Rupnik, M., & Aktories, K. (2014). *Clostridium difficile* binary toxin CDT: mechanism, epidemiology, and potential clinical importance. *Gut Microbes, 5*(1), 15-27.

Gerding, D. N., Muto, C. A., & Owens, J. R. C. (2008). Measures to control and prevent *Clostridium difficile* infection. *Clinical Infectious Diseases, 46*(Supplement_1), S43-S49.

Gil, F., Lagos-Moraga, S., Calderon-Romero, P., Pizarro-Guajardo, M., & Paredes-Sabja, D. (2017). Updates on *Clostridium difficile* spore biology. *Anaerobe, 45*, 3-9.

Gilmore, M. S., Cruz-Rodz, A. L., Leimeister-Wachter, M., Kreft, J., & Goebel, W. (1989). A *Bacillus cereus* cytolytic determinant, cereolysin AB, which comprises the phospholipase C and sphingomyelinase genes: nucleotide sequence and genetic linkage. *Journal of Bacteriology, 171*(2), 744-753.

GO, T. g. o. c. (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Research, 45*(D1), D331-d338.

Goldberg, A., Fridman, O., Ronin, I., & Balaban, N. Q. (2014). Systematic identification and quantification of phase variation in commensal and pathogenic *Escherichia coli*. *Genome Medicine, 6*(11), 112.

Goldfine, H., Johnston, N. C., & Knob, C. (1993). Nonspecific phospholipase C of *Listeria monocytogenes*: activity on phospholipids in Triton X-100-mixed micelles and in biological membranes. *Journal of Bacteriology, 175*(14), 4298-4306.

Goni, F. M., Montes, L. R., & Alonso, A. (2012). Phospholipases C and sphingomyelinases: Lipids as substrates and modulators of enzyme activity. *Progress in Lipid Research, 51*(3), 238-266.

Gould, L. H., & Limbago, B. (2010). *Clostridium difficile* in food and domestic animals: a new foodborne pathogen? *Clinical Infectious Diseases, 51*(5), 577-582.

Govind, R., & Dupuy, B. (2012). Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE. *PLoS Pathogens, 8*(6), e1002727.

Grimsley, G. R., & Pace, C. N. (2003). Spectrophotometric determination of protein concentration. *Current Protocols in Protein Science, 33*(1), 3.1.1-3.1.9.

Hachler, H., Kayser, F. H., & Berger-Bachi, B. (1987). Homology of a transferable tetracycline resistance determinant of *Clostridium difficile* with *Streptococcus (Enterococcus) faecalis* transposon *Tn*916. *Antimicrobial Agents and Chemotherapy, 31*(7), 1033-1038.

Hage, D., Anguizola, J., Li, R., Matsuda, R., Papastavros, E., Pfaunmiller, E., Sobansky, M., Zheng, X., & Fanali, S. (2017). Chapter 12: Affinity chromatography. In S. Fanali, P. R. Haddad, C. Poole, & M. L. Riekkola (Eds.), *Liquid Chromatography: Fundamentals and Instrumentation* (2 ed., Vol. 1, pp. 319-341): Elsevier Inc.

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., & Terns, M. P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell, 139*(5), 945-956.

Hall, I. C., & O'Toole, E. (1935). Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, *Bacillus difficilis*. *American Journal of Diseases of Children, 49*(2), 390-402.

Hammond, G. A., & Johnson, J. L. (1995). The toxigenic element of *Clostridium difficile* strain VPI 10463. *Microbial Pathogenesis, 19*(4), 203-213.

Hanson, M. S., Slaughter, C., & Hansen, E. J. (1992). The hbpA gene of *Haemophilus influenzae* type b encodes a heme-binding lipoprotein conserved among heme-dependent *Haemophilus* species. *Infection and Immunity, 60*(6), 2257-2266.

Haque, A., Sugimoto, N., Horiguchi, Y., Okabe, T., Miyata, T., Iwanaga, S., & Matsuda, m. (1992). Production, purification, and characterization of botulinolysin, a thiol-activated hemolysin of *Clostridium botulinum*. *Infection and Immunity, 60*(1), 71-78.

Haraldsen, J. D., & Sonenshein, A. L. (2003). Efficient sporulation in *Clostridium difficile* requires disruption of the sigmaK gene. *Molecular Microbiology, 48*(3), 811-821.

Hardegree, M. C., Palmer, A. E., & Duffin, N. (1971). Tetanolysin: in-vivo effects in animals. *Journal of Infectious Diseases, 123*(1), 51-60.

Hargreaves, K. R., & Clokie, M. R. (2014). *Clostridium difficile* phages: still difficult? *Frontiers in Microbiology, 5*, 184.

Hatheway, C. L. (1990). Toxigenic Clostridia. *Clinical Microbiology Reviews, 3*(1), 66-98.

He, H., Puerta, D. T., Cohen, S. M., & Rodgers, K. R. (2005). Structural and spectroscopic study of reactions between chelating zinc-binding groups and mimics of the matrix metalloproteinase and disintegrin metalloprotease catalytic sites: the coordination chemistry of metalloprotease inhibition. *Inorganic Chemistry, 44*(21), 7431-7442.

He, M., Sebaihia, M., Lawley, T. D., Stabler, R. A., Dawson, L. F., Martin, M. J., Holt, K. E., Seth-Smith, H. M. B., Quail, M. A., Rance, R., Brooks, K., Churcher, C., Harris, D., Bentley, S. D., Burrows, C., Clark, L., Corton, C., Murray, V., Rose, G., Thurston, S., van Tonder, A., Walker, D., Wren, B. W., Dougan, G., & Parkhill, J. (2010). Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proceedings of the National Academy of Sciences of the United States of America, 107*(16), 7527-7532.

Health Protection Agency. (2016). *Clostridium difficile* Ribotyping Network for England and Northern Ireland: 2013/15 report. Retrieved from https://www.gov.uk/government/publications/clostridium-difficile-ribotyping-network-cdrn-report

Heap, J. T., Ehsaan, M., Cooksley, C. M., Ng, Y.-K., Cartman, S. T., Winzer, K., & Minton, N. P. (2012). Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Research, 40*(8), e59-e59.

Heap, J. T., Kuehne, S. A., Ehsaan, M., Cartman, S. T., Cooksley, C. M., Scott, J. C., & Minton, N. P. (2010). The ClosTron: Mutagenesis in Clostridium refined and streamlined. *Journal of Microbiological Methods, 80*(1), 49-55.

Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P., & Minton, N. P. (2007). The ClosTron: a universal gene knock-out system for the genus Clostridium. *Journal of Microbiological Methods, 70*(3), 452-464.

Heap, J. T., Pennington, O. J., Cartman, S. T., & Minton, N. P. (2009). A modular system for Clostridium shuttle plasmids. *Journal of Microbiological Methods, 78*(1), 79-85.

Hecker, M. T., Obrenovich, M. E., Cadnum, J. L., Jencson, A. L., Jain, A. K., Ho, E., & Donskey, C. J. (2016). Fecal microbiota transplantation by freeze-dried oral capsules for recurrent *Clostridium difficile* infection. *Open Forum Infectious Diseases, 3*(2), ofw091.

Hemmasi, S., Czulkies, B. A., Schorch, B., Veit, A., Aktories, K., & Papatheodorou, P. (2015). Interaction of the *Clostridium difficile* Binary Toxin CDT and Its Host Cell Receptor, Lipolysis-stimulated Lipoprotein Receptor (LSR). *Journal of Biological Chemistry, 290*(22), 14031-14044.

Hennequin, C., Collignon, A., & Karjalainen, T. (2001). Analysis of expression of GroEL (Hsp60) of *Clostridium difficile* in response to stress. *Microbial Pathogenesis, 31*(5), 255-260.

Hennequin, C., Janoir, C., Barc, M. C., Collignon, A., & Karjalainen, T. (2003). Identification and characterization of a fibronectin-binding protein from *Clostridium difficile*. *Microbiology, 149*(Pt 10), 2779-2787.

Hensbergen, P. J., Klychnikov, O. I., Bakker, D., van Winden, V. J., Ras, N., Kemp, A. C., Cordfunke, R. A., Dragan, I., Deelder, A. M., Kuijper, E. J., Corver, J., Drijfhout, J. W., & van Leeuwen, H. C. (2014). A novel secreted metalloprotease (CD2830) from *Clostridium difficile* cleaves specific proline sequences in LPXTG cell surface proteins. *Molecular and Cellular Proteomics, 13*(5), 1231-1244.

Herbert, M., O'Keeffe, T. A., Purdy, D., Elmore, M., & Minton, N. P. (2003). Gene transfer into *Clostridium difficile* CD630 and characterisation of its methylase genes. *FEMS Microbiology Letters, 229*(1), 103-110.

Hernandez, S., Gomez, A., Cedano, J., & Querol, E. (2009). Bioinformatics annotation of the hypothetical proteins found by omics techniques can help to disclose additional virulence factors. *Current Microbiology, 59*(4), 451-456.

Higa, J. T., & Kelly, C. P. (2014). New drugs and strategies for management of *Clostridium difficile* colitis. *Journal of Intensive Care Medicine, 29*(4), 190-199.

Hinderlich, S., Stasche, R., Zeitler, R., & Reutter, W. (1997). A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Purification and characterization of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase. *The Journal of Biological Chemistry, 272*(39), 24313-24318.

Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., & Stüber, D. (1988). Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Bio/Technology, 6*, 1321.

Hochuli, E., Döbeli, H., & Schacher, A. (1987). New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *Journal of Chromatography A, 411*, 177-184.

Holdeman. (1965). ID#3647. Public Health Image Library (PHIL) Centers for Disease Control and Prevention (CDC) https://phil.cdc.gov/Details.aspx?pid=3647.

Honda, T., Shiba, A., Seo, S., Yamamoto, J., Matsuyama, J., & Miwatani, T. (1991). Identity of hemolysins produced by *Bacillus thuringiensis* and *Bacillus cereus*. *FEMS Microbiology Letters, 63*(2-3), 205-209.

Hong, W., Zhang, J., Cui, G., Wang, L., & Wang, Y. (2018). Multiplexed CRISPR-Cpf1-Mediated Genome Editing in *Clostridium difficile* toward the Understanding of Pathogenesis of *C. difficile* Infection. *ACS Synthetic Biology, 7*(6), 1588-1600.

Horvath, P., Romero, D. A., Coute-Monvoisin, A. C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., & Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *Journal of Bacteriology, 190*(4), 1401-1412.

Hotze, E. M., & Tweten, R. K. (2012). Membrane assembly of the cholesterol-dependent cytolysin pore complex. *Biochimca et Biophysica Acta, 1818*(4), 1028-1038.

Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., Jiang, W., & Gu, Y. (2016a). CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium. *ACS Synthetic Biology, 5*(12), 1355-1361.

Huang, Q., Gershenson, A., & Roberts, M. F. (2016b). Recombinant broad-range phospholipase C from *Listeria monocytogenes* exhibits optimal activity at acidic pH. *Biochimca et Biophysica Acta, 1864*(6), 697-705.

Hundsberger, T., Braun, V., Weidmann, M., Leukel, P., Sauerborn, M., & Eichel-streiber, C. (1997). Transcription analysis of the genes *tcdA-E* of the pathogenicity locus of *Clostridium difficile*. *European Journal of Biochemistry, 244*(3), 735-742.

Huseby, M., Shi, K., Brown, C. K., Digre, J., Mengistu, F., Seo, K. S., Bohach, G. A., Schlievert, P. M., Ohlendorf, D. H., & Earhart, C. A. (2007). Structure and biological activities of beta toxin from *Staphylococcus aureus*. *Journal of Bacteriology, 189*(23), 8719-8726.

Hussain, H. A., Roberts, A. P., & Mullany, P. (2005). Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. *Journal of Medical Microbiology, 54*(Pt 2), 137-141.

Idalia, V.-M. N., & Bernardo, F. (2017). *Escherichia coli* as a model organism and its application in biotechnology. In A. Samie (Ed.), *Escherichia coli: Recent Advances on Physiology, Pathogenesis and Biotechnological Applications*: Intech Open.

Ishii, F. (1960). Studies on the hemolytic property of *Escherichia coli*. *Japanese Journal of Microbiology, 4*(2), 203-218.

Jangi, S., & Lamont, J. T. (2010). Asymptomatic colonization by *Clostridium difficile* in infants: implications for disease in later life. *Journal of Pediatric Gastroenterology and Nutrition, 51*(1), 2-7.

Jank, T., & Aktories, K. (2008). Structure and mode of action of clostridial glucosylating toxins: the ABCD model. *Trends in Microbiology, 16*(5), 222-229.

Jank, T., Giesemann, T., & Aktories, K. (2007). Rho-glucosylating *Clostridium difficile* toxins A and B: new insights into structure and function. *Glycobiology, 17*(4), 15R-22R.

Janoir, C., Pechine, S., Grosdidier, C., & Collignon, A. (2007). Cwp84, a surface-associated protein of *Clostridium difficile*, is a cysteine protease with degrading activity on extracellular matrix proteins. *Journal of Bacteriology, 189*(20), 7174-7180.

Jansen, R., Embden, J. D., Gaastra, W., & Schouls, L. M. (2002a). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology, 43*(6), 1565-1575.

Jansen, R., van Embden, J. D., Gaastra, W., & Schouls, L. M. (2002b). Identification of a novel family of sequence repeats among prokaryotes. *OMICS: A Journal of Integrative Biology, 6*(1), 23-33.

Jenkins, E. M., Njoku-OBI, A. N., & Adams, E. W. (1964). Purification of the soluble hemolysins of *Listeria monocytogenes Journal of Bacteriology, 88*(2), 418-424

Jepson, M., & Titball, R. (2000). Structure and function of clostridial phospholipases C. *Microbes and Infection, 2*(10), 1277-1284.

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology, 31*(3), 233-239.

Jiang, W., Samai, P., & Marraffini, L. A. (2016). Degradation of phage transcripts by CRISPR-associated RNases enables Type III CRISPR-Cas immunity. *Cell, 164*(4), 710-721.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816-821.

Jones, G. (1980). ID#3876. Public Health Image Library (PHIL) Centers for Disease Control and Prevention (CDC) https://phil.cdc.gov/Details.aspx?pid=3876.

Just, I., Selzer, J., Wilm, M., von Eichel-Streiber, C., Mann, M., & Aktories, K. (1995a). Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature, 375*(6531), 500-503.

Just, I., Wilm, M., Selzer, J., Rex, G., von Eichel-Streiber, C., Mann, M., & Aktories, K. (1995b). The enterotoxin from *Clostridium difficile* (ToxA) monoglucosylates the Rho proteins. *Journal of Biological Chemistry, 270*(23), 13932-13936.

Kamboj, M., Khosa, P., Kaltsas, A., Babady, N. E., Son, C., & Sepkowitz, K. A. (2011). Relapse versus reinfection: surveillance of *Clostridium difficile* infection. *Clinical Infectious Diseases, 53*(10), 1003-1006.

Kao, D., Roach, B., Silva, M., Beck, P., Rioux, K., Kaplan, G. G., Chang, H. J., Coward, S., Goodman, K. J., Xu, H., Madsen, K., Mason, A., Wong, G. K., Jovel, J., Patterson, J., & Louie, T. (2017). Effect of oral capsule- vs colonoscopy-delivered fecal microbiota transplantation on recurrent *Clostridium difficile* infection: a randomized clinical trial. *JAMA: Journal of the American Medical Association, 318*(20), 1985-1993.

Karjalainen, T., Waligora-Dupriet, A. J., Cerquetti, M., Spigaglia, P., Maggioni, A., Mauri, P., & Mastrantonio, P. (2001). Molecular and genomic analysis of genes encoding surface-anchored proteins from *Clostridium difficile*. *Infection and Immunity, 69*(5), 3442-3446.

Karlsson, S., Lindberg, A., Norin, E., Burman, L. G., & Åkerlund, T. (2000). Toxins, butyric acid, and other short-chain fatty acids are coordinately expressed and down-regulated by cysteine in *Clostridium difficile*. *Infection and Immunity, 68*(10), 5881-5888.

Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P., & Siksnys, V. (2013). crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biology, 10*(5), 841-851.

Kasendra, M., Barrile, R., Leuzzi, R., & Soriani, M. (2014). *Clostridium difficile* toxins facilitate bacterial colonization by modulating the fence and gate function of colonic epithelium. *Journal Infectious Diseases, 209*(7), 1095-1104.

Kato, H., Kita, H., Karasawa, T., Maegawa, T., Koino, Y., Takakuwa, H., Saikai, T., Kobayashi, K., Yamagishi, T., & Nakamura, S. (2001). Colonisation and transmission of *Clostridium difficile* in healthy individuals

examined by PCR ribotyping and pulsed-field gel electrophoresis. *Journal of Medical Microbiology, 50*(8), 720-727.

Keessen, E. C., Gaastra, W., & Lipman, L. J. (2011). *Clostridium difficile* infection in humans and animals, differences and similarities. *Veterinary Microbiology, 153*(3-4), 205-217.

Kennedy, C. L., Krejany, E. O., Young, L. F., O'Connor, J. R., Awad, M. M., Boyd, R. L., Emmins, J. J., Lyras, D., & Rood, J. I. (2005). The alpha-toxin of *Clostridium septicum* is essential for virulence. *Molecular Microbiology, 57*(5), 1357-1366.

Kim, J., Kim, Y.-M., Koo, B.-S., Chae, Y.-K., & Yoon, M.-Y. (2003). Production and proteolytic assay of lethal factor from *Bacillus anthracis*. *Protein Expression and Purification, 30*(2), 293-300.

Kirby, J. M., Ahern, H., Roberts, A. K., Kumar, V., Freeman, Z., Acharya, K. R., & Shone, C. C. (2009). Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of *Clostridium difficile*. *Journal of Biological Chemistry, 284*(50), 34666-34673.

Kirk, J. A., & Fagan, R. P. (2016). Heat shock increases conjugation efficiency in *Clostridium difficile*. *Anaerobe, 42*, 1-5.

Kirk, J. A., Gebhart, D., Buckley, A. M., Lok, S., Scholl, D., Douce, G. R., Govoni, G. R., & Fagan, R. P. (2017). New class of precision antimicrobials redefines role of *Clostridium difficile* S-layer in virulence and viability. *Science Translational Medicine, 9*(406).

Klichko, V. I., Miller, J., Wu, A., Popov, S. G., & Alibek, K. (2003). Anaerobic induction of *Bacillus anthracis* hemolytic activity. *Biochemical and Biophysical Research Communications, 303*(3), 855-862.

Koon, H. W., Shih, D. Q., Hing, T. C., Yoo, J. H., Ho, S., Chen, X., Kelly, C. P., Targan, S. R., & Pothoulakisa, C. (2013). Human monoclonal antibodies against *Clostridium difficile* toxins A and B inhibit inflammatory and histologic responses to the toxins in human colon and peripheral blood monocytes. *Antimicrobial Agents and Chemotherapy, 57*(7), 3214.

Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology, 37*, 67-78.

Koster, S., van Pee, K., Hudel, M., Leustik, M., Rhinow, D., Kuhlbrandt, W., Chakraborty, T., & Yildiz, O. (2014). Crystal structure of listeriolysin O reveals

molecular details of oligomerization and pore formation. *Nature Communications, 5*, 3690.

Kovacs-Simon, A., Leuzzi, R., Kasendra, M., Minton, N., Titball, R. W., & Michell, S. L. (2014). Lipoprotein CD0873 is a novel adhesin of *Clostridium difficile*. *Journal of Infectious Diseases, 210*(2), 274-284.

Kuehne, S. A., Cartman, S. T., Heap, J. T., Kelly, M. L., Cockayne, A., & Minton, N. P. (2010). The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature, 467*(7316), 711-713.

Kuehne, S. A., & Minton, N. P. (2012). ClosTron-mediated engineering of Clostridium. *Bioengineered, 3*(4), 247-254.

Kuijper E J, Coignard B, Brazier J S, Suetens C, Drudy D, Wiuff C, Pituch H, Reichert P, Schneider F, Widmer A F, Olsen K EP, Allerberger F, Notermans D W, Barbut F, Delmée M, Wilcox M, Pearson A, Patel B, Brown D J, Frei R, Akerlund T, Poxton I R, & Peet, T. (2007). Update of *Clostridium difficile*-associated disease due to PCR ribotype 027 in Europe. *Euro Surveillance: Europe's journal on infectious disease surveillance, epidemiology, prevention and control, 12*(6), E1-2.

Kuijper, E. J., Barbut, F., Brazier, J. S., Kleinkauf, N., Eckmanns, T., Lambert, M. L., Drudy, D., Fitzpatrick, F., Wiuff, C., Brown, D. J., Coia, J. E., Pituch, H., Reichert, P., Even, J., Mossong, J., Widmer, A. F., Olsen, K. E., Allerberger, F., Notermans, D. W., Delmee, M., Coignard, B., Wilcox, M., Patel, B., Frei, R., Nagy, E., Bouza, E., Marin, M., Akerlund, T., Virolainen-Julkunen, A., Lyytikainen, O., Kotila, S., Ingebretsen, A., Smyth, B., Rooney, P., Poxton, I. R., & Monnet, D. L. (2008). Update of *Clostridium difficile* infection due to PCR ribotype 027 in Europe, 2008. *Euro Surveillance, 13*(31).

Kumar, S. (2016). *Essentials of Microbiology*: Jaypee Brothers Medical Publishers Ltd.

Kuwahara, T., Yamashita, A., Hirakawa, H., Nakayama, H., Toh, H., Okada, N., Kuhara, S., Hattori, M., Hayashi, T., & Ohnishi, Y. (2004). Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. *Proceedings of the National Academy of Sciences of the United States of America, 101*(41), 14919-14924.

Labbe, A. C., Poirier, L., Maccannell, D., Louie, T., Savoie, M., Beliveau, C., Laverdiere, M., & Pepin, J. (2008). *Clostridium difficile* infections in a Canadian tertiary care hospital before and during a regional epidemic associated with the BI/NAP1/027 strain. *Antimicrobial Agents and Chemotherapy, 52*(9), 3180-3187.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*(5259), 680-685.

Lawley, T. D., Clare, S., Deakin, L. J., Goulding, D., Yen, J. L., Raisen, C., Brandt, C., Lovell, J., Cooke, F., Clark, T. G., & Dougan, G. (2010). Use of purified *Clostridium difficile* spores to facilitate evaluation of health care disinfection regimens. *Applied and Environmental Microbiology, 76*(20), 6895-6900.

Lawley, T. D., Croucher, N. J., Yu, L., Clare, S., Sebaihia, M., Goulding, D., Pickard, D. J., Parkhill, J., Choudhary, J., & Dougan, G. (2009). Proteomic and genomic characterization of highly infectious *Clostridium difficile* 630 spores. *Journal of Bacteriology, 191*(17), 5377-5386.

Lawrence, S. L., Feil, S. C., Morton, C. J., Farrand, A. J., Mulhern, T. D., Gorman, M. A., Wade, K. R., Tweten, R. K., & Parker, M. W. (2015). Crystal structure of *Streptococcus pneumoniae* pneumolysin provides key insights into early steps of pore formation. *Scientific Reports, 5*, 14352.

Lawson, P. A., Citron, D. M., Tyrrell, K. L., & Finegold, S. M. (2016). Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prevot 1938. *Anaerobe, 40*, 95-99.

Le Monnier, A., Zahar, J. R., & Barbut, F. (2014). Update on *Clostridium difficile* infections. *Médecine et Maladies Infectieuses, 44*(8), 354-365.

Leffler, D. A., & Lamont, J. T. (2012). Editorial: not so nosocomial anymore: the growing threat of community-acquired *Clostridium difficile*. *American Journal of Gastroenterology, 107*(1), 96-98.

Letunic, I., & Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research, 46*(D1), D493-D496.

Li, J., Ma, M., Sarker, M. R., & McClane, B. A. (2013). CodY is a global regulator of virulence-associated properties for *Clostridium perfringens* type D strain CN3718. *mBio, 4*(5), e00770-00713.

Lin, D. M., Koskella, B., & Lin, H. C. (2017). Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World Journal of Gastrointestinal Pharmacology and Therapeutics, 8*(3), 162-173.

Liu, J., Cai, C., Guo, Q., Zhang, J., Dong, D., Li, G., Fu, L., Xu, J., & Chen, W. (2013). Secretory expression and efficient purification of recombinant anthrax toxin lethal factor with full biological activity in *E. coli*. *Protein Expression and Purification, 89*(1), 56-61.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). *Molecular Cell Biology. Section 2.3- Folding, Modification, and Degradation of Proteins* (4 ed.): New York: W. H. Freeman.

Lowy, I., Molrine, D. C., Leav, B. A., Blair, B. M., Baxter, R., Gerding, D. N., Nichol, G., Thomas, W. D., Leney, M., Sloan, S., Hay, C. A., & Ambrosino, D. M. (2010). Treatment with monoclonal antibodies against *Clostridium difficile* toxins. *The New England Journal of Medicine, 362*(3), 197-205.

Ludwig W., Schleifer KH., & W.B., W. (2009). Revised road map to the phylum Firmicutes. In Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer K-H, & W. WB (Eds.), *Bergey's Manual of Systematic Bacteriology* (Vol. 3, pp. 1-9). New York, NY: Springer.

Lyras, D., O'Connor, J. R., Howarth, P. M., Sambol, S. P., Carter, G. P., Phumoonna, T., Poon, R., Adams, V., Vedantam, G., Johnson, S., Gerding, D. N., & Rood, J. I. (2009). Toxin B is essential for virulence of *Clostridium difficile*. *Nature, 458*(7242), 1176-1179.

MacCannell, D. R., Louie, T. J., Gregson, D. B., Laverdiere, M., Labbe, A. C., Laing, F., & Henwick, S. (2006). Molecular analysis of *Clostridium difficile* PCR ribotype 027 isolates from Eastern and Western Canada. *Journal of Clinical Microbiology, 44*(6), 2147-2152.

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct, 1*, 7.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., Backofen, R., & Koonin, E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology, 13*(11), 722-736.

Maldarelli, G. A., De Masi, L., von Rosenvinge, E. C., Carter, M., & Donnenberg, M. S. (2014). Identification, immunogenicity, and cross-reactivity of type IV pilin and pilin-like proteins from *Clostridium difficile*. *Pathogens and Disease, 71*(3), 302-314.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science, 339*(6121), 823-826.

Mani, N., & Dupuy, B. (2001). Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proceedings of the National Academy of Sciences of the United States of America, 98*(10), 5844-5849.

Mani, N., Lyras, D., Barroso, L., Howarth, P., Wilkins, T., Rood, J. I., Sonenshein, A. L., & Dupuy, B. (2002). Environmental response and autoregulation of *Clostridium difficile* TxeR, a sigma factor for toxin gene expression. *Journal of Bacteriology, 184*(21), 5971-5978.

Manich, M., Knapp, O., Gibert, M., Maier, E., Jolivet-Reynaud, C., Geny, B., Benz, R., & Popoff, M. R. (2008). *Clostridium perfringens* delta toxin is sequence related to beta toxin, NetB, and *Staphylococcus* pore-forming toxins, but shows functional differences. *PLoS One, 3*(11), e3764.

Marmorek, A. (1895). Le streptocoque et le serum antistreptococcique. *Annales de l'Institut Pasteur de Lille*(9), 593-620.

Martin, J. S. H., Monaghan, T. M., & Wilcox, M. H. (2016). *Clostridium difficile* infection: epidemiology, diagnosis and understanding transmission. *Nature Reviews Gastroenterology & Hepatology, 13*(4), 206-216.

Mason, K. M., Raffel, F. K., Ray, W. C., & Bakaletz, L. O. (2011). Heme utilization by nontypeable *Haemophilus influenzae* is essential and dependent on Sap transporter function. *Journal of Bacteriology, 193*(10), 2527-2535.

Matamouros, S., England, P., & Dupuy, B. (2007). *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. *Molecular Microbiology, 64*(5), 1274-1288.

McAllister, K. N., Bouillaut, L., Kahn, J. N., Self, W. T., & Sorg, J. A. (2017). Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Scientific Reports, 7*(1), 14672.

McDonald, L. C., Gerding, D. N., Johnson, S., Bakken, J. S., Carroll, K. C., Coffin, S. E., Dubberke, E. R., Garey, K. W., Gould, C. V., Kelly, C., Loo, V., Shaklee Sammons, J., Sandora, T. J., & Wilcox, M. H. (2018). Clinical practice guidelines for *Clostridium difficile* infection in adults and children: 2017 update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clinical Infectious Diseases, 66*(7), e1-e48.

McDonald, L. C., Killgore, G. E., Thompson, A., Owens, R. C., Kazakova, S. V., Sambol, S. P., Johnson, S., & Gerding, D. N. (2005). An epidemic, toxin

gene-variant strain of *Clostridium difficile*. *The New England Journal of Medicine, 353*(23), 2433-2441.

McDonel, J. L. (1980). *Clostridium perfringens* toxins (type A, B, C, D, E). *Pharmacology and Therapeutics, 10*(3), 617-655.

McFee, R. B., & Abdelsayed, G. G. (2009). *Clostridium difficile*. *Disease-a-Month, 55*(7), 439-470.

McKee, R. W., Aleksanyan, N., Garrett, E. M., & Tamayo, R. (2018). Type IV pili promote *Clostridium difficile* adherence and persistence in a mouse model of infection. *Infection and Immunity, 86*(5), e00943-00917.

Meader, E., Mayer, M. J., Gasson, M. J., Steverding, D., Carding, S. R., & Narbad, A. (2010). Bacteriophage treatment significantly reduces viable *Clostridium difficile* and prevents toxin production in an in vitro model system. *Anaerobe, 16*(6), 549-554.

Meader, E., Mayer, M. J., Steverding, D., Carding, S. R., & Narbad, A. (2013). Evaluation of bacteriophage therapy to control *Clostridium difficile* and toxin production in an in vitro human colon model system. *Anaerobe, 22*, 25-30.

Merrigan, M., Venugopal, A., Mallozzi, M., Roxas, B., Viswanathan, V. K., Johnson, S., Gerding, D. N., & Vedantam, G. (2010). Human hypervirulent *Clostridium difficile* strains exhibit increased sporulation as well as robust toxin production. *Journal of Bacteriology, 192*(19), 4904-4911.

Merrigan, M. M., Venugopal, A., Roxas, J. L., Anwar, F., Mallozzi, M. J., Roxas, B. A., Gerding, D. N., Viswanathan, V. K., & Vedantam, G. (2013). Surface-layer protein A (SlpA) is a major contributor to host-cell adherence of *Clostridium difficile*. *PLoS One, 8*(11), e78404.

Moens, S., & Vanderleyden, J. (1996). Functions of bacterial flagella. *Critical Reviews in Microbiology, 22*(2), 67-100.

Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology, 155*(Pt 3), 733-740.

Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution, 60*(2), 174-182.

Moller, S., Croning, M. D., & Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics, 17*(7), 646-653.

Monot, M., Boursaux-Eude, C., Thibonnier, M., Vallenet, D., Moszer, I., Medigue, C., Martin-Verstraete, I., & Dupuy, B. (2011). Reannotation of the genome sequence of *Clostridium difficile* strain 630. *Journal of Medical Microbiology, 60*(Pt 8), 1193-1199.

Monot, M., Eckert, C., Lemire, A., Hamiot, A., Dubois, T., Tessier, C., Dumoulard, B., Hamel, B., Petit, A., Lalande, V., Ma, L., Bouchier, C., Barbut, F., & Dupuy, B. (2015). *Clostridium difficile*: New insights into the evolution of the pathogenicity locus. *Scientific Reports, 5*, 15023.

Moono, P., Foster, N. F., Hampson, D. J., Knight, D. R., Bloomfield, L. E., & Riley, T. V. (2016). *Clostridium difficile* infection in production animals and avian species: a review. *Foodborne Pathogens and Disease, 13*(12), 647-655.

Moono, P., Lim, S. C., & Riley, T. V. (2017). High prevalence of toxigenic *Clostridium difficile* in public space lawns in Western Australia. *Scientific Reports, 7*, 41196.

Morgan, H. J., & Neill, J. M. (1924). Methoglobin formation by Sterile Cultures Filtrates of Pneumococcus. *The Journal of Experimental Medicine, 40*(2), 269–279.

Mullany, P., Allan, E., & Roberts, A. P. (2015). Mobile genetic elements in *Clostridium difficile* and their role in genome function. *Research in Microbiology, 166*(4), 361-367.

Mullany, P., Wilks, M., Puckey, L., & Tabaqchali, S. (1994). Gene cloning in *Clostridium difficile* using Tn*916* as a shuttle conjugative transposon. *Plasmid, 31*(3), 320-323.

Mullany, P., Williams, R., Langridge, G. C., Turner, D. J., Whalan, R., Clayton, C., Lawley, T., Hussain, H., McCurrie, K., Morden, N., Allan, E., & Roberts, A. P. (2012). Behavior and target site selection of conjugative transposon Tn*916* in two different strains of toxigenic *Clostridium difficile*. *Applied and Environmental Microbiology, 78*(7), 2147-2153.

Na, X., Kim, H., Moyer, M. P., Pothoulakis, C., & LaMont, J. T. (2008). gp96 is a human colonocyte plasma membrane binding protein for *Clostridium difficile* toxin A. *Infection and Immunity, 76*(7), 2862-2871.

Nakamura, S., Nakashio, S., Yamakawa, K., Tanabe, N., & Nishida, S. (1982). Carbohydrate fermentation by *Clostridium difficile*. *Microbiology and Immunology, 26*(2), 107-111.

Nale, J. Y., Spencer, J., Hargreaves, K. R., Buckley, A. M., Trzepinski, P., Douce, G. R., & Clokie, M. R. (2015). Bacteriophage combinations significantly reduce *Clostridium difficile* growth *in vitro* and proliferation *in vivo*. *Antimicrobial Agents and Chemotherapy, 60*(2), 968-981.

Natarajan, M., Walk, S. T., Young, V. B., & Aronoff, D. M. (2013). A clinical and epidemiological review of non-toxigenic *Clostridium difficile*. *Anaerobe, 22*, 1-5.

Ng, Y. K., Ehsaan, M., Philip, S., Collery, M. M., Janoir, C., Collignon, A., Cartman, S. T., & Minton, N. P. (2013). Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using pyrE alleles. *PLoS One, 8*(2), e56051.

Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. In D. Kihara (Ed.), *Protein Function Prediction: Methods and Protocols* (pp. 59-73). New York, NY: Springer New York.

O'Connor, J. R., Johnson, S., & Gerding, D. N. (2009). *Clostridium difficile* infection caused by the epidemic BI/NAP1/027 strain. *Gastroenterology, 136*(6), 1913-1924.

O'Connor, J. R., Lyras, D., Farrow, K. A., Adams, V., Powell, D. R., Hinds, J., Cheung, J. K., & Rood, J. I. (2006). Construction and analysis of chromosomal *Clostridium difficile* mutants. *Molecular Microbiology, 61*(5), 1335-1351.

O'Reilly, M., de Azavedo, J. C. S., Kennedy, S., & Foster, T. J. (1986). Inactivation of the alpha-haemolysin gene of *Staphylococcus aureus* 8325-4 by site-directed mutagenesis and studies on the expression of its haemolysins. *Microbial Pathogenesis, 1*, 125-138.

Ochi, S., Oda, M., Matsuda, H., Ikari, S., & Sakurai, J. (2004). *Clostridium perfringens* alpha-toxin activates the sphingomyelin metabolism system in sheep erythrocytes. *Journal of Biological Chemistry, 279*(13), 12181-12189.

Oda, M., Takahashi, M., Matsuno, T., Uoo, K., Nagahama, M., & Sakurai, J. (2010). Hemolysis induced by *Bacillus cereus* sphingomyelinase. *Biochimica et Biophysica Acta, 1798*(6), 1073-1080.

Ohtani, K., & Shimizu, T. (2016). Regulation of toxin production in *Clostridium perfringens*. *Toxins (Basel), 8*(7).

Olling, A., Huls, C., Goy, S., Muller, M., Krooss, S., Rudolf, I., Tatge, H., & Gerhard, R. (2014). The combined repetitive oligopeptides of *Clostridium difficile* toxin A counteract premature cleavage of the glucosyl-transferase domain by stabilizing protein conformation. *Toxins (Basel), 6*(7), 2162-2176.

Orrell, K. E., Zhang, Z., Sugiman-Marangos, S. N., & 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), 461-473.

Palmer, M. (2001). The family of thiol-activated, cholesterol-binding cytolysins. *Toxicon, 39*, 1681-1689.

Papaparaskevas, J., Houhoula, D. P., Papadimitriou, M., Saroglou, G., Legakis, N. J., & Zerva, L. (2004). Ruling out *Bacillus anthracis*. *Emerging Infectious Diseases, 10*(4), 732-735.

Papatheodorou, P., Carette, J. E., Bell, G. W., Schwan, C., Guttenberg, G., Brummelkamp, T. R., & Aktories, K. (2011). Lipolysis-stimulated lipoprotein receptor (LSR) is the host receptor for the binary toxin *Clostridium difficile* transferase (CDT). *Proceedings of the National Academy of Sciences of the United States of America, 108*(39), 16422.

Papatheodorou, P., Zamboglou, C., Genisyuerek, S., Guttenberg, G., & Aktories, K. (2010). Clostridial glucosylating toxins enter cells via clathrin-mediated endocytosis. *PLoS One, 5*(5), e10673.

Paredes-Sabja, D., & Sarker, M. R. (2012). Adherence of *Clostridium difficile* spores to Caco-2 cells in culture. *Journal of Medical Microbiology, 61*(Pt 9), 1208-1218.

Paredes-Sabja, D., Shen, A., & Sorg, J. A. (2014). *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. *Trends in Microbiology, 22*(7), 406-416.

Pechine, S., Janoir, C., & Collignon, A. (2005). Variability of *Clostridium difficile* surface proteins and specific serum antibody response in patients with *Clostridium difficile*-associated disease. *Journal of Clinical Microbiology, 43*(10), 5018-5025.

Peng, Z., Ling, L., Stratton, C. W., Li, C., Polage, C. R., Wu, B., & Tang, Y. W. (2018). Advances in the diagnosis and treatment of *Clostridium difficile* infections. *Emerging Microbes and Infections, 7*(1), 15.

Pépin, J., Valiquette, L., Alary, M.-E., Villemure, P., Pelletier, A., Forget, K., Pépin, K., & Chouinard, D. (2004). *Clostridium difficile*-associated diarrhea in

a region of Quebec from 1991 to 2003: a changing pattern of disease severity. *Canadian Medical Association Journal, 171*(5), 466-472.

Perelle, S., Gibert, M., Bourlioux, P., Corthier, G., & Popoff, M. R. (1997). Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infection and Immunity, 65*(4), 1402-1407.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods, 8*(10), 785-786.

Pettit, L. J., Browne, H. P., Yu, L., Smits, W. K., Fagan, R. P., Barquist, L., Martin, M. J., Goulding, D., Duncan, S. H., Flint, H. J., Dougan, G., Choudhary, J. S., & Lawley, T. D. (2014). Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. *BMC Genomics, 15*(1), 160.

Pfeifer, G., Schirmer, J., Leemhuis, J., Busch, C., Meyer, D. K., Aktories, K., & Barth, H. (2003). Cellular uptake of *Clostridium difficile* toxin B. Translocation of the N-terminal catalytic domain into the cytosol of eukaryotic cells. *Journal of Biological Chemistry, 278*(45), 44535-44541.

Popoff, M. R. (2018). Toxins of *Clostridium difficile* and *Clostridium sordellii*: The proinflammatory versus cytoprotective response. *Toxicon, 149*, 90.

Popoff, M. R., & Bouvet, P. (2009). Clostridial Toxins. *Future Medicine, 4 (8)*, 1021-1064.

Popoff, M. R., Rubin, E. J., Gill, D. M., & Boquet, P. (1988). Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infection and Immunity, 56*(9), 2299-2306.

Porath, J., Carlsson, J., Olsson, I., & Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature, 258*(5536), 598-599.

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology, 151*(Pt 3), 653-663.

Pruitt, R. N., Chambers, M. G., Ng, K. K., Ohi, M. D., & Lacy, D. B. (2010). Structural organization of the functional domains of *Clostridium difficile* toxins A and B. *Proceedings of the National Academy of Sciences of the United States of America, 107*(30), 13467-13472.

Purdy, D., O'Keeffe, T. A. T., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., & Minton, N. P. (2002). Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Molecular Microbiology, 46*(2), 439-452.

Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., & Chou, C. P. (2016). Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Scientific Reports, 6*, 25666.

Rai, P., He, F., Kwang, J., Engelward, B. P., & Chow, V. T. (2016). Pneumococcal pneumolysin induces DNA damage and cell cycle arrest. *Scientific Reports, 6*, 22972.

Ramarao, N., & Sanchis, V. (2013). The pore-forming haemolysins of *Bacillus cereus*: a review. *Toxins (Basel), 5*(6), 1119-1139.

Rao, S., Prestidge, C. A., Miesel, L., Sweeney, D., Shinabarger, D. L., & Boulos, R. A. (2016). Preclinical development of Ramizol, an antibiotic belonging to a new class, for the treatment of *Clostridium difficile* colitis. *Journal of Antibiotics (Tokyo)*.

Ratledge, C., & Dover, L. G. (2000). Iron metabolism in pathogenic bacteria. *Annual Review of Microbiology, 54*, 881–941.

Raynal, B., Lenormand, P., Baron, B., Hoos, S., & England, P. (2014). Quality assessment and optimization of purified protein samples: why and how? *Microbial Cell Factories, 13*, 180.

Reineke, J., Tenzer, S., Rupnik, M., Koschinski, A., Hasselmayer, O., Schrattenholz, A., Schild, H., & von Eichel-Streiber, C. (2007). Autocatalytic cleavage of *Clostridium difficile* toxin B. *Nature, 446*(7134), 415-419.

Reynolds, C. B., Emerson, J. E., de la Riva, L., Fagan, R. P., & Fairweather, N. F. (2011). The *Clostridium difficile* cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. *PLoS Pathogens, 7*(4), e1002024.

Riedel, T., Bunk, B., Thurmer, A., Sproer, C., Brzuszkiewicz, E., Abt, B., Gronow, S., Liesegang, H., Daniel, R., & Overmann, J. (2015). Genome resequencing of the virulent and multidrug-resistant reference strain *Clostridium difficile* 630. *Genome Announcements, 3*(2).

Roberts, A. P., Hennequin, C., Elmore, M., Collignon, A., Karjalainen, T., Minton, N., & Mullany, P. (2003). Development of an integrative vector for the

expression of antisense RNA in *Clostridium difficile*. *Journal of Microbiological Methods, 55*(3), 617-624.

Roberts, A. P., & Smits, W. K. (2018). The evolving epidemic of *Clostridium difficile* 630. *Anaerobe, 53*, 2-4.

Robertson, D. L., & Leppla, S. H. (1986). Molecular cloning and expression in *Escherichia coli* of the lethal factor gene of *Bacillus anthracis*. *Gene, 44*(1), 71-78.

Robinson, P. K. (2015). Enzymes: principles and biotechnological applications. *Essays in Biochemistry, 59*, 1-41.

Rodriguez-Palacios, A., & Lejeune, J. T. (2011). Moist-heat resistance, spore aging, and superdormancy in *Clostridium difficile*. *Applied and Environmental Microbiology, 77*(9), 3085-3091.

Rodriguez-Palacios, A., Staempfli, H. R., Duffield, T., & Weese, J. S. (2007). *Clostridium difficile* in retail ground meat, Canada. *Emerging Infectious Diseases, 13*(3), 485-487.

Rowe, G. E., & Welch, R. A. (1994). Assays of hemolytic toxins. *Methods in Enzymology, 235*, 657-667.

Rupnik, M. (2007). Is *Clostridium difficile*-associated infection a potentially zoonotic and foodborne disease? *Clinical Microbiology and Infection, 13*(5), 457-459.

Rupnik, M., & Janezic, S. (2016). An update on *Clostridium difficile* toxinotyping. *Journal of Clinical Microbiology, 54*(1), 13-18.

Rupnik, M., Wilcox, M. H., & Gerding, D. N. (2009). *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nature Reviews Microbiology, 7*(7), 526-536.

Salim, K. Y., de Azavedo, J. C., Bast, D. J., & Cvitkovitch, D. G. (2007). Role for sagA and siaA in quorum sensing and iron regulation in *Streptococcus pyogenes*. *Infection and Immunity, 75*(10), 5011-5017.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (2 ed.): Cold Spring Harbor Laboratory Press, New York.

Sansonetti, P. J., Ryter, A., Clerc, P., Maurelli, A. T., & Mounier, J. (1986). Multiplication of *Shigella flexneri* within HeLa cells: lysis of the phagocytic

vacuole and plasmid-mediated contact hemolysis. *Infection and Immunity, 51*(2), 461-469.

Saraswat, M., Musante, L., Ravidá, A., Shortt, B., Byrne, B., & Holthofer, H. (2013). Preparative purification of recombinant proteins: current status and future trends. *BioMed Research International, 2013*, 312709.

Saroj, S. D., Holmer, L., Berengueras, J. M., & Jonsson, A. B. (2017). Inhibitory role of acyl homoserine lactones in hemolytic activity and viability of *Streptococcus pyogenes* M6 S165. *Scientific Repots, 7*, 44902.

Sassone-Corsi, M., & Raffatellu, M. (2015). No Vacancy: How beneficial microbes cooperate with immunity to provide colonization resistance to pathogens. *Journal of immunology (Baltimore, Md. : 1950), 194*(9), 4081-4087.

Schacherl, M., Pichlo, C., Neundorf, I., & Baumann, U. (2015). Structural basis of proline-proline peptide bond specificity of the metalloprotease Zmp1 implicated in motility of *Clostridium difficile*. *Structure, 23*(9), 1632-1642.

Schleberger, C., Hochmann, H., Barth, H., Aktories, K., & Schulz, G. E. (2006). Structure and action of the binary C2 toxin from *Clostridium botulinum*. *Journal of Molecular Biology, 364*(4), 705-715.

Schwan, C., Kruppke, A. S., Nolke, T., Schumacher, L., Koch-Nolte, F., Kudryashev, M., Stahlberg, H., & Aktories, K. (2014). *Clostridium difficile* toxin CDT hijacks microtubule organization and reroutes vesicle traffic to increase pathogen adherence. *Proceedings of the National Academy of Sciences of the United States of America, 111*(6), 2313-2318.

Schwan, C., Stecher, B., Tzivelekidis, T., van Ham, M., Rohde, M., Hardt, W. D., Wehland, J., & Aktories, K. (2009). *Clostridium difficile* toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria. *PLoS Pathogens, 5*(10), e1000626.

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., Thomson, N. R., Roberts, A. P., Cerdeno-Tarraga, A. M., Wang, H., Holden, M. T., Wright, A., Churcher, C., Quail, M. A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B., & Parkhill, J. (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genetics, 38*(7), 779-786.

Seddon, S. V., Hemingway, I., & Borriello, S. P. (1990). Hydrolytic enzyme production by *Clostridium difficile* and its relationship to toxin production and virulence in the hamster model. *Journal of Medical Microbiology, 31*(3), 169-174.

Sehr, P., Joseph, G., Genth, H., Just, I., Pick, E., & Aktories, K. (1998). Glucosylation and ADP ribosylation of rho proteins: effects on nucleotide binding, GTPase activity, and effector coupling. *Biochemistry, 37*(15), 5296-5304.

Seidler, A. (1994). Introduction of a histidine tail at the N-terminus of a secretory protein expressed in *Escherichia coli*. *Protein Engineering, 7*(10), 1277-1280.

Sekulovic, O., Mathias Garrett, E., Bourgeois, J., Tamayo, R., Shen, A., & Camilli, A. (2018). Genome-wide detection of conservative site-specific recombination in bacteria. *PLoS genetics, 14*(4), e1007332-e1007332.

Sekulovic, O., Ospina Bedoya, M., Fivian-Hughes, A. S., Fairweather, N. F., & Fortier, L. C. (2015). The *Clostridium difficile* cell wall protein CwpV confers phase-variable phage resistance. *Molecular Microbiology, 98*(2), 329-342.

Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., van der Oost, J., Brouns, S. J., & Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America, 108*(25), 10098-10103.

Shannon, J. G., Ross, C. L., Koehler, T. M., & Rest, R. F. (2003). Characterization of anthrolysin O, the *Bacillus anthracis* cholesterol-dependent cytolysin. *Infection and Immunity, 71*(6), 3183-3189.

Shepherd, M., Heath, M. D., & Poole, R. K. (2007). NikA binds heme: a new role for an *Escherichia coli* periplasmic nickel-binding protein. *Biochemistry, 46*(17), 5030-5037.

Shimizu, T., Ba-Thein, W., Tamaki, M., & Hayashi, H. (1994). The virR gene, a member of a class of two-component response regulators, regulates the production of perfringolysin O, collagenase, and hemagglutinin in *Clostridium perfringens*. *Journal of Bacteriology, 176*(6), 1616-1623.

Short, E. C., & Kurtz, H. J. (1971). Properties of the hemolytic activities of *Escherichia coli*. *Infection and Immunity, 3*(5), 678-687.

Shuman, S., & Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. *Nature Reviews Microbiology, 5*(11), 852-861.

Sleytr, U. B., Schuster, B., Egelseer, E. M., & Pum, D. (2014). S-layers: principles and applications. *FEMS Microbiology Reviews, 38*(5), 823-864.

Slimings, C., & Riley, T. V. (2014). Antibiotics and hospital-acquired *Clostridium difficile* infection: update of systematic review and meta-analysis. *Journal of Antimicrobial Chemotherapy, 69*(4), 881-891.

Smith, A. (2005). Outbreak of *Clostridium difficile* infection in an English hospital linked to hypertoxin-producing strains in Canada and the US. *Euro Surveillance: Europe's journal on infectious disease surveillance, epidemiology, prevention and control, 10*(26), pii=2735.

Smith, M. C., Furman, T. C., Ingolia, T. D., & Pidgeon, C. (1988). Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. *Journal of Antimicrobial Chemotherapy, 263*(15), 7211-7215.

Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H., & Kuijper, E. J. (2016). *Clostridium difficile* infection. *Nature Reviews Disease Primers, 2*, 16020.

Sohanpal, B. K., Friar, S., Roobol, J., Plumbridge, J. A., & Blomfield, I. C. (2007). Multiple co-regulatory elements and IHF are necessary for the control of fimB expression in response to sialic acid and N-acetylglucosamine in *Escherichia coli* K-12. *Molecular Microbiology, 63*(4), 1223-1236.

Song, Y., Garg, S., Girotra, M., Maddox, C., von Rosenvinge, E. C., Dutta, A., Dutta, S., & Fricke, W. F. (2013). Microbiota dynamics in patients treated with fecal microbiota transplantation for recurrent *Clostridium difficile* infection. *PLoS One, 8*(11), e81330.

Songer, J. G., Trinh, H. T., Killgore, G. E., Thompson, A. D., McDonald, L. C., & Limbago, B. M. (2009). *Clostridium difficile* in retail meat products, USA, 2007. *Emerging Infectious Diseases, 15*(5), 819-821.

Sorg, J. A., & Sonenshein, A. L. (2008). Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *Journal of Bacteriology, 190*(7), 2505-2512.

Sponseller, J. K., Steele, J. A., Schmidt, D. J., Kim, H. B., Beamer, G., Sun, X., & Tzipori, S. (2015). Hyperimmune bovine colostrum as a novel therapy to combat *Clostridium difficile* infection. *The Journal of Infectios Diseases, 211*(8), 1334-1341.

Spriestersbach, A., Kubicek, J., Schafer, F., Block, H., & Maertens, B. (2015). Purification of His-Tagged Proteins. *Methods in Enzymology, 559*, 1-15.

Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T. D., Sebaihia, M., Quail, M. A., Rose, G., Gerding, D. N., Gibert, M., Popoff, M. R., Parkhill, J., Dougan, G., & Wren, B. W. (2009). Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biology, 10*(9), R102-R102.

Stachowiak, R., Lyzniak, M., Budziszewska, B. K., Roeske, K., Bielecki, J., Hoser, G., & Kawiak, J. (2012). Cytotoxicity of bacterial metabolic products, including listeriolysin O, on leukocyte targets. *Journal of Biomedicine and Biotechnology, 2012*, 954375.

Stasche, R., Hinderlich, S., Weise, C., Effertz, K., Lucka, L., Moormann, P., & Reutter, W. (1997). A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Molecular cloning and functional expression of UDP-N-acetyl-glucosamine 2-epimerase/N-acetylmannosamine kinase. *Journal of Biological Chemistry, 272*(39), 24319-24324.

Steiner, E., Dago, A. E., Young, D. I., Heap, J. T., Minton, N. P., Hoch, J. A., & Young, M. (2011). Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*. *Molecular Microbiology, 80*(3), 641-654.

Sulakvelidze, A., Alavidze, Z., & Morris, J. G., Jr. (2001). Bacteriophage therapy. *Antimicrobial Agents and Chemotherapy, 45*(3), 649-659.

Summers, W. C. (1999). *Felix d'Herelle and the origins of molecular biology*: Yale University Press, New Haven, Connecticut.

Sundriyal, A., Roberts, A. K., Ling, R., McGlashan, J., Shone, C. C., & Acharya, K. R. (2010). Expression, purification and cell cytotoxicity of actin-modifying binary toxin from *Clostridium difficile*. *Protein Expression and Purification, 74*(1), 42-48.

Sundriyal, A., Roberts, A. K., Shone, C. C., & Acharya, K. R. (2009). Structural basis for substrate recognition in the enzymatic component of ADP-ribosyltransferase toxin CDTa from *Clostridium difficile*. *Journal of Biological Chemistry, 284*(42), 28713-28719.

Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J., & Huttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon

Archaeoglobus fulgidus. *Proceedings of the National Academy of Sciences of the United States of America, 99*(11), 7536-7541.

Tao, L., Zhang, J., Meraner, P., Tovaglieri, A., Wu, X., Gerhard, R., Zhang, X., Stallcup, W. B., Miao, J., He, X., Hurdle, J. G., Breault, D. T., Brass, A. L., & Dong, M. (2016). Frizzled proteins are colonic epithelial receptors for *C. difficile* toxin B. *Nature, 538*(7625), 350-355.

Tasteyre, A., Barc, M.-C., Karjalainen, T., Dodson, P., Hyde, S., Bourlioux, P., & Borriello, P. (2000). A *Clostridium difficile* gene encoding flagellin. *Microbiology, 146*(4), 957-966.

Tasteyre, A., Barc, M. C., Collignon, A., Boureau, H., & Karjalainen, T. (2001). Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infection and Immunity, 69*(12), 7937-7940.

Tay, S. T., Devi, S., Puthucheary, S. D., & Kautner, I. M. (1995). Detection of haemolytic activity of campylobacters by agarose haemolysis and microplate assay. *Journal of Medical Microbiology, 42*(3), 175-180.

Taylor, N. S., Thorne, G. M., & Bartlett, J. G. (1981). Comparison of two toxins produced by *Clostridium difficile*. *Infection and Immunity, 34*(3), 1036-1043.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology, 60*(5), 523-533.

Theriot, C. M., Bowman, A. A., & Young, V. B. (2016). Antibiotic-induced alterations of the gut microbiota alter secondary bile acid production and allow for *Clostridium difficile* spore germination and outgrowth in the large intestine. *mSphere, 1*(1).

Todar, K. (2012). Mechanisms of Bacterial Pathogenicity.Todar's Online Textbook of Bacteriology. Retrieved from http://textbookofbacteriology.net/pathogenesis_4.html.

Todd, E. W. (1932). Antigenic Streptococcal hemolysin. *The Journal of Experimental Medicine, 55*(2), 267.

Tulli, L., Marchi, S., Petracca, R., Shaw, H. A., Fairweather, N. F., Scarselli, M., Soriani, M., & Leuzzi, R. (2013). CbpA: a novel surface exposed adhesin of *Clostridium difficile* targeting human collagen. *Cell Microbiology, 15*(10), 1674-1687.

Tweten, R. K. (2005). Cholesterol-dependent cytolysins, a family of versatile pore-forming toxins. *Infection and Immunity, 73*(10), 6199-6209.

Twine, S. M., Reid, C. W., Aubry, A., McMullin, D. R., Fulton, K. M., Austin, J., & Logan, S. M. (2009). Motility and flagellar glycosylation in *Clostridium difficile*. *Journal of Bacteriology, 191*(22), 7050-7062.

Urbina, P., Flores-Diaz, M., Alape-Giron, A., Alonso, A., & Goni, F. M. (2009). Phospholipase C and sphingomyelinase activities of the *Clostridium perfringens* alpha-toxin. *Chemistry and Physics of Lipids, 159*(1), 51-57.

van Eijk, E., Anvar, S. Y., Browne, H. P., Leung, W. Y., Frank, J., Schmitz, A. M., Roberts, A. P., & Smits, W. K. (2015). Complete genome sequence of the *Clostridium difficile* laboratory strain 630Deltaerm reveals differences from strain 630, including translocation of the mobile element CTn5. *BMC Genomics, 16*, 31.

Varga, J. J., Nguyen, V., O'Brien, D. K., Rodgers, K., Walker, R. A., & Melville, S. B. (2006). Type IV pili-dependent gliding motility in the Gram-positive pathogen *Clostridium perfringens* and other Clostridia. *Molecular Microbiology, 62*(3), 680-694.

Vassylyeva, M. N., Klyuyev, S., Vassylyev, A. D., Wesson, H., Zhang, Z., Renfrow, M. B., Wang, H., Higgins, N. P., Chow, L. T., & Vassylyev, D. G. (2017). Efficient, ultra-high-affinity chromatography in a one-step purification of complex proteins. *Proceedings of the National Academy of Sciences of the United States of America, 114*(26), E5138-E5147.

von Eichel-Streiber, C., Boquet, P., Sauerborn, M., & Thelestam, M. (1996). Large clostridial cytotoxins-a family of glycosyltransferases modifying small GTP-binding proteins. *Trends in Microbiology, 4*(10), 375-382.

von Eichel-Streiber, C., Sauerborn, M., & Kuramitsu, H. K. (1992). Evidence for a modular structure of the homologous repetitive C-terminal carbohydrate-binding sites of *Clostridium difficile* toxins and *Streptococcus mutans* glucosyltransferases. *Journal of Bacteriology, 174*(20), 6707-6710.

Wade, K. R., Hotze, E. M., & Tweten, R. K. (2015). 24- Perfringolysin O and related cholesterol-dependent cytolysins: mechanism of pore formation. In J. Alouf, D. Ladant, & M. R. Popoff (Eds.), *The Comprehensive Sourcebook of Bacterial Protein Toxins (Fourth Edition)* (pp. 719-738). Boston: Academic Press.

Walev, I., Palmer, M., Valeva, A., Weller, U., & Bhakdi, S. (1995). Binding, oligomerization, and pore formation by streptolysin O in erythrocytes and

fibroblast membranes: detection of nonlytic polymers. *Infection and Immunity, 63*(4), 1188-1194.

Waligora, A. J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A., & Karjalainen, T. (2001). Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. *Infection and Immunity, 69*(4), 2144-2153.

Wang, H., Smith, M. C., & Mullany, P. (2006). The conjugative transposon Tn*5397* has a strong preference for integration into its *Clostridium difficile* target site. *Journal of Bacteriology, 188*(13), 4871-4878.

Wang, S., Hong, W., Dong, S., Zhang, Z. T., Zhang, J., Wang, L., & Wang, Y. (2018). Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. *Clinical Microbiology and Infection, 24*(10), 1095-1099.

Wang, S., Xu, M., Wang, W., Cao, X., Piao, M., Khan, S., Yan, F., Cao, H., & Wang, B. (2016a). Systematic review: Adverse events of fecal microbiota transplantation. *PLoS One, 11*(8), e0161174.

Wang, Y., Li, X., Milne, C. B., Janssen, H., Lin, W., Phan, G., Hu, H., Jin, Y.-S., Price, N. D., & Blaschek, H. P. (2013). Development of a gene knockout system using mobile group II introns (targetron) and genetic disruption of acid production pathways in *Clostridium beijerinckii*. *Applied and Environmental Microbiology, 79*(19), 5853-5863.

Wang, Y., Zhang, Z. T., Seo, S. O., Choi, K., Lu, T., Jin, Y. S., & Blaschek, H. P. (2015). Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *Journal of Biotechnology, 200*, 1-5.

Wang, Y., Zhang, Z. T., Seo, S. O., Lynn, P., Lu, T., Jin, Y. S., & Blaschek, H. P. (2016b). Bacterial genome editing with CRISPR-Cas9: Deletion, integration, single nucleotide modification, and desirable "clean" mutant selection in *Clostridium beijerinckii* as an example. *ACS Synthetic Biology, 5*(7), 721-732.

Warny, M., Pepin, J., Fang, A., Killgore, G., Thompson, A., Brazier, J., Frost, E., & McDonald, L. C. (2005). Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *The Lancet, 366*(9491), 1079-1084.

Washington, J. A. (2012). *Laboratory Procedures in Clinical Microbiology*: Springer Science & Business Media.

Wattam, A., Davis, J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E., Disz, T., Gabbard, J., Gerdes, S., Henry, C., Kenyon, R., Machi,

D., Mao, C., Nordberg, E., Olsen, G., Murphy-Olson, D., Olson, R., Overbeek, R., Parrello, B., Pusch, G., Shukla, M., Vonstein, V., Warren, A., Xia, F., Yoo, H., & Stevens, R. (2017). *Nucleic Acids Res*. 45(D1):D535-D542. https://www.patricbrc.org/webpage/website/cite_patric.html (Publication no. 10.1093/nar/gkw1017).

Waugh, D. S. (2011). An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification, 80*(2), 283-293.

Weld, J. T. (1935). Further studies with toxic serum extracts of hemolytic Streptococci. *The Journal of Experimental Medicine, 61*(4), 473.

Westra, D. F., Welling, G. W., Koedijk, D. G. A. M., Scheffer, A. J., The, T. H., & Welling-Wester, S. (2001). Immobilised metal-ion affinity chromatography purification of histidine-tagged recombinant proteins: a wash step with a low concentration of EDTA. *Journal of Chromatography B: Biomedical Sciences and Applications, 760*(1), 129-136.

Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., Westphal, W., Heck, A. J., Boekema, E. J., Dickman, M. J., & Doudna, J. A. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the National Academy of Sciences of the United States of America, 108*(25), 10092-10097.

Wilcox, M. H., Gerding, D. N., Poxton, I. R., Kelly, C., Nathan, R., Birch, T., Cornely, O. A., Rahav, G., Bouza, E., Lee, C., Jenkin, G., Jensen, W., Kim, Y. S., Yoshida, J., Gabryelski, L., Pedley, A., Eves, K., Tipping, R., Guris, D., Kartsonis, N., Dorr, M. B., Modify, I., & Investigators, M. I. (2017). Bezlotoxumab for prevention of recurrent *Clostridium difficile* infection. *New England Journal of Medicine, 376*(4), 305-317.

Wilcox, M. H., Shetty, N., Fawley, W. N., Shemko, M., Coen, P., Birtles, A., Cairns, M., Curran, M. D., Dodgson, K. J., Green, S. M., Hardy, K. J., Hawkey, P. M., Magee, J. G., Sails, A. D., & Wren, M. W. (2012). Changing epidemiology of *Clostridium difficile* infection following the introduction of a national ribotyping-based surveillance scheme in England. *Clinical Infections Diseases, 55*(8), 1056-1063.

Wiseman, G. M. (1975). The hemolysins of *Staphylococcus aureus*. *Bacteriological Reviews, 39*(4), 317.

Wolfe, C., Pagano, P., Pillar, C. M., Shinabarger, D. L., & Boulos, R. A. (2018). Comparison of the *in vitro* antibacterial activity of Ramizol, fidaxomicin, vancomycin, and metronidazole against 100 clinical isolates of

Clostridium difficile by broth microdilution. *Diagnostic Microbiology and Infectious Disease, 92*(3), 250-252.

Wootton, J. C. (1994). Non-globular domains in protein sequences: automated segmentation using complexity measures. *Computers and Chemistry, 18*(3), 269-285.

Wu, A. G., Alibek, D., Li, Y. L., Bradburne, C., Bailey, C. L., & Alibek, K. (2003). Anthrax toxin induces hemolysis: An indirect effect through polymorphonuclear cells. *Journal of Infectious Diseases, 188*(8), 1138-1141.

Wu, J., & Filutowicz, M. (1999). Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. *Acta Biochimica Polonica, 46*(3), 591-599.

Wust, J., & Hardegger, U. (1983). Transferable resistance to clindamycin, erythromycin, and tetracycline in *Clostridium difficile*. *Antimicrobial Agents and Chemotherapy, 23*(5), 784-786.

Yosef, I., Goren, M. G., & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Research, 40*(12), 5569-5576.

Young, J. A., & Collier, R. J. (2007). Anthrax toxin: receptor binding, internalization, pore formation, and translocation. *Annual Review of Biochemistry, 76*, 243-265.

Yutin, N., & Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environmental Microbiology, 15*(10), 2631-2641.

Zeiser, J., Gerhard, R., Just, I., & Pich, A. (2013). Substrate specificity of clostridial glucosylating toxins and their function on colonocytes analyzed by proteomics techniques. *Journal of Proteome Research, 12*(4), 1604-1618.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., & Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell, 163*(3), 759-771.

Zhong, J., Karberg, M., & Lambowitz, A. M. (2003). Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Research, 31*(6), 1656-1664.