The Function and Biogenesis of the *Clostridium difficile* S-layer

By:

Joseph Adam Kirk

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Science
Department of Molecular Biology and Biotechnology

August 2017

Abstract

Clostridium difficile is the most common cause of antibiotic-associated diarrhoea and the severity and incidence of *C. difficile* infections (CDIs) have risen dramatically in the last decade. *C. difficile* is only able to cause disease once the microbiota of the gut has been damaged, usually through the use of broad spectrum antibiotics. CDIs are treated using antibiotics which further damage the microbiota, resulting in high rates of recurrence. This highlights the need for more specific CDI therapies that do not damage the natural flora.

Avidocin-CDs are a potential alternative therapy for CDI that resemble the R-type Pyocins of *Pseudomonas aeruginosa,* and kill *C. difficile* by inserting a needle like core through the bacterial cell envelope. We have demonstrated that these complexes bind SlpA, the major component of the S-layer. Spontaneous mutants resistant to Avidocin-CDs, that lacked a functional S-layer, were isolated. These S-layer mutants, and a strain in which functional *slpA* was returned to the chromosome of these mutants, provided a unique opportunity to study the function of the S-layer. We have now demonstrated that the S-layer affects the growth and morphology of cells, sporulation efficiency, resistance to innate immune effectors and CwpV expression. Interestingly, we also demonstrated that the S-layer mutants are avirulent in the hamster model of infection, although maintain the ability to persist in the gut. Sequencing of the S-layer mutant genome also allowed the identification of a new phase variable promoter.

In addition to studying S-layer function we also investigated S-layer biogenesis and the role of SecA2. Using pull-down experiments, we have identified potential SecA2 binding partners including members of the Cell Wall Protein family and the cell wall synthesis machinery components MreB and FtsZ. We provide evidence that S-layer biogenesis is localised to areas of cell wall growth and suggest a model in which SecA2 interacts with MreB or FtsZ to localise the secretion of SlpA to these areas.

Publications

Kirk, J.A., Banerji, O. & Fagan, R.P., 2017. Characteristics of the *Clostridium difficile* cell envelope and its importance in therapeutics. Microbial biotechnology, 10(1), pp.76–90.

Kirk, J.A. & Fagan, R.P., 2016. Heat shock increases conjugation efficiency in *Clostridium difficile*. Anaerobe, 42, pp. 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, doi:10.1126/scitranslmed.aah6813.

Acknowledgements

Firstly, I must thank my supervisor Dr Robert Fagan for his constant support through my PhD. Anyone who knows Rob will agree that there are few people who share his enthusiasm for research and debate, especially after a pint or two. Without Rob and his seemingly bottomless font of knowledge, my PhD would have been less enjoyable and I wouldn't have accomplished nearly as much as I have. I would also like to thank Rob for giving me the freedom to follow up on my own ideas and projects. In addition, I would like to thank my collaborators Dr Paula Salgardo and Dr Gillian Douce who have both enabled me to take my research further. I must also thank the University of Sheffield, which I have been visiting since I was 13 years old, and funded my PhD.

Without my friends, inside and outside of the lab, I would have gone insane years ago. I started my PhD with Bartek, and I'm glad we get to finish together. Our trips to the whiskey bar, fishing excursions, and his fantastic cooking really kept my spirits high. Bartek and Kasia are two of the friendliest people I have been fortunate enough to meet and I'm glad we will be working together for at least another year. Sophie joined our lab during the second year of my PhD and I have thoroughly enjoyed our coffee mornings and meme wars. Although I don't have enough room to express my gratitude to every member of the lab, Stéphane, Luz, Smitty, Nicola, Peter, Oishik, Nadia, Rebecca, and Alison have all provided me with wonderful experiences. I don't know how I got any work done surrounded by such wonderful people. I'd also like to thank Kyle. I've known Kyle for over a decade and he always, for better or worse, inflates my ego and makes me feel I can do anything. I would like to thank my parents and grandparents for getting me this far and my brother for being the one person in the family that gets it.

Lastly, and most importantly, I would like to thank my wife, Cindy. Cindy and I started to rent our first apartment together at the beginning of my PhD and got married near the end. Cindy reminds me that work is not everything and without her I would have burnt out long ago. Cindy is the one person who has shared both my good days and bad, celebrating the wins and cheering me up after the losses and I couldn't have done this without her. Cindy has shown unwavering patience with me and my research, with many weekends and evenings sacrificed, and everyday I try to make her proud. I dedicate this work to you.

'But then science is nothing but a series of questions that lead to more questions,

which is just as well, or it wouldn't be much of a career path, would it?'

-Terry Pratchet

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Atc	Anhydrotetracycline
BHI	Brain heart infusion
Ca-DPA	Calcium Dipicolinic acid
cAMP	Cyclic AMP
CDI	*C. difficile*
c-di-GMP	Cyclic-di-GMP
CDMM	*C. difficile* minimal medium
CDT	*C. difficile* binary toxin
CFU	Colony forming units
CROP	C-terminal combined repetitive oligopeptides
CTD	C-terminal Domain
CWB2	Cell wall binding domain 2
CWP	Cell wall protein
DMSO	Dimethyl sulfoxide
dNTP	Deoxynucleotide
DTT	Dithiothreitol
FMT	Faecal microbiota transplantation
gDNA	Genomic DNA
GI	Gastro intestinal
HMW	High molecular weight
HSD	Helical scaffold domain
HWD	Helical wing domain
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRA1	intramolecular regulator of ATP hydrolysis 1
LC-MS	Liquid chromatography–mass spectrometry
LMW	Low molecular weight
LPS	Lipopolysaccharide
*Mtb*SecA1/2	*M. tuberculosis* SecA1/2

MWCO	Molecular weight cut off
NBD1/2	Nucleotide binding domain 1/2
O/N	Over night
OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PPXD	Substrate crosslinking domain
RAM	Retrotransposition-acivated marker
RBP	Receptor binding protein
RT	Room temperature
SCWP	Secondary cell wall polysaccharide
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
S-layer	Surface layer
SLCT	S-layer cassette type
SLH	S-layer homology domain
SLP	S-layer protein
SNP	Single nucleotide polymorphism
SOEing	Splicing by overlap extension
SRP	Signal recognition particle
TEM	Transmission electron microscopy
TLR4	Toll like receptor 4
TY	Tryptose, yeast extract
UPE	Upstream promoter element
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter I

Introduction

1.1 *Clostridium difficile*

Clostridium difficile is a Gram-positive, spore forming, obligate anaerobe and is the major cause of antibiotic-associated diarrhoea. Originally named *Bacillus difficilis* due to the difficulty in culturing this organism, *C. difficile* was first isolated in 1935 from the stool of a healthy infant (Hall and O'Toole, 1935). Although pseudomembranous colitis, one of the most severe complications of *C. difficile* infection (CDI), was first described in 1893 (Bartlett, 1994), it wasn't until 1978 that *C. difficile* was associated with human disease, being identified as the cause of the majority of antibiotic-associated diarrhoea cases (Bartlett et al., 1978; George et al., 1978). *C. difficile* has now been recognised as the leading cause of nosocomial diarrhoea and is being increasingly associated with community acquired disease (Freeman et al., 2010). Asymptomatic carriage of *C. difficile* occurs in 2-18% of healthy adults, with rates higher in hospital settings compared to the community, and 25-35% in children under 1 year old (Smits et al., 2016). Despite this frequent asymptomatic carriage, *C. difficile* can cause disease ranging in severity from mild, self-limiting diarrhoea to the more severe cases characterised by severe diarrhoea, abdominal pain, and leucocytosis (Burke and Lamont, 2014). These diseases can progress into the more severe inflammatory complication, pseudomembranous colitis, which can, in turn, progress to toxic megacolon (Fig.1.1). Such severe CDIs are associated with high rates of mortality (Martin et al., 2016).
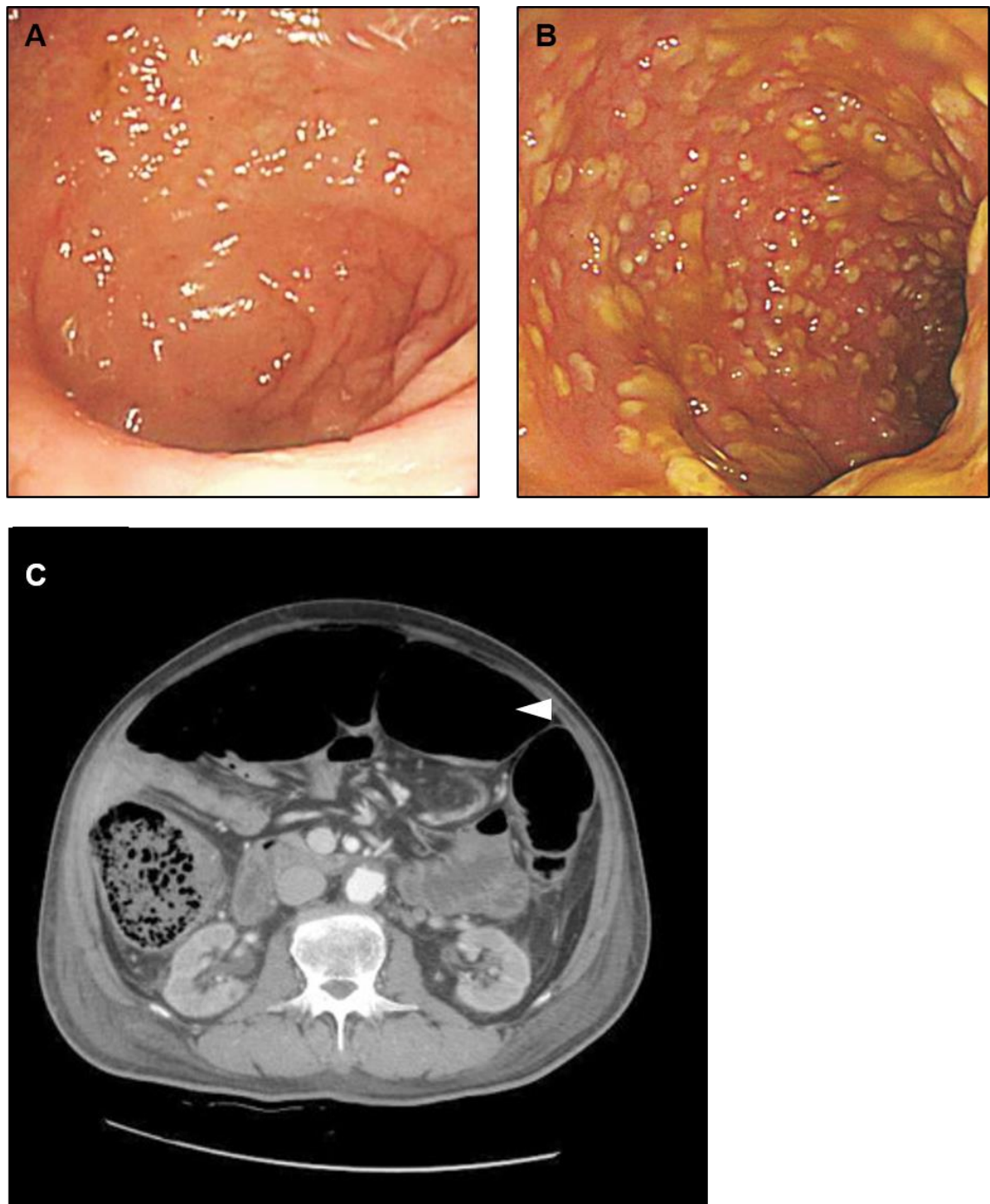
Fig.1.1: Manifestations of severe CDI. (A) Colonoscopic examination of a healthy colon, showing normal mucosa. **(B)** Sigmoidoscopy of a 74-year-old man showing symptoms of severe CDI. Notable are the 'volcanic' yellow lesions present on the mucosa, characteristic of pseudomembranous colitis. **(C)** Computed tomography (transverse plane) of the patient's abdomen showing clear dilation of the colon (white arrow), characteristic of toxic megacolon. Modified from (Gweon et al., 2015).

1.1.1 Epidemiology of *C. difficile* Infection

From its identification as a pathogen in the 1970s, CDI was considered a common, easily managed complication of antibiotic use. Before 2003, no systematic surveillance on circulating *C. difficile* types was performed. However, since 2003 a large increase in the severity and incidence of CDI was observed in North America, several European countries including the UK, and, after 2008, Australia, Asia, and Central America. This was associated with emergence of the epidemic PCR ribotype 027 strains, and to a lesser extent ribotype 078 (Goorhuis et al., 2008; He et al., 2013).

This increase in incidence and mortality due to CDI continued in the UK until 2007, when more than 55,000 cases of CDI resulting in 8,324 deaths were reported, more than five times the number of deaths caused by MRSA infection (https://www.ons.gov.uk/). Due to the huge threat posed by CDI, increased surveillance and mandatory reduction targets for CDI were implemented and the incidence has fallen steadily since 2007 in England and Wales, with only 1,646 deaths involving CDI reported in 2012 (Wilcox et al., 2012) (https://www.ons.gov.uk/).

Despite these infection control successes in the UK, the European Centre for Disease Prevention and Control (ECDC) 2011 and 2012 point-prevalence survey estimated that within the European Union approximately 124,000 individuals develop nosocomial CDI each year and, as many cases are overlooked, this figure is probably an underestimate (ECDC, 2013). Although the accuracy of the approximation is uncertain, the US Centres for Disease Control and Prevention (CDC) estimated that there were approximately 500,000 cases of CDI, resulting in 29,000 deaths, in the US in 2011 (Lessa et al., 2015). *C. difficile* poses such as threat to human health that in 2013 the CDC labelled the risk posed as 'urgent' (CDC, 2013).

CDI typically occurs in health-care settings due to the abundance of antibiotic use, high concentrations of elderly patients with compromised immune systems, abundance of spores in the environment, and poor health care provider hygiene. For these reasons, elderly, hospitalised

patients undergoing prolonged treatment with antibiotics are the most common sufferers of CDI. Broad spectrum antibiotic treatment increases the risk of CDI as it disrupts the intestinal microbiota allowing *C. difficile* to establish and cause disease. *C. difficile* is the most common cause of antibiotic associated diarrhoea, estimated to cause 15-39% of cases (Viswanathan et al., 2010). Although many antibiotics have been associated with CDI, cephalosporins, clindamycin and fluoroquinolones carry a higher risk compared to others (Johnson et al., 1999; Pepin et al., 2005). Unlike cephalosporins and clindamycin, to which almost all *C. difficile* strains are resistant, the risk of fluoroquinolone use is more recent due to the increased incidence of fluoroquinolone resistant ribotype 027 strains (Pepin et al., 2005).

Although CDI is typically a nosocomial infection, community-acquired disease is increasingly observed and, atypically, these infections occur in individuals lacking the classic risk factors, including prior antibiotic exposure (Gupta and Khanna, 2014). In North America and Europe, it is estimated that 20-27% of CDI cases are community-associated (Gupta and Khanna, 2014). A recent study of over 15,000 CDI cases in the USA concluded that community-acquired infection occurred in 30-120 individuals per 100,000 (Lessa et al., 2015). In the Netherlands, community-acquired CDI alone is estimated to occur in 390–730 individuals per 100,000, making it more common than *Salmonella* infection (Bouwknegt et al., 2015). Individuals with community acquired CDI are less likely to have been exposed to antibiotics prior to infection compared to nosocomial infections (78% and 98% respectively) (Khanna et al., 2012). In addition, it seems that community acquired CDI occurs more often in younger patients, with the median age of community-acquired CDI patients being 50, compared with 72 for patients with nosocomial infections (Khanna et al., 2012).

1.1.2 Tools Available for *C. difficile* Genetic Manipulation

C. difficile research has been hampered due to the lack of tools available for the genetic modification of this organism. Recently however, several methods for manipulation of the *C.*

difficile genome have been described. The first was ClosTron, which allowed the inactivation of genes through targeted insertion of a group II intron (Heap et al., 2007). However, the use of ClosTron carries several disadvantages, including an inability to precisely manipulate the genome, make clean knock-outs, and polar effects due to the insertion of the group II intron.

These problems were overcome with the development of pseudosuicide homologous recombination vectors (Cartman et al., 2012). Areas of homology flanking the genomic site to be modified are introduced into these vectors and the genome is then modified via two homologous recombination events with the vector, a process known as allele exchange. These vectors contain sub-optimal origins of replication that limit the plasmid replication rate. Under antibiotic selection, the growth rate of *C. difficile* bearing these plasmids is limited by the rate of plasmid replication. Upon the first recombination event, the plasmid is incorporated into the genome; the replication of the plasmid DNA is therefore no longer dependent on the origin of replication, and the growth rate of the bacterium becomes normal. Cells in which the first recombination event has occurred can therefore be selected by their increased growth rate relative to those in which the plasmid has not yet recombined with the genome. The second recombination event, which results in the excision of the plasmid from the genome and subsequent plasmid loss, can then be selected by growth of the bacteria on media containing 5-fluorocytosine. The recombination vectors contain a gene encoding the cytosine deaminase CodA, an enzyme that converts cytosine to uracil. Plasmid-bearing bacteria cannot grow on plates containing 5-fluorocytosine as CodA converts it into the toxic 5-fluorouracil (Cartman et al., 2012). Therefore, bacteria that grow on this media have lost the plasmid and either reverted to WT or contain the desired mutation, theoretically a 50% chance either way. The use of allele exchange allows for precise manipulation of the genome, including insertion, deletion, and substitution of sequences.

1.1.3 *C. difficile* Sporulation and Germination

C. difficile is transmitted by the faecal-oral route with oxygen-tolerant endospores acting as the transmissive agent (Deakin et al., 2012). These spores are highly resistant to environmental stresses including high temperatures and many disinfectants. As spores are metabolically inactive, they are also unaffected by antibiotic treatment. Despite the clinical importance of *C. difficile,* our understanding of sporulation by this organism is relatively poor (as reviewed in (Paredes-Sabja et al., 2014)). Although the sporulation pathways of *C. difficile* and *B. subtilis* are mechanistically similar, little homology is shared by the spore-coat proteins of these organisms. Components in the sporulation pathway shared between these two species also appear to be regulated differently. The stimuli leading to sporulation initiation in *C. difficile* are yet to be identified, although the cascade is likely to respond to similar stimuli as in other organisms, such as nutrient starvation and quorum sensing (Higgins and Dworkin, 2012). As in *B. subtilis,* initiation of sporulation in *C. difficile* is caused by the activation of the master regulator Spo0A. This leads to the activation of RNA polymerase sigma factors σ^F, σ^E, σ^G, and σ^K (Fimlaid et al., 2013). Although the activation of these sigma factors is regulated differently in *C. difficile* relative to *B. subtilis*, they do display compartment-specific activation with σ^F and σ^G being localised to the forespore, and σ^E and σ^K to the mother cell (Pereira et al., 2013). These sigma factors control the expression of proteins involved in spore formation and mother cell lysis. The overall structure of the *C. difficile* spores shows great similarity to the spores of other organisms (Fig.1.2).

Once ingested, the spore interacts with bile salts initiating germination (Sorg and Sonenshein, 2008). Primary bile salts, notably cholate conjugated with either taurine or glycine, are produced by the liver and are secreted into the duodenum. They then pass into the jejunum and ileum, where they are largely absorbed and recycled (Ridlon et al., 2006). However, some passes into the anaerobic environment of the caecum, where members of the microbiota expressing bile salt hydrolases remove the bile salt from its conjugated amino acid (Ridlon et al.,

2006). The unconjugated bile salts are then converted to secondary bile salts such as deoxycholate by members of the microbiota (Ridlon et al., 2006). Some secondary bile salts can initiate germination of spores, but prevent the outgrowth of vegetative cells (Sorg and Sonenshein, 2008). In healthy individuals, *C. difficile* spores pass through the stomach to the jejunum, rich in nutrients and primary bile acids, begin to germinate, and pass into the caecum where secondary bile salts inhibit the outgrowth of vegetative cells. Growth may also be limited by competition for nutrients and immune-modulation by the intact microbiota. Antibiotic treatment disrupts the microbiota reducing the capability of the flora to convert primary bile salts into secondary bile salts. This would lead to an increase in the concentration of primary bile salts, and decrease in secondary bile salts, in the caecum. This change in bile salt composition may facilitate the germination and outgrowth of *C. difficile* spores (Sorg and Sonenshein, 2008). The bile salt receptor has recently been identified as the protease CspC, although it is unclear how it activates downstream components of the germination pathway (Francis et al., 2013). Sensing of primary bile salts by CspC causes a proteolytic cascade which results in degradation of the spore's peptidoglycan, rehydration of the spore, calcium dipicolinic acid (Ca-DPA) loss, and subsequent outgrowth of the bacterium. Glycine, through an unknown mechanism, is also able to act as a *C. difficile* co-germinant (Sorg and Sonenshein, 2008).
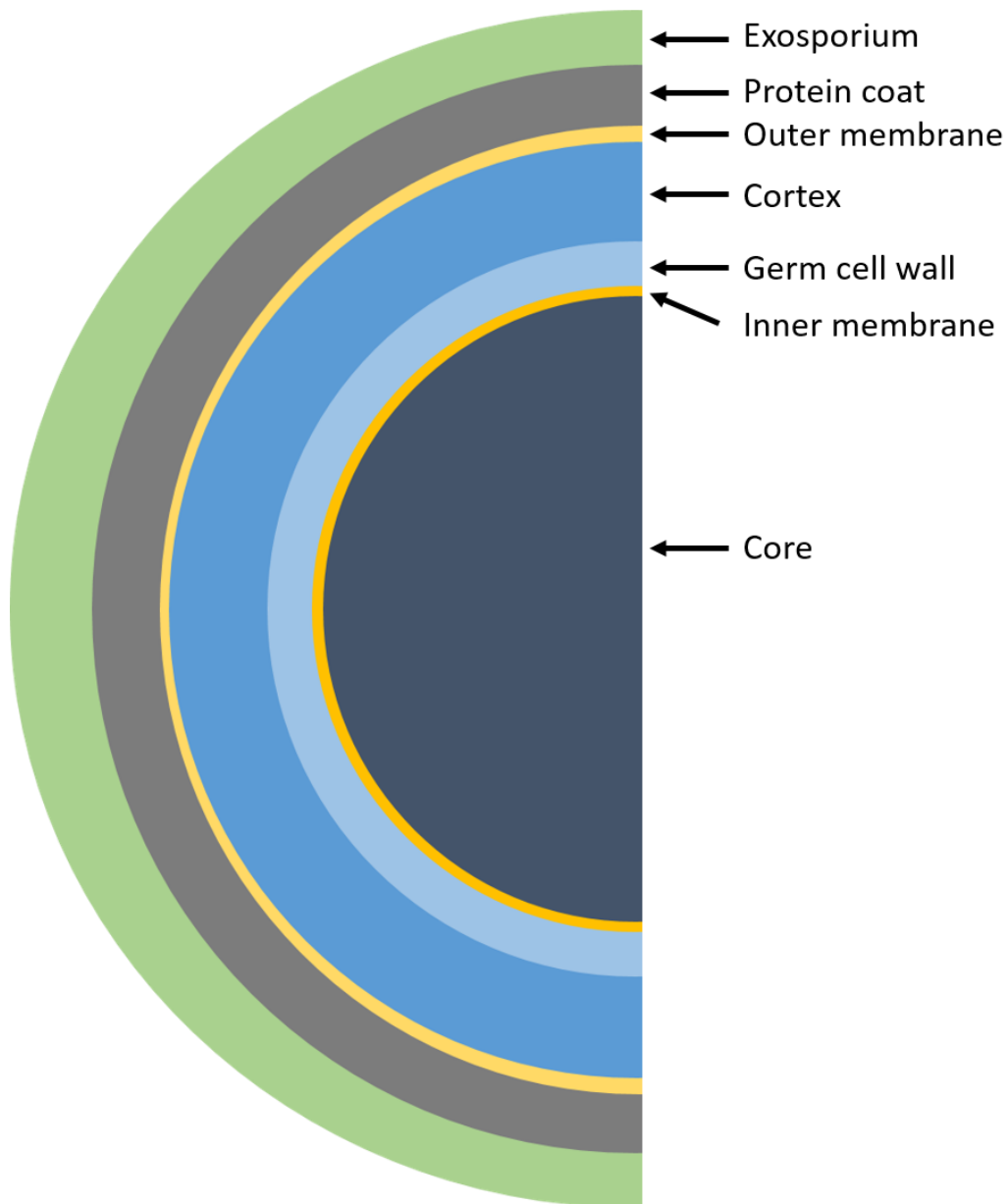
Fig.1.2: The structure of the *C. difficile* spore. The innermost compartment of the spore, the core, contains the genome in a largely dehydrated cytosol. Ca-DPA constitutes approximately 25% of the dry weight of the core (Paredes-Sabja et al., 2014). The core is encased by a low-permeability inner membrane which protects the core from DNA damaging molecules. Surrounding the inner membrane is the germ cell wall, which becomes the cell wall of the resulting vegetative cell. A thick modified-peptidoglycan layer, the cortex, surrounds the germ cell wall. Modifications to the cortex peptidoglycan allow specific recognition by cortex lytic enzymes required for germination, without hydrolysis of the germ cell wall. A mother cell derived outer membrane surrounds the cortex, on which a proteinaceous coat is assembled. The outer-most layer of the spore is the exosporium (Paredes-Sabja et al., 2014).

1.1.4 The *C. difficile* Toxins

Once established in the gut, *C. difficile* proliferates and releases toxins. The two major toxins of *C. difficile*, encoded by *tcdA* and *tcdB*, are widely accepted as the major causes of CDI disease symptoms. In animal models, administration of purified TcdA (toxin A) and TcdB (toxin B) is sufficient to illicit all symptoms of CDI (Shen, 2012). *tcdA* and *tcdB* are located within a 19.6 kb pathogenicity locus (PaLoc) which also contains the toxin regulatory genes *tcdR*, *tcdE,* and *tcdC* (Hammond and Johnson, 1995)*.* TcdR is an alternative sigma factor essential for toxin production, TcdC is predicted to be an anti-sigma factor that sequesters TcdR and negatively regulates toxin production, and TcdE is a predicted holin thought to be involved in toxin secretion (Dupuy et al., 2008; El Meouche et al., 2013; Tan et al., 2001). Interestingly, in 027 strains, *tcdC* contains a nonsense mutation which has been suggested to contribute to the increased virulence of these strains (Curry et al., 2007), although this is disputed (Cartman et al., 2012) (Bakker et al., 2012).

The enterotoxin, toxin A, and the cytotoxin, toxin B, are two of the largest identified bacterial toxins, belonging to a family of large, glycosylating clostridial toxins. This family of toxins targets Rho and Ras GTPases, modifying them by mono-*O*-glucosylation (e.g. TcdA, TcdB) or by mono-O-N-acetylglucosaminylation (e.g. *Clostridium novyi* α-toxin) (Schirmer and Aktories, 2004). These modifications result in the inhibition of the regulatory function of these targets. Toxin A and toxin B both target Rho GTPases which act as molecular switches controlling host signal pathways involved in many cellular processes, including cytoskeleton assembly (reviewed in (Hall, 2012)), disruption of which leads to loss of intercellular tight-junctions, cell rounding, fluid secretion, and eventual epithelium cell death (Shen, 2012). In addition, the toxins trigger the release of pro-inflammatory cytokines from immune cells and intestinal epithelial cells. This leads to an influx of neutrophils and further damage to the intestinal epithelia. Additionally, poor prognosis is associated with an influx of neutrophils into the peripheral blood, which may

suggest that over-stimulation of the immune system during infection may be responsible for much of the CDI pathology (Shen, 2012).

The *C. difficile* toxins are composed of four functional domains, an N-terminal glucosyltransferase, an autoprotease domain, a hydrophobic pore-forming delivery domain, and a C-terminal combined repetitive oligopeptides (CROPs) domain (Chumbler et al., 2016). The CROPs are believed to be involved in receptor binding. The crystal structure of toxin A, lacking the CROPs domain, has now been resolved to 3.25 Å (Chumbler et al., 2016). This structure has been modelled into a previous EM structure revealing the overall organisation of the four domains of toxin A. The structure of toxin B is predicted to be similar (Chumbler et al., 2016; Pruitt et al., 2010).

The exact contributions of the two major toxins to virulence has been a topic of great debate. Early experiments suggested that toxin A was the more potent, being sufficient to cause the symptoms of CDI in rodents, while toxin B required co-administration with a sub-lethal concentration of toxin A (Lyerly et al., 1985). However, toxin B was shown to be able to cause symptoms of CDI in humanised mice (Savidge et al., 2003). This may be due to poor expression of the toxin B receptors in standard rodent models. More recent analysis, using isogenic toxin mutants, showed that toxin B is essential for virulence in the hamster model of infection (Lyras et al., 2009). TcdA$^+$/TcdB$^-$ mutants were found to avirulent (Lyras et al., 2009) although this is partially disputed (Kuehne et al., 2010). Although naturally occurring TcdA$^-$/TcdB$^+$ strains are frequently isolated, TcdA$^+$/TcdB$^-$ strains have never been isolated from patients with CDI. Interestingly, for unknown reasons, TcdA$^-$/TcdB$^+$ strains may cause more severe disease than TcdA$^+$/TcdB$^+$ strains (Drudy et al., 2007). The exact contributions of toxin A and toxin B to virulence remains unclear.

Several *C. difficile* strains also produce an additional toxin, the *C. difficile* binary toxin (CDT), which belongs to the ADP-ribosyltransferase toxin family. These toxins are composed of

two peptide subunits, an ADP-ribosyltransferase (encoded by *cdtA*) and a binding/translocation

subunit (encoded by *cdtB*) (Aktories and Barth, 2004). The exact contribution of CDT to virulence

is not currently understood and CDT alone is not sufficient to cause symptoms of CDI, although

it has been associated with increased adherence to the host (Schwan et al., 2014) and can induce

pathogenic inflammation via TLR2 (Cowardin et al., 2016). However, patients infected with *C.

difficile* strains expressing CDT have 60% greater mortality rates than those infected with CDT

deficient strains (Bacci et al., 2011) (Kuehne et al., 2014).

1.1.5 Treatment of CDI

Until recently, there has been little interest in the development of new treatments for

CDI as severe infections were uncommon and the disease usually responded well to antibiotic

treatment. Although recurrent infections were common, they were easily managed. However,

due to the emergence of epidemic strains, and the increased incidence and severity of CDI,

research into new treatments has increased over the past decade.

1.1.5.1 Antibiotic Treatment of CDI

Typically, metronidazole and vancomycin are used to treat CDI. Metronidazole is used

to treat moderate cases of CDI, while vancomycin is reserved for severe, complicated, and

recurrent infections (Debast et al., 2014). Metronidazole is less effective, with cure rates of

approximately 70% compared to 80% for vancomycin (Johnson et al., 2014). Vancomycin use

has been restricted in the past due to the fear of selecting for outgrowth of vancomycin-resistant

Enterococcus, although evidence suggests that these restrictions are not warranted (Debast et

al., 2014). As both these antibiotics are broad spectrum, their use results in further damage to

the microbiota, leaving the patient susceptible to recurrent infection. 20-30% of CDI cases result

in recurrent infection either through relapse or re-infection (Figueroa et al., 2012). As prolonged

use of metronidazole can result in neurotoxicity, it is not recommended for use beyond first

recurrence of CDI. Instead, tapered/pulsed doses of vancomycin are used to prevent *C. difficile*

growth and allow recovery of the microbiota (Debast et al., 2014). However, after 2-3 recurrences, the chance of further recurrence increases to 40-60% (Johnson, 2009).

In 2011, the macrocyclic antibiotic, fidaxomicin, was approved by the European medicines Agency and the US Food and Drug Administration Agency for treatment of CDI (Venugopal and Johnson, 2012). Fidaxomicin use has been shown to be as effective as vancomycin, with relatively minor effects on the microbiota and a significantly reduced risk of recurrence (Louie et al., 2012; Mullane et al., 2011). The use of fidaxomicin, however, may be limited due to the high cost of a full course of this antibiotic (Bartsch et al., 2013).

1.1.5.2 Faecal Transplantation

Faecal microbiota transplantation (FMT) involves the restoration of the natural microbiota using faecal matter donated from a healthy individual. FMT is not a new concept, with reports indicating that diarrhoea treatments using orally introduced faecal suspensions were used by fourth century Chinese physicians (Zhang et al., 2012). FMT was first used to treat pseudomembranous colitis in 1958 by the administration of faeces by enema (Eiseman et al., 1958). However, it has only been in the last decade that FMT has gained popularity as a treatment for CDI. The use of FMT for recurrent CDI infections has been reported to have success rates of over 90% (Brandt et al., 2012; Oprita et al., 2016; van Nood et al., 2013).

Although FMT shows great promise as a treatment for CDI, its use is limited for several reasons. Perhaps understandably, due to the unconventional nature of this treatment, there is reluctance amongst the general public to accept this treatment and for medics to perform it. In addition, there are obvious dangers that must be addressed in the clinical use of faecal matter. FMT involves introducing many bacteria and viruses into patients, necessitating the need for thorough screening of donor material to avoid secondary infection transmission (as reviewed in (Gupta et al., 2016)). Standardisation of FMT application must also be performed, including the time of sample collection and preparation, sample storage, donor selection, and route of

administration. This has led to the establishment of stool banks, such as the Netherlands Donor Faeces Bank, which has attempted to establish procedures for sample preparation and storage (as reviewed in (Terveer et al. 2017)). In addition, the contribution of the microbiota to human health is practically unknown, although there is a lot of evidence showing a role in human metabolism (as reviewed in (Clemente et al., 2012)). The long-term effects of manipulation of the microbiota may be drastic. Indeed, there are reports of patients exhibiting significant weight gain after FMT (Alang and Kelly, 2015).

Research is now being undertaken to identify the constituents of the microbiota required for CDI resolution, to make this treatment more reproducible and palatable. Administration of six phylogenetically distinct, intestinal bacterial species is sufficient to clear CDI in the mouse model (Lawley et al., 2012). Complementing this observation, a recent study showed that administration of a preparation of thirty-three bacterial species was sufficient to cure two patients infected with ribotype 078 *C. difficile.* This treatment was comically referred to as rePOOPulating the gut (Petrof et al., 2013). However, the first treatment based on this observation, a pill containing stool derived bacterial spores, recently failed stage II clinical trials (Ratner, 2016).

1.1.5.3 Diffocins and Avidocin-CDs

Due to the high rates of recurrence, there is a need for *C. difficile* specific therapies that do not cause damage to the microbiota. One potential therapy is the use of Diffocins. These are high molecular weight bacteriocins that resemble the R-type Pyocins of *Pseudomonas aeruginosa,* naturally synthesised by *C. difficile* to kill competing strains (Gebhart et al., 2012; Nakayama et al., 2000).

These structures resemble *Myoviridae* phage tails, consisting of a nanotube core and contractile sheath coupled to receptor binding proteins (RBPs) via a baseplate and tail fibres. Recognition of a cell surface receptor by the RBPs triggers contraction of the sheath, which in

turn pushes the needle-like core through the envelope of the bacterial cell. The core forms a small pore in the cell envelope causing leakage of ions from the cell, dissipation of the membrane potential, and eventually death (Scholl et al., 2009). A single Diffocin binding to the cell surface is therefore theoretically sufficient to result in killing. The genomes of *C. difficile* strains were mined, identifying a novel phage RBP (PtsM) which was used to modify the Diffocin structure. The modified Diffocin, termed Avidocin-CD, showed increased stability compared to naturally occurring Diffocins (Gebhart et al., 2015).

Avidocin-CDs administered to mice via oral gavage remained active during transit in the mouse GI tract and were effective prophylactic agents. In addition, treatment with Avidocin-CDs did not result in a significant alteration to the microbiota of the mouse (Gebhart et al., 2015). These agents represent a promising alternative therapy for CDI.

1.2 Bacterial S-layers

Bacterial S-layers are proteinaceous, paracrystalline arrays that coat the outside of the cell. They are commonly found on Gram-positive and Gram-negative bacteria, and are ubiquitously in the Archaea. S-layers are composed of a single, or multiple, proteins. These S-layer proteins (SLPs) are often glycosylated, and spontaneously self-assemble to form a 2D array containing regularly repeating units (Fagan and Fairweather, 2014). S-layers were first discovered in the 1950s (Houwink, 1953) and since then, structural data has been obtained for a small number of S-layers (see below). Despite their ubiquity, little is understood about the functions of these structures, partly due to the model organisms *E. coli* and *B. subtilis* lacking S-layers. However, the S-layer shows potential as an antimicrobial target (as reviewed in (Kirk et al., 2017)).

1.2.1 S-layer Structure

Since the discovery of S-layers their structures have been investigated with varying techniques. Some of the most interesting structural information was obtained using electron microscopy of cells which has shown that different S-layers form arrays with oblique (P1 & P2), square (P4), and hexagonal (P6) symmetries (Fagan and Fairweather, 2014). Information on intact S-layer structure is limited however, largely due to the difficulties in obtaining 3D crystals of SLPs and the relatively low resolution obtained from electron microscopy experiments.

Recently, the structure of the *Caulobacter crescentus* S-layer has been resolved (Bharat et al., 2017). Electron cryotomography was used to determine the native structure of the S-layer to 7.4 Å, and X-ray crystallography enabled the structure of residues 249-1026 of the major SLP of *C. crescentus,* RsaA, to be resolved to 2.7Å (Fig.1.3A). The X-ray structure was fit into the electron cryotomography map, providing a near complete structure of the native S-layer to high resolution. The structure is arranged in a repeating hexameric lattice, with a central pore of approximately 20 Å and other gaps of no more than 27 Å. Within the structure there are bound Ca^{2+} ions, found at the lattice interfaces. Without addition of Ca^{2+} or Sr^{2+}, no RsaA crystals were obtained and no sheets were observed in EM experiments, suggesting that Ca^{2+} is necessary for S-layer formation and stabilisation (Bharat et al., 2017).

The S-layer of *C. difficile* is mainly composed of two proteins, the high molecular weight (HMW) and low molecular weight (LMW) SLPs, created through the cleavage of a pre-protein encoded by *slpA*. The HMW and LMW SLPs form a heterodimer non-covalently anchored to the cell wall by three cell wall binding domains (CWB2) in the HMW SLP, with the LMW SLP exposed to the environment. Although the structure of the intact *C. difficile* S-layer has not been determined, the structure of the N-terminal of the LMW SLP is known (Fagan et al., 2009). In addition, the structure of the cell wall-binding domains of Cwp8 and Cwp6 which show high sequence similarity to the CWB2 domains of the HMW SLP, have been recently determined

(Usenik et al., 2017). Like the *C. crescentus* S-layer, the HMW and LMW SLPs are able to form 2D arrays *in vitro* in the presence of Ca^{2+} (Takumi et al., 1992). The LMW SLP contains three distinct domains. Domain 2 contains a novel protein fold, is exposed to the environment and shows the most sequence variability between *C. difficile* strains (see section 1.2.3). Domain 1 forms a two-layer sandwich motif and the third, interaction domain, anchors the LMW SLP to the HMW SLP (Fig.1.3B) (Fagan et al., 2009). The HMW SLP contains three CWB2 domains, similar to Cwp8 and Cwp6. The CWB2 domains of Cwp8 and Cwp6 form pseudo-trimers, forming a rough triangular disk shape with a diameter of approximately 60 Å (Fig.1.3C). As this structure is thought to anchor the proteins to the cell wall, it is suggested that the pseudo-trimeric structure is an example of convergent evolution as the Surface Layer Homology (SLH) anchoring domains, that anchor the cell wall proteins of *Bacillus* to the cell wall (see section 1.2.4.1), also form pseudo-trimers (Usenik et al., 2017).

A

B

C
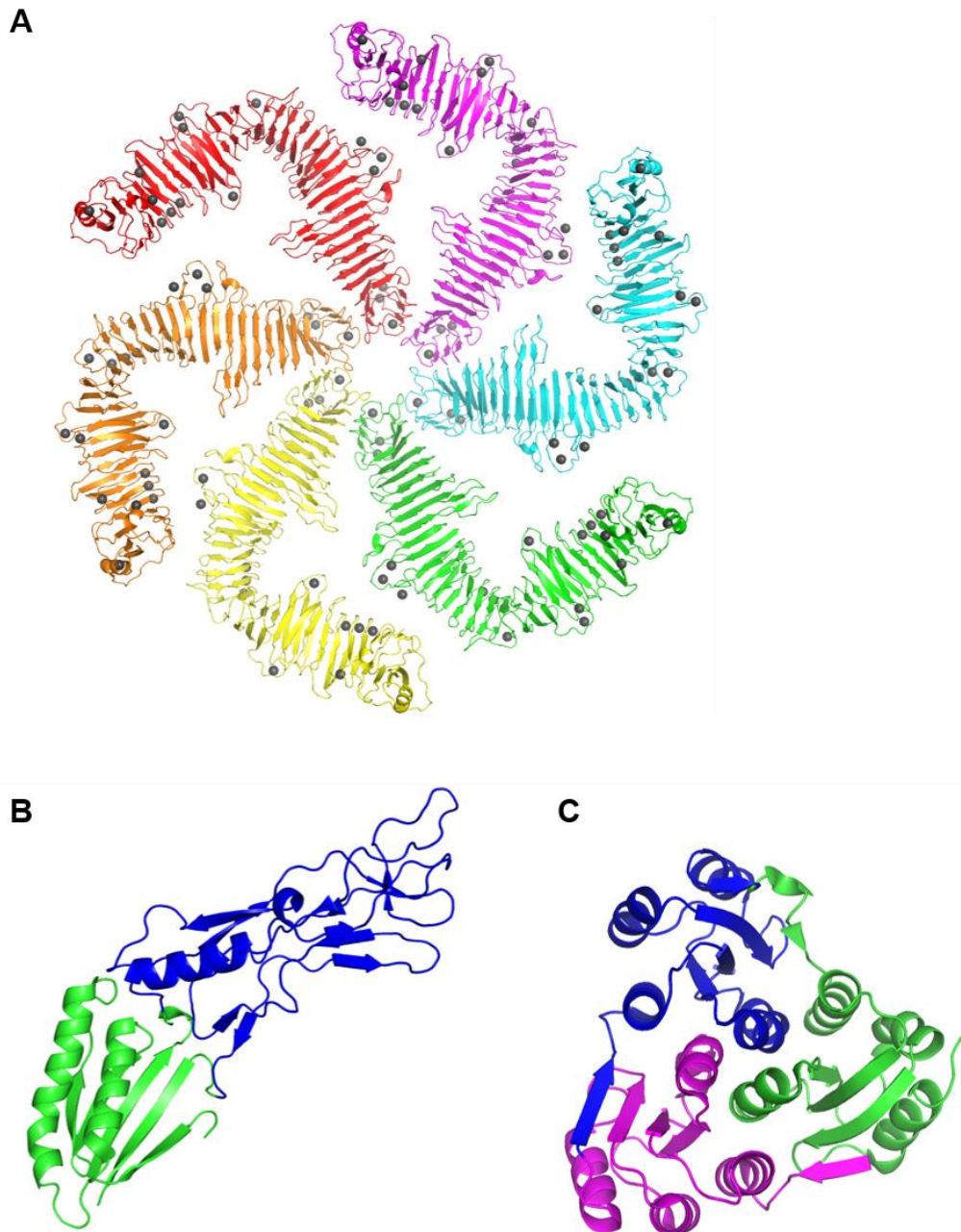
Fig.1.3: SLP structures. (A) Structure of the native S-layer of *C. crescentus*. The *C. crescentus* S-layer is the most complete, native structure of an S-layer to date, forming a hexameric array with a 20 Å central pore. Interaction between each subunit is stabilised by Ca^{2+} atoms (grey) (Bharat et al., 2017). **(B)** Incomplete structure of the *C. difficile* LMW SLP. Shown are the surface exposed domain 2 (blue) and domain 1 (green). The structure of the interaction domain has not yet been resolved (Fagan et al., 2009). **(C)** The CWB2 trimer of Cwp6. Each colour represents a single CWB2 domain (Usenik et al., 2017). Images were created using data obtained from the Protein Data Bank. PBD IDs 5N97, 3CVZ, and 5J72 respectively. Images were created using PyMOL molecular graphics system.

1.2.2 The Cell Wall Protein Family

C. difficile encodes 28 proteins that belong to the Cell Wall Protein (CWP) family (Fig.1.4). These proteins are paralogues of SlpA and contain three CWB2 motifs (Fagan et al., 2011). In addition to the cell wall binding domains, these proteins generally contain an additional domain, thought to provide extra function to the S-layer. Although only a few of these proteins have been characterised experimentally, many are predicted to provide functions typical of cell wall proteins, such as adhesion to the host (e.g. the adhesin Cwp66) (Waligora et al., 2001) and peptidoglycan modification (e.g. the L,D transpeptidase Cwp22) (Peltier et al., 2011).

Cwp84 is a cysteine protease that is involved in S-layer biogenesis, cleaving SlpA into the two SLPs. However, purified Cwp84 also shows enzymatic activity against fibronectin, type IV collagen, and laminin (Janoir et al., 2007). This observation suggests that Cwp84 may play a role during infection. However, as a *cwp84* mutant shows no loss of virulence in animal models of infection*,* this observation may not translate *in vivo.* Cwp13 is a homologue of Cwp84, sharing 63% protein identity. However, this protein is only partially able to complement loss of Cwp84 and recognises an additional site in the HMW SLP. It has therefore been suggested that Cwp13 is involved in proteolysis of misfolded cell surface proteins (de la Riva et al., 2011).

Cwp66 is an adhesin capable of mediating adherence to Vero cells (Waligora et al., 2001), although animal studies using mutants lacking this protein have not been reported. Cwp2 has also been suggested to function as an adhesin (Bradshaw et al., 2017). However, *cwp2* mutants showed no reduction in virulence and were not complemented.

Finally, CwpV, the largest member of the CWP family, provides resistance to infection by bacteriophages and promotes bacterial aggregation (Reynolds et al., 2011; Sekulovic et al., 2015). *cwpV* is under the control of a phase variable promoter, and as a result, only 5% of cells in a population express CwpV under normal lab growth conditions (Emerson et al., 2009). In addition to the cell wall anchoring domains at its N-terminal, CwpV also contains a region of

unknown function linked to a C-terminal region of 4-9 repeats of 79-120 residues (Reynolds et al., 2011). After secretion, CwpV undergoes auto-proteolysis forming a 42 kDa N-terminal fragment and a 90-120 kDa C-terminal fragment that, in a manner analogous to SlpA, form a heterodimeric complex on the cell surface (Dembek et al., 2012). The sequence of the C-terminal domain of this protein varies between strains and, to date, five distinct sequence types have been identified (Reynolds et al., 2011).
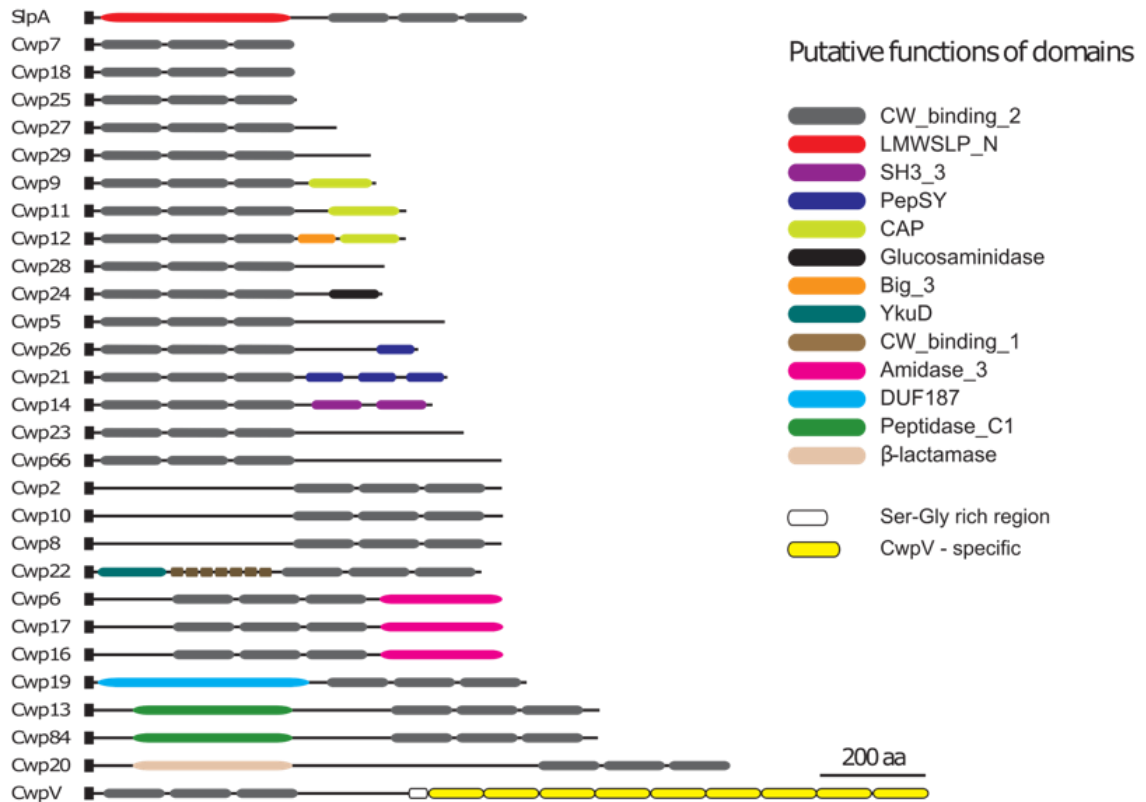
Fig.1.4: The Cell Wall Protein family. (A) SlpA and the 28 members of the Cell Wall protein Family. Each of these proteins contains an N-terminal secretion signal peptide (black boxes) and three CWB2 motifs (PF04122) (grey boxes). Shown are the predicated Pfam domains of each protein. LMWSLP_N (PF12211), SH3_3 (PF08239), PepSY (PF03413), CAP (PF00188), Glucosaminidase (PF01832), Big_3 (PF07523), YkuD (PF03734), CW_binding_1 (PF01473), Amidase_3 (PF01520), DUF187 (PF02638), Peptidase_C1 (PF00112) and β-lactamase (PF00144). Republished with permission of the Microbiology Society, (Fagan et al., 2011), permission conveyed through Copyright Clearance Center, Inc.

1.2.3 S-layer Variation

In addition to the diversity shown by S-layers of different species, S-layers may also vary between strains of a single species. In the *C. difficile* genome *slpA* is located within an approximately 10 kb cassette, along with *secA2, cwp2,* and *cwp66* (Dingle et al., 2013)*.* This cassette shows considerable genetic diversity amongst *C. difficile* strains. To date, 12 divergent S-layer cassette types (SLCT) have been identified. One of these cassettes is considerably larger than the others due to the insertion of a 24 kb region containing genes predicated to be involved in S-layer glycosylation (see section 1.2.4.2).

The S-layer cassette is hypothesised to undergo recombinational switching, enabling antigenic variation. The high degree of variation between *slpA*s from different SLCTs suggests that the S-layer is under selective pressure, potentially due to identification by the host immune system or by acting as the receptor for bacteriophages. This is reflected by the fact that most of the variation between different *slpA*s occurs within the LMW SLP encoding region and the LMW SLP is predicted to be exposed to the environment. It is therefore possible that the LMW SLP acts as the receptor for binding of host immune system effectors and bacteriophages (Dingle et al., 2013). A recent study suggested that the SLCT of a strain may influence its virulence, with the SLPs from ribotype 027 and 078 strains inducing a more potent immune system response (Lynch et al., 2017). This may contribute to the increased virulence observed with these strains. The diversity of SlpA between strains also highlights the challenges of developing antimicrobials that target this structure.

1.2.4 Biosynthesis of the S-layer

Many S-layer-containing bacteria possess a secretion pathway dedicated to the secretion of SLPs, perhaps unsurprising due to the large burden associated with secretion of such an abundant protein. For example, *C. difficile* is estimated to secrete 164 SlpA molecules per second, during exponential growth, to maintain overall S-layer structure (Kirk et al., 2017).

In Gram-negative bacteria, several mechanisms have evolved to overcome this problem. For example, specific type I secretion systems are required for SLP translocation in *C. crescentus* and *C. fetus* (Awram and Smit, 1998; Thompson et al., 1998)*, while *Aeromonas* species use a specific type II system (Noonan and Trust, 1995). The secretion of SLPs has been investigated in two Gram-positive bacteria, *C. difficile* and *B. anthracis,* both of which require the accessory Sec secretion system.

After secretion, *C. difficile* SlpA is cleaved to form the HMW and LMW subunits by the cysteine protease Cwp84 (Kirby et al., 2009). Cwp84 cleavage is essential for S-layer formation, as mutants in the gene encoding this protein have a severe growth defect, abnormal S-layer, and shedding of surface proteins (Kirby et al., 2009). However, *cwp84* mutants do not show any loss of virulence *in vivo,* likely due to host-proteases, such as trypsin, being able to functionally complement the loss of Cwp84. SLPs presumably spontaneously self-assemble, driven thermodynamically (Chung et al., 2010), and anchor to the cell wall. The formation of 2D crystals is mediated by crystallisation domains present in the some SLPs (Baranova et al., 2012), although such a domain has not been identified in either of the *C. difficile* SlpA subunits. Our understanding of S-layer biogenesis in *C. difficile* is summarised in Fig.1.5.
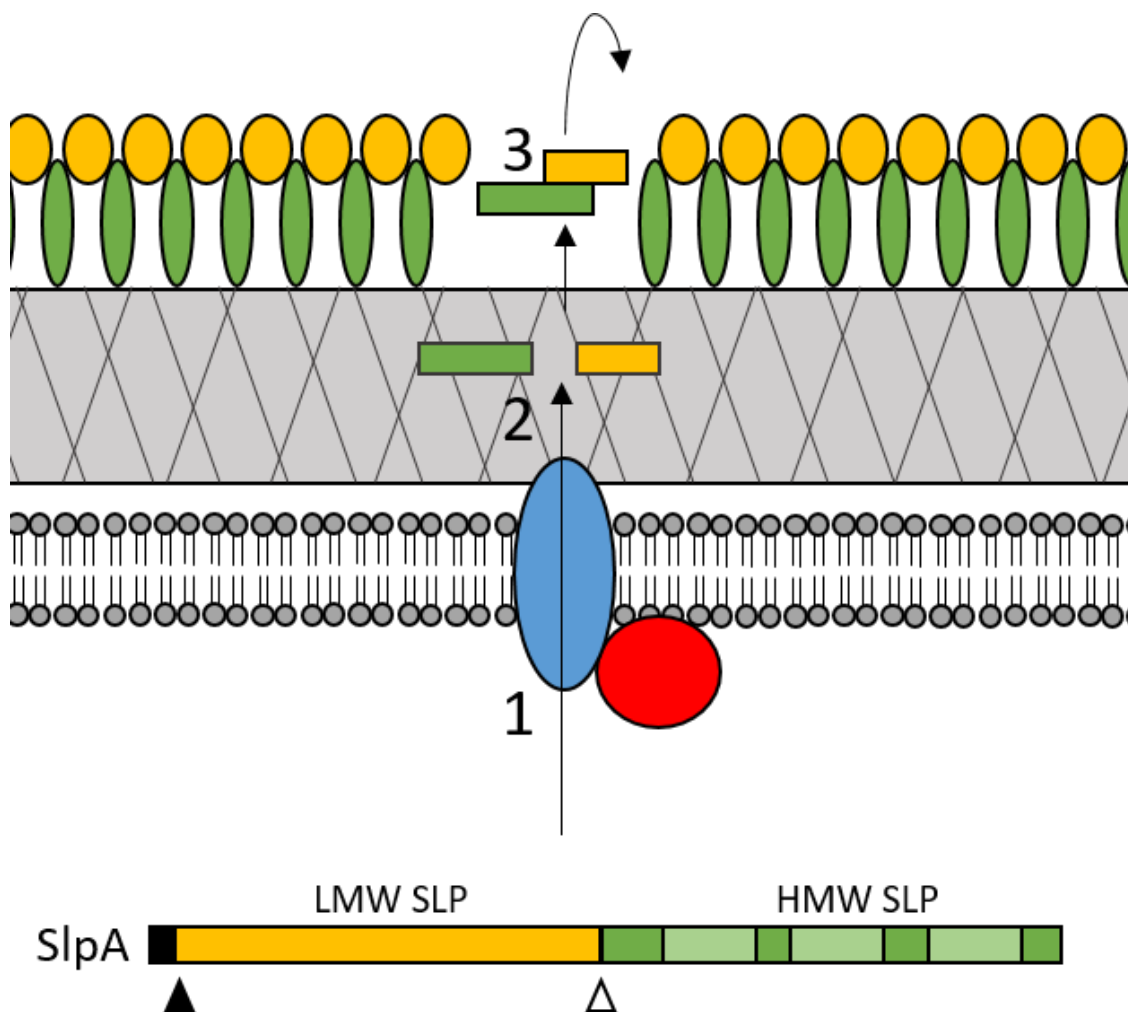
Fig.1.5: Model of *C. difficile* S-layer biogenesis. The *C. difficile* S-layer is largely composed of two proteins created through the cleavage of the preprotein SlpA. **(1)** SlpA is targeted to the Sec secretion channel (blue) by the accessory secretory protein, SecA2 (red). SecA2 is believed to use the canonical SecYEG complex for secretion as *C. difficile* does not express an accessory SecY homologue (SecY2) (Fagan and Fairweather, 2011). SlpA contains a signal peptide (Black) at its N-terminal. **(2)** After SlpA is translocated across the cytoplasmic membrane, its signal peptide is removed and SlpA is cleaved into the LMW SLP (orange) and the HMW SLP (green) by the cysteine protease Cwp84 (not shown) (Kirby et al., 2009). **(3)** The two subunits then spontaneously self-assemble to form a heterodimeric complex. This complex then becomes anchored to the cell wall (crosslinked grey) via the secondary cell wall polysaccharide, PS-II (not shown), via the three cell wall binding domains present in the HMW SLP (Willing et al., 2015). HMW-LMW complexes form a 2D array on the surface of the cell, with the LMW subunit exposed to the environment.

1.2.4.1 Anchoring the S-layer to the Cell Wall

S-layers are non-covalently attached to the cell wall usually via interaction with lipopolysaccharide (LPS) in Gram-negative bacteria and secondary cell wall sugars in Gram-positive bacteria (Fagan and Fairweather, 2014). The anchoring mechanism is poorly understood in Gram-negative bacteria. However, it is known that in *C. crescentus,* RsaA requires an intact O-antigen on the LPS for interaction. Interaction between RsaA and the LPS is mediated by approximately 225 amino acids at its N-terminus (Awram and Smit, 2001; Ford et al., 2007). The N-terminal is also responsible for LPS binding of the *Campylobacter fetus* SLP (Dworkin et al., 1995).

The anchoring mechanisms of Gram-positive S-layers are better characterised. There are two major anchoring modules in Gram-positive SLPs, the SLH and CWB2 domains, both of which act as functional pseudo-trimers. The SLH domains are the most wide-spread, present in the SLPs of *Clostridium thermocellum, B. anthracis,* and *Deinococcus radiodurans* among others*.* This best characterised of which are the major SLPs of *B. anthracis*, Sap and EA1 (Etienne-Toumelin et al., 1995; Mesnage et al., 1997). The three SLH domains present in these proteins act cooperatively to bind pyruvylated secondary cell wall polymers (SCWP) (Mesnage et al., 2000).

The HMW SLP of *C. difficile,* and other members of the Cell Wall Protein family, contain three CWB2 domains. CWB2 domains were first identified in an autolysin, CwlB, from *B. subtilis* and are found within cell wall proteins of many members of the Firmicute family (Fagan and Fairweather, 2014). Recently, it was discovered that the anionic SCWP, PS-II, is responsible for anchoring the S-layer of *C. difficile* to the cell wall (Willing et al., 2015). Interestingly, immediately downstream of the genes encoding several of the major cell wall proteins are several genes essential for growth, thought to be responsible for sugar biosynthesis (Dembek et al., 2015). It is likely that these genes are responsible for PS-II synthesis (Willing et al., 2015).

1.2.4.2 Glycosylation of S-layer Proteins

The S-layers of many bacterial species are glycosylated. The glycosylation patterns of these S-layers show high diversity and varying complexity between species (as reviewed in (Schaffer and Messner, 2017)). Although bacterial surface proteins and archaeal SLPs can be modified with either O-linked or N-linked glycans, bacterial SLPs have only been identified with O-linkages (Ristl et al., 2011).

An S-layer glycan synthesis gene cluster (*slg*) was first identified in *Geobacillus stearothermophilus*, and has been identified in several organisms since. This gene cluster encodes a glycosyltransferase, glycan processing enzymes, and transport machinery required for protein glycosylation. This discovery led to the development of a model for S-layer glycosylation involving the transfer of galactose, from uridine diphosphate α-D-Galactose, to a lipid carrier (Ristl et al., 2011). A linker unit is then formed by the addition of glycans onto which the repetitive glycan chain is assembled. The repetitive glycan chain can be linear or branched and typically contains 15-50 repeat units each composed of 2-6 sugars (Ristl et al., 2011). The glycan chain is then translocated across the cytoplasmic membrane and ligated to the target SLP (Ristl et al., 2011). The overall structure of a typical glycan chain is similar to the LPS of Gram-negative bacteria, which may suggest a common evolutionary origin for LPS and S-layer glycan synthesis pathways (Schaffer et al., 1996).

The S-layer of *C. difficile* is usually not glycosylated (Qazi et al., 2009). However, one SLCT (SLCT-11) was shown to contain an insertion of approximately 24 kb, resembling the *slg* gene cluster (Dingle et al., 2013). The LMW subunit of SLCT-11 SlpA is much smaller than other SlpA types. Confirmation of S-layer glycosylation in these strains, and investigation of its contribution to virulence, are yet to be performed.

1.2.5 Functions of the S-layer

As the S-layer is the major proteinaceous component of the cell surface, it is unsurprising that many different functions have been assigned to these structures. Despite this, no single function has been assigned to all the S-layers and its function is unknown in most bacterial species. This is especially true for *C. difficile,* partly due to the previous lack of genetic tools for *C. difficile* research, but also because the S-layer of *C. difficile* appears to be essential for growth *in vitro* (Dembek et al., 2015). There has been some research into its contribution to adherence to the host, but research into its function has been seriously hampered due to the lack of *slpA* mutants. However, due to the large metabolic sacrifice required to create and maintain these structures, coupled with the observation that it is essential for growth, the S-layer likely provides vital functions to the cell.

1.2.5.1 S-layers as a Protective Coat

As the S-layer form a contiguous array coating the cell, it is believed that it forms a molecular sieve, essentially acting as a size exclusion barrier (Sara and Sleytr, 1987). The atomic structure of the *C. crescentus* S-layer revealed pores no larger than 27Å (Bharat et al., 2017). This would theoretically prevent some harmful material such as phage and large globular enzymes such as lysozyme from reaching the cell wall. The S-layer may therefore provide resistance to environmental stresses. Reported pore sizes of S-layers vary significantly (Fagan and Fairweather, 2014). Further investigation into the atomic structure of these complexes may provide further insight into S-layer barrier function. The S-layer of *C. crescentus* has also recently been shown to provide resistance to antimicrobial peptides, including the human cathlicidin LL-37 (de la Fuente-Nunez et al., 2012). These peptides kill bacteria by oligomerising and forming toroidal pores in the cytoplasmic membrane (Vandamme et al., 2012). Due to the small size of LL-37, it is likely that this peptide could pass through the pores of the S-layer. However, as SLPs tend to have low isoelectric points (3.46 for RsaA of *C. crescentus*), this resistance may be

mediated through interaction of the cationic antimicrobial peptide with the negatively charged residues of the S-layers. This is thought to sequester the peptide at the S-layer, preventing them from reaching the cell membrane (de la Fuente-Nunez et al., 2012).

The parasite *Bdellovibrio bacterovirus* is a bacterium that enters the periplasm of other bacteria where it replicates, digesting and eventually lysing its prey (Sockett, 2009). *B. bacterovirus* is not normally able to parasitise *C. crescentus.* However, S-layer mutants are susceptible to predation indicating that the S-layer provides protection from *B. bacterovirus* killing (Koval and Hynes, 1991).

1.2.5.2 S-layers in Pathogenesis

As the S-layer coats the cell surface, it is unsurprising that in many bacteria it has been associated with a role in pathogenesis. For example, *C. fetus,* the most common cause of abortion in cattle and also a cause of human systemic infection, expresses an S-layer essential for virulence (Blaser, 1993). Loss of its S-layer results in increased phagocytosis and sensitivity to killing by the complement system (Blaser, 1993). The S-layer of *C. fetus* is also antigenically variable, enabling evasion of the host immune system (Thompson, 2002). S-layers may also function as adhesins, facilitating binding of the bacterium to host tissue. The major SLP of *Lactobacillus crispatus*, present in the avian and human gut microbiota, can bind to type I and IV collagen (Sillanpaa et al., 2000).

Although the exact contribution of the *C. difficile* S-layer to virulence remains elusive, aspects of its role in pathogenicity have been investigated. Purified *C. difficile* SLPs can induce the release of pro-inflammatory cytokines from human monocytes, maturation of monocyte derived dendritic cells, and enhance allogenic T-cell proliferation (Ausiello et al., 2006). More recent studies using the mouse model of infection have shown that the SLPs induce maturation of bone marrow derived dendritic cells, important for the induction of helper T-cells, and cytokine production (Ryan et al., 2011). Toll like receptor 4 (TLR4) was responsible for dendric

cell maturation and TLR4 knock-out mice show increased sensitivity to *C. difficile* infection presenting more severe symptoms (Ryan et al., 2011). Activation of dendritic cells required both the HMW and LMW SLPs, suggesting that TLR4 recognises the intact complex. Based on these observations, it appears that TLR4 is important for the clearance of *C. difficile* infections, mediated by recognition of the S-layer complex (Ryan et al., 2011). Recently it has been suggested that the S-layers of different SLCTs may elicit different immune responses (Lynch et al., 2017). More potent macrophage responses were observed using the SLPs of ribotype 027 and 078 relative to 001 and 014 (Lynch et al., 2017). In addition, the S-layer may function as an adhesin. Although the lack of a knock-out mutant makes the study of this function difficult, SlpA has been shown to bind to fixed gut enterocytes (Calabi et al., 2002). This was complemented by the observation that isolated SLPs can inhibit binding of *C. difficile* to Caco-2 cells (Merrigan et al., 2013).

1.2.5.3 Additional Roles of the S-layer

S-layers have, due to their great diversity, been associated with many additional functions. The structure of the *C. crescentus* S-layer revealed similarities to antifreeze proteins and a surface lined with threonine residues, which suggests a role in surface hydration and ice crystal avoidance (Bharat et al., 2017). The *B. anthracis* SLPs Sap and EA1 have peptidoglycan hydrolase activity despite lacking obvious functional domains (Ahn et al., 2006). Indeed, *sap* mutants display an increased cell chain length. This phenotype can be complemented through exogenous addition of purified BslO, a protein involved in cell division, to cell culture medium. BslO also contains three SLH domains, and is a member of the *B. anthracis* S-layer associated protein (BSL) family, along with 23 other proteins including EA1 and Sap (Kern et al., 2012). Sap localises on the lateral cell wall, while BslO localises to the septum. This suggests that the Sap S-layer may exclude other SLH containing proteins, enabling localisation of other functional proteins to areas not occupied by Sap, such as the septum (Kern et al., 2012). It has also been speculated that while S-layers could not determine bacterial shape, they may have a role in its

maintenance (Engelhardt, 2007). Although the underlying mechanism remains unknown, two SLPs from marine cyanobacterium, *Synechococcus*, are essential for swimming motility. These proteins are SwmA, a 130 kDa glycosylated SLP, and SwmB, a 1.12 MDa SLP. (Brahamsha, 1996; McCarren and Brahamsha, 2007). Additional genes required for motility encode glycosyltransferases and an ABC transporter (McCarren and Brahamsha, 2009).

1.3 The Sec Secretion and Accessory Pathways

Over 30% of bacterial proteins are exported to extra-cytoplasmic locations where they perform their functions (Tsirigotaki et al., 2017). These proteins, which constitute the secretome, are embedded in the membrane, bind to the cell wall, or localise in the periplasmic space or the extracellular environment. Several bacterial pathways have evolved to translocate proteins across the cytoplasmic membrane including the Tat system and the type I-VI systems (as reviewed in (Green and Mecsas, 2016)). However, most members of the secretome are translocated via the Sec system, a secretion pathway found in all domains of life.

1.3.1 Targeting Proteins to the Sec Pathway

The Sec secretion system translocates proteins via two distinct pathways. Membrane proteins generally utilise the co-translational pathway, in which nascent proteins emerging from a ribosome are targeted to the SecYEG channel and translocated whilst still being synthesised (Xie and Dalbey, 2008). Other proteins are translocated post-translationally, where they are maintained in an unfolded state by a chaperone, and targeted to the SecYEG channel by their N-terminal signal peptide.

Integral membrane proteins secreted co-translationally are recognised by the signal recognition particle (SRP) by their hydrophobic transmembrane domains. The ribosome and the nascent peptide chain are directed to the SecYEG complex by the SRP through interaction with its membrane associated receptor, FtsY (Parlitz et al., 2007; Valent et al., 1997). Once inserted into

the SecY channel, a lateral gate within the pore opens allowing the insertion of transmembrane helices into the membrane (Egea and Stroud, 2010).

Peptides secreted post-translationally are synthesised as pre-proteins, and are distinguished from the other members of the proteome by the presence of a N-terminal signal peptide. The signal peptide is largely conserved, containing a short positively charged N-terminus, a hydrophobic core, and a polar C-terminus (von Heijne, 1990). The C-terminus contains a signal peptidase cleavage site to allow removal of the signal peptide post-secretion (Dalbey and Wickner, 1985). As proteins must be unfolded to allow Sec-mediated secretion, post-translationally secreted proteins must be maintained in an unfolded conformation. This can be achieved by the binding of cytoplasmic chaperones such as trigger factor, SecA, housekeeping chaperones, or specialised Sec chaperones such as SecB in Gram-negative bacteria. SecA then targets the protein to SecYEG (Tsirigotaki et al., 2017).

1.3.2 Secretion Through the Sec System

The secretory machinery is composed of two major elements, the channel complex SecYEG, through which proteins are translocated, and the motor protein SecA, which provides energy for translocation. SecY contains ten transmembrane helices which form the channel pore in the cytoplasmic membrane (Van den Berg et al., 2004). In its resting conformation, hydrophobic residues in the middle of the pore restrict its diameter to 5 Å, and the pore is capped by a helical plug. These two elements may act as channel gates (Park and Rapoport, 2011). SecE is bound to the cytoplasmic protrusions of SecY, stabilising its structure. Both SecE and SecY are essential for protein translocation. The non-essential SecG forms hydrophobic interactions with two of the transmembrane helices of SecY, with its cytoplasmic loop sealing the cytoplasmic side of the SecY pore (Tanaka et al., 2015).

Several SecA structures have now been resolved to high resolution, the first being that of *B. subtilis* (Hunt et al., 2002). SecA is composed of several domains. The DEAD motor box

domain, composed of Nucleotide binding domain 1 and 2 (NBD1 and NBD2), catalyses the hydrolysis of ATP, providing the energy for proteins translocation (Hunt et al., 2002). Stemming off of NBD1 is the peptide crossing linking domain (PPXD) which is involved in binding to protein substrates (Papanikou et al., 2005). The helical scaffold domain (HSD) forms a long alpha helix that binds all SecA domains and connects the motor domains with the specificity domain. The helical wing domain (HWD) connects the intra-molecular regulator of ATP hydrolysis 1 (IRA1), believed to function in inserting the protein substrate into the SecY channel (Erlandson et al., 2008). Finally, the C-terminal domain (CTD) is believed to interact with the SecB chaperone in Gram-negative bacteria and phospholipids (Breukink et al., 1995).

Although the exact mechanism by which proteins are translocated through the SecYEG channel remains elusive, decades of research have illuminated some of the key features of Sec secretion (as reviewed in (Tsirigotaki et al., 2017)). During post-translational export, SecA recognises the signal peptide on the pre-protein substrate and targets the complex to the SecYEG channel through interactions with SecY and membrane lipids (Breukink et al., 1995; van der Sluis et al., 2006). The signal peptide is then inserted into the SecY channel and SecA undergoes cycles of ATP binding and hydrolysis. ATP hydrolysis is associated with large changes in the conformation of SecA, which facilitates the translocation of the pre-protein (Economou and Wickner, 1994; Robson et al., 2009). The IRA1 domain of SecA is thought to insert into the SecY channel during these cycles to facilitate translocation (Zimmer et al., 2008). Once the protein has been translocated, the peptide signal is removed by specific signal proteases and the mature protein folds into its functional conformation (Dalbey and Wickner, 1985).

1.3.3 The Accessory Sec Secretion Pathway

Recently it has come to light that many bacterial species contain an accessory Sec secretion pathway (as reviewed in (Feltcher and Braunstein, 2012)). This is characterised by the presence of an additional SecY and/or SecA, termed SecY2 and SecA2. The accessory system

secretes a small subset of proteins while the canonical Sec machinery acts as the housekeeping pathway, secreting the majority of proteins.

SecA2-SecY2 systems are well conserved in *Streptococcus* and *Staphylococcus* species. *secA2* and *secY2* are located on a genetic locus along with a gene encoding a large serine rich protein secreted by the accessory system, genes encoding glycosylation machinery responsible for the glycosylation of this substrate, and genes encoding export machinery of unknown function (Feltcher and Braunstein, 2012). It appears that the SecA2-SecY2 systems have evolved to secrete these large serine rich proteins, which contain an abnormally large N-terminal peptide sequence and are involved in adhesion to the host (Siboo et al., 2005). Interestingly, one such protein, GspB, is heavily glycosylated, and this modification occurs in the cytoplasm (Bensing et al., 2004). *B. anthracis* also contains a SecA2-SecY2 system, although it is phylogenetically distinct from the *Staphylococcus* and *Streptococcus* systems, and *secY2* is not found in the same locus as *secA2* (Rigel and Braunstein, 2008). SecA2 in *B. anthracis* is required for the efficient secretion of the S-layer proteins Sap and EA1 (Nguyen-Mau et al., 2012).

C. difficile contains a SecA2-only system, lacking a SecY2 homologue. SecA2 of *C. difficile* is the only identified SecA2 that is essential for growth *in vitro.* SecA1 and SecA2 function independently in *C. difficile.* Knock-down experiments showed that depletion of SecA1 did not affect the export of SecA2 substrates (Fagan and Fairweather, 2011). Due to the lack of a SecY2 homologue, it is believed that SecA2 secretion utilises the canonical SecY channel complex. In support of this model, experiments were performed using dominant negative variants of SecA, SecA1K106R and SecA2K106R. These variants are unable to hydrolyse ATP and are therefore predicted to be unable to dissociate from the SecYEG complex, effectively blocking them. Expression of SecA1K106R resulted in diminished secretion of SecA2 substrates, providing evidence that SecA1 and SecA2 share the canonical SecYEG channel (Fagan and Fairweather, 2011). The only identified substrates of SecA2 in *C. difficile* to date are SlpA and CwpV although

it has been suggested that SecA2 dependent secretion may be a characteristic of more, if not all, of the members of the CWP family (Fagan and Fairweather, 2011). However, the exact mechanism for the specificity of SecA2 for its substrates remains elusive, as SlpA and CwpV contain what appear to be normal signal peptide sequences. In addition, the only shared motif of SlpA and CwpV are their CWB2 domains. However, the CWB2 domains of SlpA are located on the C-terminus, whereas they are N-terminal in CwpV and are therefore unlikely recognition sites for SecA2 binding. One of the best characterised SecA2-only system is that of *Mycobacterium tuberculosis*. *M. tuberculosis* SecA2 (*Mtb*SecA2) is responsible for the secretion of several virulence factors, and is not essential for growth *in vitro*; mutants show diminished growth in models of infection (Braunstein et al., 2003). The crystal structure of *Mtb*SecA2, the only SecA2 structure solved to date, revealed that it is smaller than SecA1 due to truncations in the HWD and CTD (Swanson et al., 2015) which may have effects on substrate specificity.

1.4 Project Aims

The S-layer of *C. difficile* is likely to provide vital functions to the cell, owing to the metabolic cost associated with its synthesis, and the fact that it is essential for growth *in vitro.* The lack of an SlpA-deficient mutant has severely hindered research into the function of this array. However, challenging growing cultures of *C. difficile* with the recently created Avidocin-CDs has resulted in the isolation of spontaneous mutants resistant to these antimicrobials that lack a functional S-layer, allowing analysis of S-layer function for the first time.

During this project, we aimed to identify the major functions of the S-layer, and its contribution to virulence, through characterisation of the S-layer deficient mutant. In addition, we also investigated the properties of this mutant that allow its growth without an S-layer, to gain insight into why the S-layer of *C. difficile* is usually essential for growth. In addition, as little is known about the biogenesis of the S-layer, crystallographic analysis of the SecA2 structure and microscopic analysis of its intracellular localisation was performed. This was done in the

hope of identifying features of SecA2 allowing its substrate specificity and to determine where the S-layer is assembled.

"New Class of Precision Antimicrobials Redefines Role of *Clostridium difficile* S-layer in Virulence and Viability", a manuscript we have recently had accepted for publication in Science Translational Medicine, has been included as chapter III of this thesis. This paper describes the characterisation of the recently isolated S-layer mutant, FM2.5, and led to many of the observations described in chapter IV of this thesis. The version of the manuscript included is the one accepted for publication posted with permission from the AAAS.

Our published paper "Heat shock increases conjugation efficiency in *Clostridium difficile*" is also included in the Appendices. Work presented in this paper describes the optimisation of a conjugation protocol essential for the mutagenesis described in chapters IV and V.

Chapter II

Materials and Methods

2.1 Handling of Bacterial Strains

2.1.1 Bacterial Strains and Growth Conditions

Bacterial strains used are described in table A1. *C. difficile* strains were routinely cultured on pre-reduced brain heart infusion (BHI) agar (Sigma), BHIS (BHI supplemented with 0.5% w/v yeast extract and 0.1% w/v cysteine) agar, TY (Dupuy and Sonenshein, 1998) (3% tryptose, 2% yeast extract (Bacto)) agar or *Clostridium difficile* Minimal Medium (CDMM) (table 2.1)(Karlsson et al., 1999). Liquid cultures of *C. difficile* were grown using pre-reduced TY broth, TYG broth (TY broth containing 0.5% glucose), or BHIS broth. Cultures were supplemented with anhydrotetracycline (Atc) (4-700 ng/ml), thiamphenicol (15 µg/ml), cycloserine (250 µg/ml) or lincomycin (5 µg/ml) when appropriate. *C. difficile* was grown in an anaerobic work station (Don Whitley Scientific) containing an environment composed of 80% N_2, 10% CO_2 and 10 % H_2 at 37°C.

E. coli strains were routinely grown on LB agar (Sigma) or in LB broth (Fisher Chemicals). Cultures were supplemented with chloramphenicol (15 µg/ml), kanamycin (50 µg/ml) and/or carbenicillin (50 µg/ml) when appropriate. *E. coli* strain CA434 was used as a conjugation donor for transferring plasmids into *C. difficile. E. coli* stains Neb5α (New England Biolabs) and Top10 were used routinely for cloning and plasmid propagation. Expression for protein purification was carried out using *E. coli* strain Rosetta (Novagen). Cultures were grown with agitation (225 rpm) at 37°C.

Bacterial strains were stored in 20% glycerol at -80°C.

Table 2.1: *C. difficile* minimal medium composition

	Stock concentration (mg/ml)	Final concentration (mg/ml)
Amino acids (5X)		
Casamino acids	50	10
L-Tryptophan	2.5	0.5
L-cysteine	2.5	0.5
Salts (10X)		
Na_2HPO_4	50	5.0
$NaHCO_3$	50	5.0
KH_2PO_4	9.0	0.9
NaCl	9.0	0.9
Trace salts (50X)		
$(NH_4)_2SO_4$	2.0	0.04
$CaCl_2 \cdot 2H_2O$	1.3	0.026
$MgCl_2 \cdot 6H_2O$	1.0	0.02
$MnCl_2 \cdot 4H_2O$	0.5	0.01
$CoCl_2 \cdot 6H_2O$	0.05	0.001
Glucose (20X)		
D-glucose	200	10
Iron (100X)		
$FeSO_4 \cdot 7H_2O$	0.4	0.004
Vitamins (1000X)		
D-biotin	1.0	0.001
Calcium-D-panthothenate	1.0	0.001
Pyridoxine	1.0	0.001
5-fluorocytosine	10	0.05

2.1.2 *Clostridium difficile* Spore Isolation

15 ml TGY broth (3% tryptic soy broth (Bacto), 1% yeast extract, 2% glucose, 0.1% L-cysteine) was inoculated with a single colony of *C. difficile* and incubated O/N. The cultures were then sub-cultured 1:10 in 15 ml of SMC broth (9% Bacto peptone (Bacto), 0.5% proteose peptone (Sigma), 0.1% ammonium sulphate, 0.15% tris base) and allowed to grow to OD_{600nm} 0.4-0.7. 100 µl of the SMC culture was then spread on SMC agar plates (6 for R20291 or FM2.5RW, or 30 for FM2.5). The plates were anaerobically incubated for a further 7 days before growth was harvested in 3 ml of sterile water. The harvested material was then washed by resuspending the pellet in 0.5 ml of ice-cold sterile water. Samples were the centrifuged (5,000 x g, 1 min). The wash step was repeated a further 4 times. Before the final wash, growth from 3-4 plates was pooled. Pellets were then resuspended in 0.5 ml of 20% (w/v) HistoDenz (Sigma). This was then layered on top of 1 ml of 50% (w/v) HistoDenz and centrifuged (15 min at 15,000 x g). Using this method, the spores form a pellet, while vegetative cells and debris remain suspended in the HistoDenz density gradient. The HistoDenz was then removed and the spores washed an additional five times with ice-cold sterile water as before. Purified spores were resuspended in sterile water and were quantified by both absorbance at OD_{600nm} and by counting CFU on BHIS supplemented with 0.1 % sodium taurocholate. Purity of each preparation was checked by phase contrast microscopy. Purified spores were stored at -80°C.

2.2 DNA Manipulation

2.2.1 Genomic DNA Isolation

High quality *C. difficile* genomic DNA (gDNA) was isolated using a modified version of the phenol-chloroform method previously described (Wren and Tabaqchali, 1987). 1.5 ml of O/N culture was harvested via centrifugation (2 min, 12,000 x g) and resuspended in 200 µl PBS supplemented with 10 µl of the purified bacteriophage endolysin CD27L catalytic domain (Mayer et al., 2008), followed by incubation at 37°C for 1 h. 10 µl pronase (final concentration 1 mg/ml)

was added to the sample and incubated at 55°C for 1 h. 80 µl of N- lauroylsarcosine (final concentration 2%) was then added to the sample and incubated at 37°C for 1 h. Finally, 200 µl of an RNase solution (final concentration 0.2 mg/ml) was added to the sample and incubated at 37°C for 1 h. 500 µl of Phenol:chloroform:isoamyl alcohol (25:24:1) was added to the sample and gently mixed. The sample was centrifuged in a phase lock gel tube (13,000 x g, 2 min) to allow separation of the organic phase and DNA containing aqueous phase. The aqueous phase was recovered and treated with Phenol:chloroform:isoamyl alcohol as before. Excess phenol was removed from the recovered aqueous phase by treatment with 500 µl chloroform:isoamyl alcohol (24:1) and centrifugation in a gel lock tube as before. The resulting aqueous phase was then added to an equal volume of ice-cold isopropanol and incubated at -20°C O/N to precipitate the gDNA. gDNA was harvested by centrifugation (4,000 x g, 15 min) and washed once in 70% ethanol. The gDNA was then harvested again (4,000 x g, 10 min) and air dried to remove residual ethanol. gDNA was resuspended in 50 µl nuclease free water and quantified using a Nanodrop spectrometer. Purity was determined through analysis of A_{260}/A_{280} and A_{260}/A_{230} ratios.

Low-quality gDNA preparations for use in PCR were also prepared using Chelex 100 resin (Sigma). A single colony of *C. difficile* was resuspended in 100 µl of nuclease free water containing a small amount of Chelex resin, before being boiled for 10 min and briefly centrifuged. The supernatant was used as a template for PCR. These preparations were made fresh before each PCR reaction.

2.2.2 Polymerase Chain Reaction

2x Phusion polymerase high fidelity master mix (Thermo) was used routinely as per the manufacturer's instructions, for reactions that required high fidelity polymerisation as well as for most PCRs using *C. difficile* chelex preparations. 20 µl reactions contained either 2 µl of a *C. difficile* chelex prep, 200 ng of purified gDNA or 1 ng of plasmid DNA, as the template. 5% (v/v) DMSO was also added to the reaction when required, i.e. when using primers with long over-lapping regions or A/T rich stretches. Initial denaturation was carried out at 98°C for 30 sec

followed by 32 cycles of denaturation at 98°C for 30 sec, annealing at 2°C below the annealing temperature of the primers for 30 sec, and extension at 72°C for 20 sec per kb for DNA fragments up to 3 kb in length or 30 sec for DNA fragments longer than 3 kb. A final extension was performed at 72°C for 5 min.

Taq polymerase was used for PCRs where high fidelity was not required, including colony PCRs. A 2x master mix, prepared in house, was routinely used. The 2x master mix contained 0.01% cresol red, 0.5 M sucrose, 400 μM of each dNTP, purified *Taq* polymerase (estimated to be 0.1 U/μl), 20 mM tris-HCl pH9, 100 mM KCl, 0.02% gelatin, 0.02% tween 20, 4 mM MgCl$_2$. Initial denaturation was carried out at 95°C for 3 min followed by 35 cycles of denaturation at 95°C for 30 sec, annealing at 2°C below the annealing temperature of the primers for 30 sec, and extension at 72°C for 1 min per kb. A final extension was performed at 72°C for 5 min.

PCR reactions were performed using a T100 Thermo Cycler (BIORAD). Primers were synthesised by Eurofins and are listed in table A2.

2.2.3 Agarose Gel Electrophoresis

0.8-2% (w/v) agarose (VWR) was melted in TAE buffer (40 mM tris-acetate pH8, 1 mM EDTA). The percentage of agarose used was dependent on the size of the DNA fragments to be resolved, as higher percentage gels are better able to resolve small DNA fragments, while lower percentage gels are better able to resolve large fragments. The agarose was cooled to approximately 55°C and cast in a horizontal Perspex tray with the addition of SybrSafe DNA stain (1/10,000 final volume) (Invitrogen). When the gel had set, trays were submerged in TAE buffer in an electrophoresis gel tank. DNA loading buffer was added to the DNA samples, either standard DNA loading buffer (New England Biolabs) or UView (Bio-rad) for gel extraction, and samples were loaded into the gel wells. Electrophoresis was performed at 130 V for 15-30 min. DNA bands were visualised using a BioRad ChemiDoc MP imaging system (Bio-Rad).

2.2.4 Purification of PCR Products

Purification of PCR products was accomplished using either a GeneJET PCR Purification kit (Thermo) per the manufactures guidelines, or via gel extraction (see below). DNA was eluted in 20 µl nuclease free water and quantified via A_{260} spectrometry.

2.2.5 Gel Extraction of DNA

DNA samples were resolved on agarose gels (see above). DNA was visualised using a UV transilluminator and bands were cut out using a scalpel. DNA was extracted from the gel using the GeneJET Gel Extraction kit (Thermo), as per the manufacturer's instructions. DNA was eluted in nuclease free water and quantified via A_{260} spectrometry.

2.2.6 Restriction Endonuclease Digestion of DNA

Endonuclease restriction of DNA was performed using restriction enzymes as per the manufacturer's guidelines (New England Biolabs). Restriction endonuclease and its appropriate buffer was added to solutions containing purified DNA and incubated at 37°C for 1 h. For removal of template plasmid DNA from completed Phusion PCR reactions, 2 µl of *Dpn*I was added to the Phusion PCR reaction mix, without prior DNA purification, and incubated at 37°C for 2 h. Digested DNA was purified as described above.

2.2.7 Asymmetric Endonuclease Digestion Assay

Amplification of fragments used in asymmetric digest analysis was performed using 2x *Taq* PCR master mix (Thermo), as per the manufacturer's instructions, from purified gDNA. 10 µl of each PCR reaction was combined with equal volumes of solutions containing 1x concentrations of the endonuclease and its appropriate buffer. Digestion of DNA was then performed at 37°C for 1-2 h using a T100 thermocycler (Bio-Rad). DNA loading buffer was added to a final concentration of 1x before the DNA was resolved using 2% agarose gel electrophoresis. DNA was visualised using a BioRad ChemiDoc MP imaging system. Image Lab software (Bio-Rad)

was used to perform densitometric analysis on the resulting image to quantify the relative amounts of DNA in each band.

2.2.8 Ligation of DNA

DNA fragments were ligated using T4 DNA ligase (New England Biolabs) as per the manufacturer's instructions. Briefly, enzyme reactions contained 1x T4 ligase buffer, 25 ng vector DNA, insert DNA at a molar ratio of 1:3 vector to insert, 0.5 µl T4 ligase, and nuclease free water to make a 10 µl final volume. Reactions were incubated at RT for 1 h before being used in *E. coli* transformations (see below). Linearised vectors created by inverse PCR were first phosphorylated using T4 polynucleotide kinase (New England biolabs) (37°C, 30 min), as per the manufacturer's guidelines, and PCR purified prior to ligation.

2.2.9 Splicing by Overlap Extension

To combine two PCRs fragments, splicing by overlap extension (SOEing) was performed (Horton et al., 1989). Briefly, PCR fragments to be spliced were amplified using primers containing 30 bp overlapping regions, at the ends to be joined. The resulting fragments were PCR purified and 10-20 ng of each fragment were used as DNA templates in a 20 µl Phusion PCR reaction. The forward primer of the 5' fragment and the reverse primer of the 3' fragment were included in the reaction at a final concentration of 0.5 µM. 5% (v/v) DMSO was added when required. The reaction was then treated as a standard Phusion PCR reaction and the combined fragment was purified by gel extraction (see above).

2.2.10 Gibson Assembly of DNA Fragments

Multiple fragments of DNA were ligated in one step through the use of Gibson assembly (Gibson et al., 2009). This was achieved using the NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs). Vector DNA was linearised using Phusion PCR. Insert fragments were amplified using primers so that the resulting fragments contained 30 bp of overlapping sequence with the adjacent DNA fragment. PCR fragments gel extracted as previously described. Reactions were

41

performed in a final volume of 10 µl, containing either 25 ng vector and a two-fold molar excess of insert (for fewer than three fragments), or approximately 200 ng of vector DNA and equimolar insert (more than three fragments). HiFi assembly master mix was added to a final concentration of 1x and the reactions were made to volume using nuclease free water. Reactions were then incubated at 50°C for 0.5 -1 h before being used in *E. coli* transformations.

2.1.11 Production of Chemically Competent *E. coli*

O/N cultures of *E. coli* were sub-cultured 1:100 and grown to log-phase (OD_{600nm} 0.4-06) before being harvested by centrifugation (10 min, 4,000 x g at 4°C). Pellets were resuspended in 5 ml ice cold 100 mM $CaCl_2$ and incubated on ice for 15 min. Cells were harvested as before and resuspended in 1 ml 100 mM $CaCl_2$ with 15% (V/V) glycerol before being incubated on ice for 2 h. The cell suspension was then aliquoted in 50 µl samples and stored at -80°C.

2.2.12 Transformation of *E. coli*

2 µl of ligation or HiFi assembly reaction, or approximately 0.25 µl of a plasmid miniprep, was added to approximately 25 µl of chemically competent *E. coli* and incubated on ice for 30 min. Cells were then heat-shocked at 42°C for 30 sec and incubated on ice for a further 5 min. 475 µl of SOC medium was then and the mixture was incubated at 37°C for 1 h before being spread on LB agar containing the appropriate antibiotic(s).

2.2.13 Isolation of Plasmid DNA

5 ml of O/N *E. coli* culture was harvested by centrifugation (4,000 x g, 10 min) and plasmid DNA was extracted using the GeneJET Plasmid Miniprep kit (Thermo) as per the manufacturer's instructions. DNA was eluted in 60 µl nuclease free water. Plasmids used in this project are outline in table A3.

2.2.14 Sequencing of DNA

Routine sequencing of plasmid and PCR-amplified DNA was carried out using the GATC Biotech (Konstanz, Germany) SUPREMERUN or LIGHTRUN Sanger sequencing services. Geneious

sequence analysis software (Biomatters) was used for the analysis of sequencing data. Illumina sequencing of bacterial genomes was performed by MicrobesNG (Birmingham, UK) or by the Welcome Trust Sanger Institute (Hinxton, UK). Genome assembly and analysis was performed by Dr Roy Chaudhuri (University of Sheffield, UK). PacBio sequencing of bacterial genomes was performed by Genome Québec (Quebec, Canada). Data analysis and genome assembly was performed by Nadia Fernandes (University of Sheffield).

2.2.15 Conjugative Transfer of Plasmid DNA to *C. difficile*

The protocol for conjugative plasmid transfer into *C. difficile* was optimised as part of this PhD studentship and the results published in the journal Anaerobe (appendix IV) (Kirk and Fagan, 2016). Briefly, 200 μl of *C. difficile* O/N culture was heat-treated (50-52°C, 5-15 min) and used to resuspend a pellet of plasmid bearing *E. coli* strain CA434 obtained by centrifugation (4,000 x g, 2 min) of 1 ml O/N culture. The resulting cell suspension was spotted onto BHI agar and incubated for 8-24 h at which point the growth was harvested using approximately 900 μl TY broth and a spreader. The harvested growth was then spread on BHI containing cycloserine, to prevent *E. coli* growth, and thiamphenicol for plasmid selection. To prevent *E. coli* outgrowth, transconjugants were restreaked onto selective BHI containing cycloserine twice before further use.

2.2.16 ClosTron Mutagenesis

Due to the availability of more precise tools for the manipulation of the *C. difficile* genome, ClosTron insertional mutagenesis (Heap et al., 2007) was only used to inactivate a single gene in R20291, FM2.5 and FM2.5RW to replicate the published method of *recV* inactivation described in (Reynolds et al., 2011). pASF075, carrying group II intron targeted to *recV,* was transferred into *C. difficile* via conjugation as previously described. Transconjugants were restreaked on BHI containing lincomycin (5 μg/ml) to select for successful insertion based

on the activity of the retrotransposition-activated marker, *ermB*. Successful mutants were confirmed through PCR using primers flanking the intron insertion site.

2.2.16 *C. difficile* Allele Exchange Mutagenesis

Precise manipulation of the *C. difficile* genome was achieved through allele exchange using the *codA* heterologous counterselection system described previously (Cartman et al., 2012). 600-1200 bp fragments, upstream and downstream of the sequence to be altered, were amplified by PCR and purified. PCR products were ligated into pMTLSC7215 or pMTLSC7315 using either restriction/ligation or a combination of SOEing PCR and Gibson assembly. The resulting plasmids were transferred into *C. difficile* strains R20291 or 630, respectively, via conjugation (see above). pMTLSC7215 and pMTLSC7315 contain sub-optimal origins of replication and thus, under antibiotic selection, the growth rate of plasmid bearing *C. difficile* is limited by the rate at which these pseudosuicide vectors can be replicated. The first recombination event was identified by the appearance of large colonies on selective agar, as the growth rate of single-crossover integrant is no longer limited by the rate of plasmid replication. Single recombination was confirmed by PCR and colonies restreaked on non-selective BHI agar and incubated for 72 h to facilitate the second recombination event and plasmid loss. The bacterial growth was then harvested from the BHI plates using 1 ml of sterile, pre-reduced PBS. The resulting cell suspension was then serially diluted in PBS and plated on CDMM containing the counter-selection agent, 5-fluorocytosine. Colonies that arose after 48-72 h were screened by PCR using primers flanking the altered site. The resulting fragment was then sequenced.

2.3 Protein Purification

2.3.1 Recombinant Protein Expression

C. difficile strain 630 bearing the appropriate plasmid was grown O/N and used to inoculate 10 l TYG to OD_{600nm} 0.1. The culture was grown to OD_{600nm} 0.5 and induced with 700

ng/ml of Atc. Cultures were then incubated for approximately 5 h and harvested via centrifugation (4,000 x g, 10 min) and cell pellets stored at -20°C.

2.3.2 Purification of Recombinant Proteins

Proteins purified during this project were all designed to contain a *strep*-tag at the N-terminus to facilitate a one-step purification process (Schmidt and Skerra, 2007). Pelleted cells were resuspended in 25 ml of 2x binding buffer (200 mM tris-HCl pH8, 300 mM NaCl, 2 mM EDTA) containing approximately 1 ml of purified the bacteriophage endolysin CD27L catalytic domain and cOmplete protease inhibitor cocktail (Roche). The resulting mixture was incubated at 37°C for 30 min followed by centrifugation at 20,000 x g for 3 min to remove cell debris. The cell lysate was then filtered through a 0.44 µm filter, diluted to 50 ml using 1x binding buffer and loaded onto 2x strep-tactin columns (GE healthcare) using an ÄKTAprime (GE healthcare). The columns were then washed using 10 column volumes of 1x binding buffer and proteins eluted by addition of an elution buffer (binding buffer containing 2.5 mM desthiobiotin) gradient. Protein content of fractions was analysed by SDS-PAGE (section 2.5.2) and the fractions containing the desired protein were pooled.

2.3.3 Dialysis and Concentration of Purified Proteins

Purified proteins were dialysed using 10 MWCO Slide-A-Lyzer dialysis cassettes (Thermo). Cassettes were submerged in dialysis buffer (25 mM tris-HCl pH8, 50 mM NaCl) for 2 min to wet the membrane. 3 ml of protein solution was then injected in between the membranes of the cassette. The cassettes were then submerged in 1 l of dialysis buffer and incubated at 4°C with gentle stirring, initially for 2 h, then for 3-4 hours and finally O/N. In-between incubations the buffer was exchanged for fresh dialysis buffer. The following morning the protein was removed from the cassette and concentrated using a 10 MWCO Amicon ultra-4 centrifugal filter. Protein concentrations were quantified using A_{280} spectrometry.

2.3.4 Crude S-layer Extraction

S-layer extracts were prepared using the low pH glycine method as described previously (Calabi et al., 2001). 5-20 ml of *C. difficile* cultures were harvested via centrifugation (4,000 x g, 10 min). Pelleted cells were resuspended to an approximate OD_{600nm} of 50 in 0.2 M glycine-HCl pH 2.2 and incubated with rotation at RT for 15 min. The suspensions were centrifuged (20,000 x g, 2 min) and the supernatants containing the S-layer proteins were recovered and neutralised using 2 M tris.

2.5 Protein Characterisation

2.5.1 Protein Crystallisation and 3D Modelling

Solutions of purified SecA2 were concentrated to approximately 1 mg/ml. Crystal screening was initially carried out using PACT, Cryo, Screen 1, and Morpheus suites (Qiagen) using the sitting drop vapour diffusion method with a 200 nl drop size (1:1 protein/precipitant ratio). After 7 days, the plates were screened for crystal growth. Optimisation of crystal growth was performed by Dr Paula Salgado (University of Newcastle) using a range of concentrations of $MgCl_2$ (100-250 mM) and PEG6K (15-30%) buffered with either 100 mM MES pH 6.0-6.5 or 100 mM HEPES pH 7.0-7.5 using a hanging drop vapour diffusion method. Diffraction data was obtained from crystals using the I24 microfocus X-ray source (Diamond light source, Oxford). Data was processed and analysed entirely by Dr Paula Salgado.

2.5.2 SDS-PAGE

Electrophoretic analysis of proteins was carried out using denaturing polyacrylamide gels (Laemmli, 1970). 30% acrylamide:bisacrylamide (37:5:1) gels were prepared according to standard protocols. The final acrylamide concentration used was dependent on the size of the proteins to be resolved, i.e. lower percentages were used for larger proteins. Samples were

combined with an equal volume of 2x Laemmli loading buffer (150 mM Tris-HCl, pH 6.8, 30% (v/v) glycerol, 1.5% (w/v) SDS, 15% (v/v) β-mercaptoethanol, 2 µg/ml bromophenol blue) and electrophoresis was performed using Mini-Protean tetra cell apparatus (Bio-rad). Electrophoresis was performed using a constant voltage (190 V), until the dye front approached the bottom of the gel. When required, gels were stained O/N in 45% (v/v) methanol, 10% (v/v) acetic acid, 0.25% (w/v) brilliant blue R-250 and destained using 45% (v/v) methanol, 10% (v/v) acetic acid.

2.5.3 In-gel Fluorescence

C. difficile strains expressing SNAP protein fusions were grown in liquid broth and SNAP-TMR star added to a final concentration of 250 nM. Bacteria were then harvested via centrifugation (20,000 x g, 2 min) and pellets were resuspended and lysed. Lysates were combined 1:1 with 2x Laemmli loading buffer and conventional SDS-PAGE performed. Proteins bands were visualised using a BioRad ChemiDoc MP imaging system.

2.5.4 Transfer to PVDF Membrane

Following SDS-PAGE, proteins were transferred to PVDF membrane in one of two ways. For relatively small proteins (<100 kDa) the semi-dry transfer method was employed. Gels were equilibrated in Cation buffer (40 mM glycine, 25 mM tris-HCl, 10% (v/v) methanol, ph 9.4) for 10 min. During this time, PVDF membranes were briefly soaked in methanol, submerged in water for 2 min and equilibrated in Anode buffer II (25 mM tris-HCl, 10% (v/v) methanol, pH 9.4) for 8 min. Transfer stacks were then assembled, anode to cathode, as follows. Two pieces of 3 MM Whatman filter paper soaked in Anode buffer I (300 mM tris-HCl, 10% (v/v) methanol, pH 10.4), one piece of filter paper soaked in Anode buffer II, the Anode II buffer equilibrated membrane, the Cathode buffer equilibrated gel, and finally three pieces of filter paper soaked in Cathode buffer. Transfer was carried out using a TransBlot Turbo transfer system (Bio-Rad), at a constant voltage (15 V) for 15 min.

When analysing large proteins (>100 kDa) the wet transfer method was employed. PVDF membranes were soaked briefly in methanol, submerged in water for 2 min and equilibrated in transfer buffer (25 mM tris, 192mM glycine, 10% (v/v) methanol). Gels were also equilibrated in transfer buffer. The membrane and gel were sandwiched between three transfer buffer soaked pieces of filter paper on each side. Transfer stacks were then incorporated into Mini Trans-Blot Modules (Bio-rad) and submerged in ice-cold transfer buffer in a Mini-Protean Tetra Cell (Bio-Rad). Transfer was carried out at a constant voltage (100 V) for two hours.

To determine the quality of protein transfer, membranes were covered with Ponceau red stain (0.1% (w/v) Ponceau S, in 5% acetic acid) and washed using distilled water until protein bands could be easily visualised. Membranes were then dried by submerging in methanol followed by incubation at room temperature for a minimum of 30 min.

2.5.5 Western Blot Analysis

Dried membranes were incubated with primary antibody in a 3% skimmed milk PBS solution, either O/N at 4°C or at RT for 1 h, with gentle rocking. Membranes were washed three times with PBS and incubated with horseradish peroxidase (HRP) conjugated secondary antibodies in 3% skimmed milk at RT for 1 h. Membranes were washed an additional four times with PBS and developed using Clarity Western ECL Blotting Substrate (Bio-Rad). Blots were imaged using the BioRad ChemiDoc MP imaging system.

2.5.6 Preparation of Samples for Peptide Mass Fingerprinting

Protein samples were first reduced, alkylated and trypsinized as follows. 5 µl of protein solution were added to 35 µl digestion buffer (final concentration 100 mM NH_4HCO_3). Samples were reduced through the addition of 2 µl DTT (final concentration 10 mM) and incubation for 1 h at 56°C. Samples were cooled to RT and alkylated by adding 2 µl iodoacetamide (final concentration 10 mM). Samples were vortexed and incubated in the dark at RT for 30 min. 1 µl of 500 mM DTT was then added to quench the alkylation reaction. Trypsin was then added to a

final ratio of 25:1 protein to protease, and digestion carried out at 37°C O/N. The following morning samples were desalted using C18 spin columns (Thermo), eluted in 40 μl of 70% acetonitrile and dried in a vacuum evaporator.

2.5.7 Peptide Mass Fingerprinting

Peptide mass fingerprinting was carried out by Dr Richard Beniston at the Biological Mass Spectrometry Service (Department of Chemistry, University of Sheffield). Peptide mass was determined via LC-MS/MS using a Thermo Orbitrap Elite. Mass spectrometry experiments were performed using biological duplicates. Proteins were analysed using two different concentrations and once without prior reduction/alkylation.

2.6 Microscopy

2.6.1 Fixing Cells With Paraformaldehyde

1 ml of sample was harvested (20,000 x g, 2 min) and washed three times with PBS. Cells were then resuspended in 3.2% paraformaldehyde in PBS and incubated on a rotary shaker for 15 min at RT. Samples were then washed twice in PBS. The cross-linker was then quenched by the addition of 0.5 ml 20 mM NH_4Cl and incubated for 15 min at RT. Samples were washed a further two times in PBS and stored as pellets at 4°C.

2.6.2 Phase Contrast and Epifluorescence Microscopy

For phase contrast microscopy, fixed cells were resuspended in sterile water to an appropriate cell density and dried onto glass slides. Slides were rinsed using distilled water and dried. Cover slips were mounted in 80% glycerol.

Samples for use in SNAP-tag fluorescence microscopy were stained with SNAP TMR-Star (New England Biolabs) prior to being fixed. SNAP TMR-Star (final concentration 250 nM) was

added to 1 ml samples of growing *C. difficile* cultures and incubated at 37°C anaerobically for 30 min, washed twice using 0.5 ml sterile PBS, and fixed and mounted as described above.

For HADA labelling of the cell wall, O/N cultures of *C. difficile* were used to inoculate fresh broth to OD_{600nm} 0.05. 1.5 h post-inoculation, HADA (final concentration 0.6 mM) was added to the growing culture. When the cultures reached OD_{600nm} 0.6, 1 ml samples were harvested by centrifugation (2,000 x g) and washed 3x with sterile, pre-reduced TY broth. Bacterial pellets were then resuspended in TY broth and incubated anaerobically for a further 30 min. Samples were then processed as above.

For immunofluorescence, samples were fixed and dried onto glass slides as above. Slides were blocked O/N in 3% BSA in PBS at 4°C. Slides were then drained and rinsed in PBS. 15 µl of primary antibody in 3% BSA in PBS was spotted onto the dried cells and incubated for 1 h at RT. Slides were rinsed in PBS before 15 µl secondary antibody in 3% BSA in PBS was spotted onto the slide. Slides were incubated for a further 1 h at RT. Slides were thoroughly rinsed in PBS and sterile water, dried and mounted using 5 µl *SlowFade* gold (Thermo). All samples were visualised using a Nikon Ti Eclipse inverted microscope.

2.6.3 Electron Microscopy

TEM analysis was performed by Christopher Hill at the electron microscopy unit of the University of Sheffield. After fixation with paraformaldehyde, cells were further fixed with 3% glutaraldehyde in 0.1 M cacodylate buffer followed by 1% osmium tetroxide. Samples were dehydrated in ethanol, embedded in araldite resin and sectioned at 85 nm on a Leica UC6 ultramicrotome. Sections were then transferred onto copper coated grids and stained with uranyl acetate and lead citrate. Samples were visualised using a FEI Technai BioTWIN TEM at 80 kV fitted with a Gatan MS600CW camera.

2.7 Phenotypic Assays

2.7.1 Sporulation Efficiency Determination

Sporulation assays were performed as previously described (Dembek et al., 2015). Single colonies of *C. difficile* were used to inoculate pre-reduced BHIS media. The following morning, 5 ml of fresh BHIS media was inoculated to OD_{600nm} 0.01 and incubated for 8 h. This culture was then used to inoculate 10 ml of fresh BHIS media to OD_{600nm} 0.0001. This was performed to ensure minimal carry over of spores from the original culture. For total CFU counts, the following morning (T_0) samples were serially diluted in PBS and 10 µl of the dilutions were spotted in triplicate on BHIS plates containing 0.1% (w/v) sodium taurocholate. For total spore counts, 0.5 ml samples of the culture were heated to 65°C for 30 min. This treatment kills all vegetative cells in the sample. Heated samples were serially diluted and plated on sodium taurocholate containing BHIS plates. CFU enumeration was performed the following day. This process was repeated every day for a further 5 days (T_1-T_5). Samples were also taken, fixed, mounted, and visualised using phase contrast and electron microscopy.

2.7.2 Spore Thermal Resistance

Sporulation assays were prepared as above. At T_5 total counts and spore counts were performed as above. In addition, 0.5 ml samples were also heated to 75°C for 30 min. CFU enumeration was performed as above.

2.7.3 Germination Efficiency

Germination assays were performed as previously described (Dembek et al., 2015). Purified spores were resuspended in BHIS containing 0.5% sodium taurocholate to OD_{600nm} 1.0. Ca^{2+}-DPA release and core-rehydration was then monitored over time by measuring the A_{600nm}.

2.7.4 *C. difficile* Growth Analysis

O/N cultures of *C. difficile* grown in TY broth were used to inoculate fresh, pre-reduced TY to OD_{600nm} 0.05. Growth was then measured by recording the OD_{600nm} every hour. In addition, 200 µl samples were serially diluted and spotted in triplicate on BHI agar. CFU enumeration was performed after O/N incubation.

2.7.5 Lysozyme and LL37 resistance

Cultures for lysozyme and LL-37 challenge experiments were prepared as described above (section 2.7.4). Lysozyme (500 µg/ml) and LL-37 (5 µg/ml) were added 2.5 h post inoculation and the OD_{600nm} was recorded hourly.

2.7.6 Avidocin-CD Killing

Broth based Avidocin-CD killing assays were performed as above (section 2.7.5) in TY supplemented with 1 mM $CaCl_2$. Avidocin-CDs were added 2.5 h post inoculation and the OD_{600nm} was then measured hourly. Avidocin-CDs were added in excess to ensure complete killing.

2.7.7 Comparative Toxin Production and Toxin Release Assays

For comparative analysis of toxin production, O/N cultures of *C. difficile* were used to inoculate fresh, pre-reduced TY to an OD_{600nm} of 0.01. Cultures were incubated for 8 h and sub-cultured in TY broth to OD_{600nm} 0.0001. After O/N incubation, the culture was subcultured to OD_{600nm} 0.05. 1 ml samples were harvested (20,000 x g, 2 min) after 5, 24, 48 and 72 h. Pelleted cells were washed once in 0.5 ml PBS and stored at -20°C until required. Supernatants were passed through a 0.44 µm filter and stored at -20°C. Before analysis, supernatants were thawed and concentrated to a relative OD_{600nm} of 20. Pelleted cells were resuspended to OD_{600nm} 20 in PBS containing the CD27L bacteriophage endolysin, followed by incubation at 37°C for 30 min.

To determine if Avidocin-CD killing causes the release of intracellular toxin, Av-CD291.2 challenge experiments were performed. This was done using cultures incubated for 24 h, as this

was determined to the time-point where cells contain the highest concentration of intracellular toxin. After 24 h incubation, 1 ml samples were centrifuged in the anaerobic work station (2,000 x g, 3 min) and washed once with 0.5 ml pre-reduced TY media. Pelleted bacteria were then resuspended to OD_{600nm} 20 in TY supplemented with 10 mM $CaCl_2$. Av-CD291.2 was added in equal volumes to samples at different ratios of bacterial cells: Avidocin-CD. Av-CD684.1 was added (1:500 cells to Avidocin-CD) to one sample and one sample was left untreated to act as negative controls. Samples were incubated anaerobically at 37°C for 1 h. Serial dilutions of samples were then spotted in triplicate on BHI agar containing taurocholate. A sample of the culture was also heated to 65°C and spore counts calculated as previously described. CFU was enumerated after 24 h to enable an accurate amount of killing by the Avidocin-CD. Treated samples were harvested by centrifugation (13,000 x g, 3 min). Supernatants were recovered and centrifuged again (16,000 x g, 3 min) to prevent contamination by bacterial cells.

The cell lysates and supernatant samples were combined 1:1 with 2x Laemmli loading buffer and loaded on 6% SDS-PAGE gels ready for western blot analysis using an anti-toxin B antibody (Thermo Fisher).

2.8.8 β- Glucuronidase Quantification

These assays were performed as previously described (Dupuy and Sonenshein, 1998). *C. difficile* cultures were grown to exponential phase (approximately OD_{600nm} 0.7). 5 ml samples were harvested (4,000 x g, 10 min) and pelleted cells were stored at -20°C. Each 5 ml sample was resuspended in 100 µl Z-buffer (60 mM Na_2HPO_4, 60 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM 2-mercaptoethanol, pH 7.0) containing CD27L bacteriophage endolysin. Samples were incubated at 37°C for 30 min and centrifuged at 20,000 x g for 10 minutes. 700 µl Z-buffer was added to each sample and incubated at 37°C for 5 min. Enzyme reactions were initiated by the addition of 160 µl of 6 mM p-nitrophenyl-β-D-glucuronide. The start time of each reaction was recorded. After the development of a yellow colour to the solutions, reactions were stopped

by the addition 0.4 ml 1 M Na_2CO_3. Samples were then centrifuged (20,000 x g, 10 min) to remove cell debris and the OD_{405nm} recorded. Enzyme activity was calculated using the formula $(OD_{405nm}$ x 1000$)/(OD_{600nm}$ x reaction time (min) x 1.25 x vol of culture (ml)).

2.8.9 Bacterial Two-Hybrid

BACTH reporter *E. coli* strains DHM1 and BTH101 were co-transformed with plasmids expressing bait and prey proteins fused to the T25 and T18 fragments of the catalytic domain of adenylate cyclase (CyaA). Transformants were grown O/N in LB broth supplemented with the appropriate antibiotics. 10 µl of O/N culture was spotted onto LB agar containing 0.5 mM IPTG and 40 µg/ml X-gal. Plates were incubated for up to 48 h at either 37°C or 30°C, and monitored for the development of a blue colour to the *E. coli* growth.

2.8 Bioinformatics

Geneious sequence analysis software 10.0.9 (Biomatters) was used routinely for analysis of protein and nucleotide sequences, including sequence alignment, *in silico* cloning, and experimental design. Bacterial promoter prediction was performed using SoftBerry BPROM (www.softberry.com). Primers to amplify fragments used in either SOEing or Gibson assembly were designed using NEBuilder (http://nebuilder.neb.com/). Calculation of protein parameters including the estimated extinction coefficient was performed using the ExPASy protParam tool (Wilkins et al., 1999). Analysis of the inverted repeat sequences conservation was performed by aligning the left and right inverted repeat of each genetic switch. A 21 bp region of high consensus identified and used to create a logo-plot. Logo plots were created using the WebLogo online application (Crooks et al., 2004).

2.9 Statistical Analysis

Statistical analysis of data was performed using a combination of one way ANOVA, two tailed t-tests using Welch's correction, and Mann-Whitney U tests. Statistical tests were performed using GraphPad Prism 7 (GraphPad Software Inc).

Chapter III

New Class of Precision Antimicrobials Redefines Role of *Clostridium difficile* S-layer in Virulence and Viability

As part of my PhD thesis, I am including my published paper "New Class of Precision Antimicrobials Redefines Role of *Clostridium difficile* S-layer in Virulence and Viability", which was originally published in Science Translational Medicine, DOI: 10.1126/scitranslmed.aah6813.

The work contained in this paper was performed in collaboration with Dr Gillian Douce (University of Glasgow) and our industrial partner AvidBiotics Corp. (South San Francisco, USA). Approximately 75% of the experimental work described in the paper was carried out by myself. This includes all *C. difficile* molecular biology and microscopy, as well as the Western blotting and half of the *in vitro* testing. Work that was performed by others includes experiments using the hamster model of infection (University of Glasgow) and design, construction, and purification of Avidocin-CDs, the remainder of the *in vitro* testing and isolation of the S-layer mutants (AvidBiotics).

Figures 1A-F, 2A&B, 3A-D, 4E, S1 (A-C), S3 (A-C), S4 (A-D), S5 (A-C), and S6 represent work that I personally performed. Dr Gillian Douce's group was responsible for the data shown in Figure 4A-D. The research team at AvidBiotics, headed by Dr Gregory Govoni, was responsible for the data shown in S2 and S7 in addition to the work mentioned previously.

I drafted the sections of the paper (Methods and Results) describing my work and prepared the figures. The text was subsequently edited by my supervisor, Dr Robert Fagan and the second corresponding author, Dr Gregory Govoni.

Yours sincerely,

Joseph Kirk (PhD candidate)

Dr Robert Fagan

Dr Gregory Govoni

New Class of Precision Antimicrobials Redefines Role of *Clostridium difficile* S-layer in Virulence and Viability

Joseph A. Kirk[1], Dana Gebhart[2], Anthony M. Buckley[3], Stephen Lok[2], Dean Scholl[2], Gillian R. Douce[3], Gregory R. Govoni[2*], Robert P. Fagan[1*]

1: Krebs Institute, Department of Molecular Biology and Biotechnology University of Sheffield, Sheffield S10 2TN, UK.
2: AvidBiotics Corp. South San Francisco, California, USA.
3: Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow G12 8TA, UK.

*To whom correspondence should be addressed: Gregory R. Govoni: ggovons@gmail.com;

Robert P. Fagan: r.fagan@sheffield.ac.uk

Abstract:

Avidocin-CDs are a new class of precision bactericidal agents that do not damage resident gut microbiota and are unlikely to promote the spread of antibiotic resistance. The precision killing properties result from the fusion of bacteriophage receptor binding proteins (RBPs) to a lethal contractile scaffold from an R-type bacteriocin. We recently described the prototypic AvidocinCD, Av-CD291.2, that specifically kills *C. difficile* ribotype 027 strains and prevents colonization of mice. We have since selected two rare Av-CD291.2 resistant mutants of strain R20291 (RT027; S-layer cassette type-4, SCLT-4). These mutants have distinct point mutations in the *slpA* gene that result in an S-layer null phenotype. Reversion of the mutations to wild-type restored normal SLCT-4 S-layer formation and Av-CD291.2 sensitivity; however, complementation with other SCLT alleles did not restore Av-CD291.2 sensitivity despite restoring S-layer formation. Using newly identified phage RBPs, we constructed a panel of new Avidocin-CDs that kill *C. difficile* isolates in an SLCT-dependent manner, confirming the Slayer as the receptor in every case. In addition to bacteriophage adsorption, characterization of the S-

layer null mutant also uncovered important roles for SlpA in sporulation, resistance to lysozyme and LL-37, and toxin production. Surprisingly, the S-layer-null mutant was found to persist in the hamster gut despite its completely attenuated virulence. Avidocin-CDs have significant therapeutic potential for the treatment and prevention of *C. difficile* Infection (CDI) given their exquisite specificity for the pathogen. Furthermore, the emergence of resistance forces mutants to trade virulence for continued viability and, therefore, greatly reduce their potential clinical impact.

3.1 Introduction

New antibacterial agents are needed to counteract the impending loss of effective treatment options for multi-drug resistant bacteria. Furthering this need is the realization that dysbiosis caused by broad-spectrum antibiotic use contributes to the prevalence of diseases/disorders such as inflammatory bowel disease, obesity and gastrointestinal infections (*1*). Strategies to overcome these threats include use of narrow spectrum or precision agents and the design of drugs that target virulence instead of *in vitro* viability (*2, 3*). One pathogen for which alternative treatment approaches are needed is *C. difficile*. This spore-forming, obligate anaerobe is the leading cause of nosocomial infections worldwide. Approximately 450,000 cases and 29,000 deaths each year are attributed to this pathogen in the US alone (*4*). As a result, the Centers for Disease Control and Prevention has identified *C. difficile* as an urgent threat to human health (*5*). This opportunistic pathogen exploits a reduction in gut microbiota diversity that often follows broad spectrum antibiotic use to proliferate, release toxins, and cause life-threatening colitis (*6*). Although the toxins have been studied in great detail, other aspects of *C. difficile* virulence, including colonization of the gut, are not well understood (*6*). The *C. difficile* cell surface is covered by a paracrystalline surface layer (S-layer) largely comprised of SlpA and sparsely interspersed by 28 related cell wall proteins (*7*). The S-layer precursor SlpA is proteolytically processed on the cell surface to generate the Low and High Molecular Weight S-

layer Proteins (LMW and HMW SLPs). The SLPs interact with high-affinity to form a heterodimer, the basic unit of the mature S-layer (*8*). The *slpA* gene is located within a highly variable S-layer cassette consisting of 5 genes; 13 distinct S-layer cassette types (SLCTs) have been described to date (*9*). The variation that defines individual cassette types is largely confined to the LMW SLPencoding region of *slpA* (*4*). The HMW SLP region is highly conserved and includes the cell wall binding motifs that anchor the S-layer to the cell wall (*10*). The S-layer and several associated cell wall proteins have been implicated in colonization of host tissues (*7*) and in stimulation of the host immune response via TLR4 signaling (*11*).

Whereas *C. difficile* is not significantly resistant to the frontline antibiotics used to treat CDIs (vancomycin, metronidazole and fidaxomicin), the use of these antibiotics causes further disruption of the resident microbiota leading to frequent CDI relapse (*6*). To be safe and effective, new agents to treat and prevent CDIs must not harm the diverse gut microbiota and the colonization resistance it provides. Avidocin-CDs represent one such potential agent (*12*). These bactericidal proteins are genetically modified versions of natural R-type bacteriocins (a.k.a. diffocins) produced by *C. difficile* to kill competing *C. difficile* strains (*13*). Diffocins resemble *myoviridae* phage tails and consist of a contractile sheath, nanotube core, baseplate and tail fiber structures. However, instead of delivering DNA across the bacterial membrane as does a bacteriophage, R-type bacteriocins function as killing machines by injecting a nanotube core through the bacterial cell envelope and creating a small pore that dissipates the cell's membrane potential (*14*). Killing specificity is determined by the receptor binding proteins (RBPs) located at the tail fiber tips that trigger sheath contraction upon binding with a cognate receptor on the bacterial cell surface. Genetic replacement or fusion of the RBP gene with homologues from other strains or RBP sources (i.e*. C. difficile* bacteriophages and prophage insertions) make it possible to retarget killing (*12, 15-17*). These modified bacteriocins are known as Avidocin-CDs. An Avidocin-CD prototype, Av-CD291.2, constructed with a bacteriophage RBP identified within a prophage insertion was found to be more stable than the

natural parent diffocin. AvCD291.2 had a modified killing spectrum that included all hypervirulent RT027 *C. difficile* strains tested, blocked *C. difficile* colonization in a mouse model of spore transmission and did not disrupt the resident gut microbiota (*12*). These properties encourage the further development of Avidocin-CDs as oral human therapeutics.

Here we further characterize the Av-CD291.2 mechanism of action and describe an expanded panel of Avidocin-CDs that cover all clinically relevant *C. difficile* strain types using newly identified bacteriophage RBPs. Rare resistant mutants were isolated *in vitro* under Av-CD291.2 selection. Analysis of these mutants enabled the identification of SlpA as the cell surface receptor for all tested Avidocin-CDs and, by extension, the corresponding bacteriophage and prophage RBP-sources. We also identified previously unsuspected roles for SlpA in sporulation, resistance to lysozyme and the natural anti-microbial peptide LL-37, toxin production and virulence. Surprisingly, despite their complete attenuation of virulence, resistant mutants could colonize and persist in the hamster gut for the duration of a 14-day study. These findings imply that antimicrobial agents that force pathogens to trade virulence for continued viability would have clear clinical advantages should resistance emerge during treatment.

3.2 Results

3.2.1 Av-CD291.2-resistant mutants lack an S-layer

Resistance to any antimicrobial agent can occur and be exploited to understand its mode of action. We isolated two spontaneous mutants of *C. difficile* strain R20291 (ribotype 027) that were resistant to killing by Av-CD291.2. These mutants appeared at a frequency of $< 1 \times 10^{-9}$ and were found to encode independent point mutations in the *slpA* gene (Fig.3.1A). Both mutations were predicted to truncate SlpA at a site N-terminal to the post-translational cleavage site and, thereby, prevent formation of an S-layer. Both FM2.5 and FM2.6 lacked detectable cell surface S-layer protein subunits as predicted but still expressed minor cell wall proteins including Cwp2 and Cwp6 (Fig.3.1B). These mutations did not affect the growth rate of the bacteria *in vitro*

(Fig.S4B); however, FM2.5 displayed a slight, but statistically significant earlier entry into stationary phase (maximum OD$_{600nm}$: FM2.5 = 2.2; R20291 = 3.2, *p* =

0.000012). We attempted to complement the FM2.5 and FM2.6 mutations with a plasmid-borne wild-type R20291 *slpA*; however, unexpected homologous recombination restored the wild type *slpA* gene to the chromosome (Fig.3S1). Accordingly, we created genetically identifiable recombinants "watermarked" with synonymous substitutions in the R20291 *slpA* allele (Fig.3.1A). The resulting strains, FM2.5RW and FM2.6RW, were found to have LMW and HMW SLPs on their surface (Fig.3.1B) and regained sensitivity to Av-CD291.2 (Fig.3.1C and D). These results confirmed that an intact S-layer is required for Av-CD291.2 killing but did not determine whether SlpA itself or a protein dependent on the S-layer for surface localization was the receptor for Av-CD291.2. To address this question, *slpA* alleles from the most common nonR20291 SLCTs were individually cloned into an inducible expression plasmid and transferred into the S-layer deficient FM2.5 (Fig.3.1E). Each resulting strain expressed the expected LMW and HMW SLPs on the cell surface, indicative of S-layer formation (*9*), but did not regain sensitivity to Av-CD291.2, thus, ruling out the possibility that the simple formation of an S-layer was responsible for sensitivity to Av-CD291.2. To address whether Av-CD291.2 directly interacts with SLCT-4 SlpA, plasmids encoding SlpA from 8 different SLCTs were introduced into the non-isogenic laboratory strain 630 (SLCT-7; ribotype 012), which is insensitive to AvCD291.2 (Table 3S1). In this fully S-layer competent SLCT-7 wild type strain, only induction of SLCT-4 SlpA was sufficient to confer sensitivity to Av-CD291.2 (Fig.3S3). Moreover, the degree of sensitivity was dependent on the level of induction with only 10 ng/mL of inducer required

(Fig. S3A). Taken together these observations clearly demonstrate that the SLCT-4 variant of SlpA is the cell surface target of Av-CD291.2.
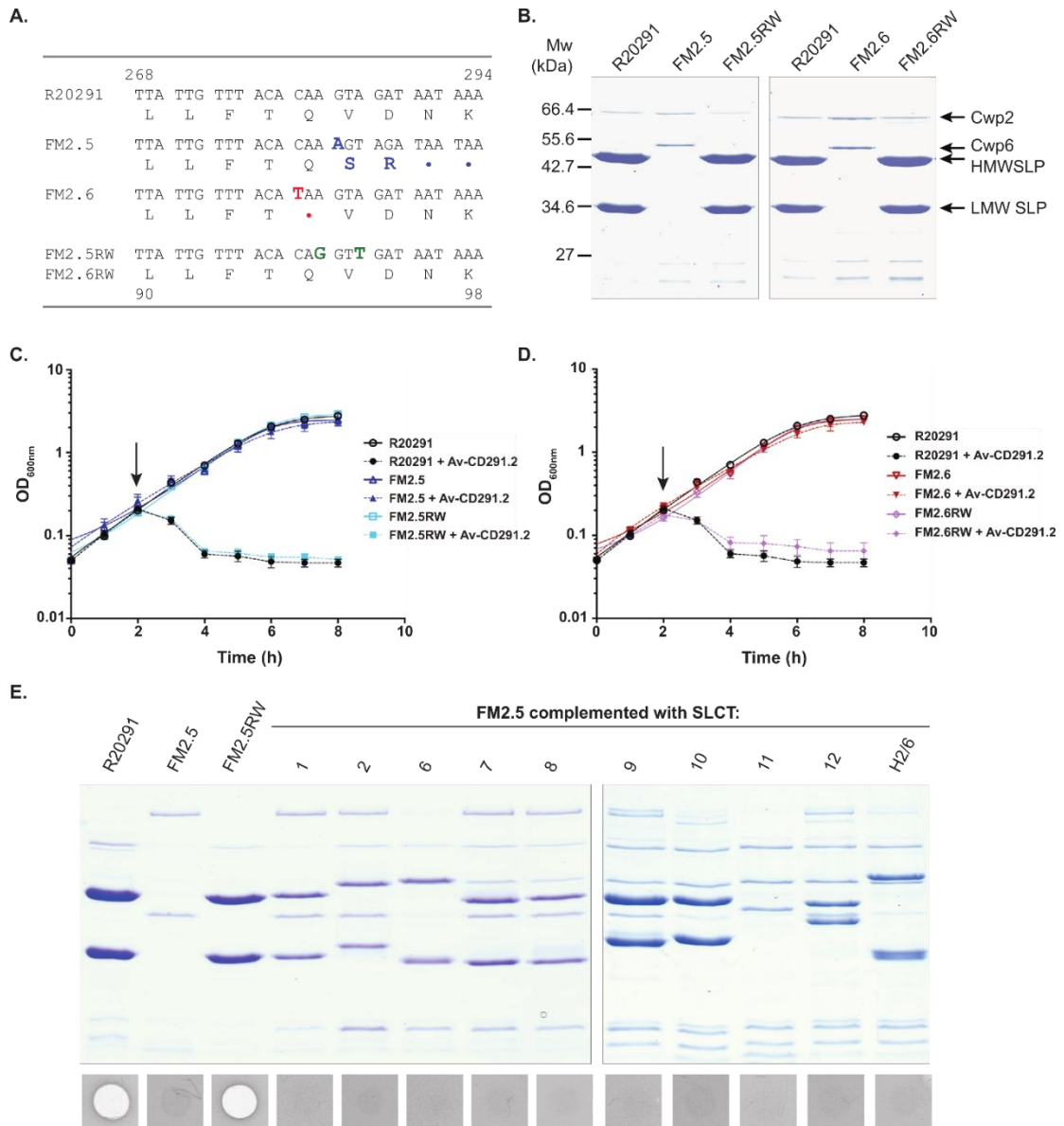
Fig.3.1: Mutations in *slpA* confer Av-CD291.2 resistance. (A) Alignment of the *slpA* sequence (nucleotides 268-294) from R20291, FM2.5, FM2.6, FM2.5RW and FM2.6RW. A nucleotide insertion at position 283 of FM2.5 *slpA* results in a frameshift and premature stop codon (in blue). A nucleotide substitution at position 280 of FM2.6 *slpA* results in a nonsense mutation (in red). To allow differentiation from the wild type sequence, two synonymous mutations were introduced into *slpA* in FM2.5RW and FM2.6RW (in green) **(B)** SDS-PAGE analysis of S-layer extracts from R20291, FM2.5, FM2.6, FM2.5RW and FM2.6RW. The positions of the LMW and HMW SLPs and minor cell wall proteins Cwp2 and Cwp6 are indicated. **(C and D)** The impact of Av-CD291.2 on exponentially growing R20291, FM2.5, FM2.5RW and FM2.6RW was monitored by measuring the optical density at 600 nm. Av-D291.2 addition is indicated with an arrow. Experiments were carried out in triplicate on biological duplicates. Means and standard deviations are shown. **(E)** SDS-PAGE analysis and Av-CD291.2 sensitivity of FM2.5 complemented with *slpA* alleles from multiple SLCTs following induction with anhydrotetracycline (20 ng/ml). R20291 and FM2.5RW are included as controls. A zone of clearance in the agar lawn indicates killing.

3.2.2 Sensitivity to all Avidocin-CDs is slpA allele-specific

These observations suggested that each variant of SlpA may serve as a specific receptor for additional *C. difficile* bacteriophage RBPs. In an attempt to expand Avidocin-CD coverage beyond SLCT-4, we constructed ten new Avidocin-CDs using predicted bacteriophage RBPs mined from the genome sequences of *C. difficile* clinical isolates or newly isolated bacteriophages (Table 3S1). Source strains for RBP sequences were chosen based on their SLCT since strain typing information for the RBP source often correlates with sensitivity to the corresponding Avidocin (*e.g.* ribotype 027 and sensitivity to Av-CD291.2; SLCT-1 of phi-147 propagating strain and sensitivity to Av-CD147.1) (*12, 15, 16*). Preparations of Diffocin-4 (scaffold for all Avidocin-CDs), Av-CD291.2, and each of the new Avidocin-CDs were tested for killing activity on a panel of 62 *C. difficile* isolates containing all 13 known SLCTs and a newly identified 14th SLCT (Fig.3S2). Only two strains, representing SLCTs 3 and 5, were not killed by a single Avidocin-CD (Fig.3S2); otherwise, every isolate from all 12 other SLCTs was killed by at least one Avidocin-CD. A near perfect correlation was observed between a strain's SLCT and its sensitivity to each of the Avidocin-CDs. The only exceptions were SLCT-2 isolates with Av-CD027.2 and Av-CD685.1. A strong correlation was also observed between ribotype and sensitivity to a particular Avidocin-CD since all ribotypes in the panel, except 012, 014 and 015, were found to associate exclusively with a single SLCT. Similar correlations with strain sensitivity and SLCT or ribotype were not observed for killing with Diffocin-4, suggesting this natural R-type bacteriocin binds to *C. difficile* via another receptor.

Having made these observations, we wanted to determine if sensitivity to specific Avidocin-CDs was directly dependent upon the variant of SlpA present. The panel of isogenic FM2.5 strains complemented with *slpA* alleles from the 10 most common SLCTs was tested for sensitivity to a panel of the most potent Avidocin-CDs (Fig.3.2). The corresponding parental strains from which the *slpA* alleles were obtained were used as controls. Each complemented strain became sensitive to the same Avidocin-CDs as the parental strain. To confirm that this

engineered sensitivity did not result from altered cell surface architecture, an analogous experiment was performed in a second panel of isogenic strains created when plasmids encoding SlpAs from 8 SLCTs were introduced into the S-layer competent strain 630 (Fig.S3B and C). As before, the spectrum of killing was identical to that of the parental strains from which the *slpA* alleles were obtained. These data conclusively demonstrate that the polymorphic SlpA acts as the binding receptor for each of the Avidocin-CDs tested and, therefore, the corresponding bacteriophage RBPs.
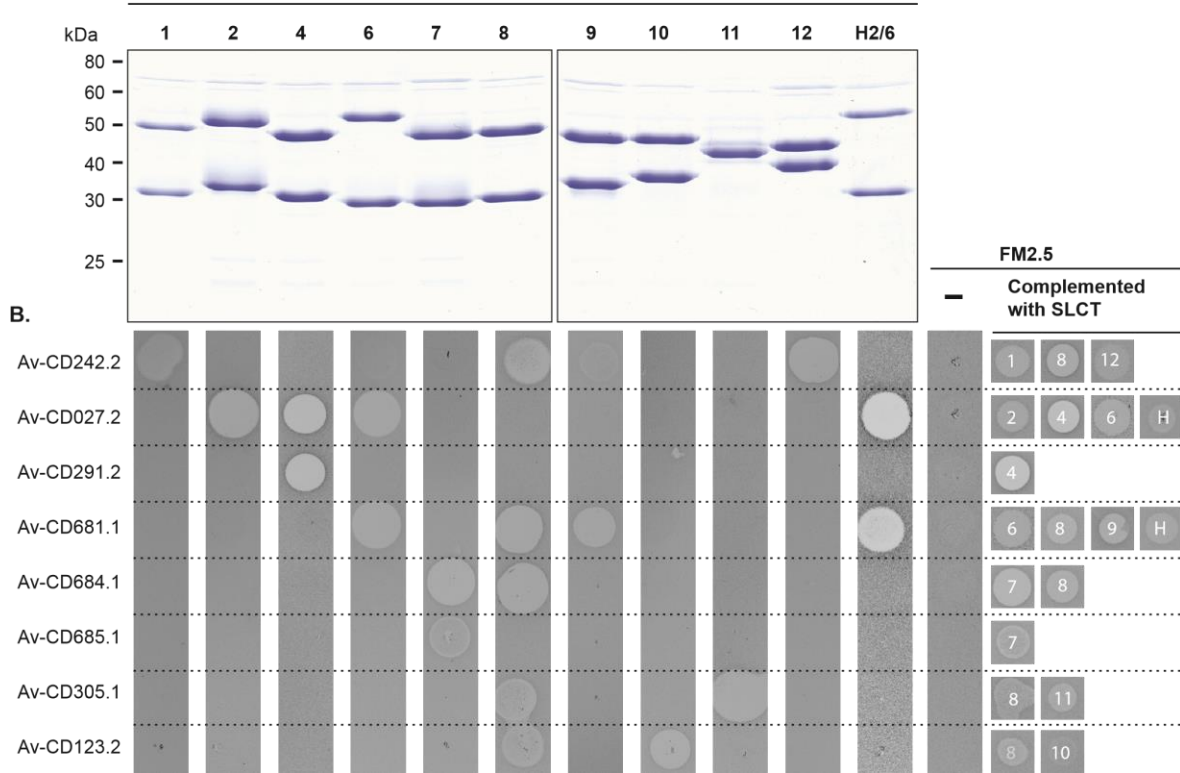
Fig.3.2: Avidocin-CD sensitivity correlates with SLCT. (A) SDS-PAGE analysis of SLPs extracted from a panel of strains representing the 11 most commonly isolated SLCTs. **(B)** Spot bioassays with 8 Avidocin-CDs on the *C. difficile* strains used in panel A, as well as FM2.5 alone (-) and FM2.5 complemented with *slpA* alleles from 10 SLCTs following induction with anhydrotetracycline (20 ng/ml). The zone of clearance caused by each Avidocin-CD is shown along with SLCT (H = Hybrid 2/6).

3.2.3 S-layer null mutants are abnormally sensitive to innate immune effectors

Bacterial S-layers serve many critical cellular functions (*7*). Given its location on the cell surface, the S-layer has been proposed to act as a molecular sieve to selectively limit exposure of the underlying cell envelope to large biomolecules such as the innate immune effector lysozyme (*18*). Until now, analysis of *C. difficile* S-layer function has been hampered by an inability to isolate *slpA*-deficient mutants (*19*). While *C. difficile* is highly resistant to killing by this enzyme, resistance has been attributed to extensive peptidoglycan deacetylation (*20*). To determine if the S-layer also plays a role in lysozyme resistance, we treated exponentially growing bacteria with a high concentration of lysozyme (500 µg/ml) and monitored the effects on growth (Fig.3.3A). Upon addition of lysozyme an immediate decrease in optical density consistent with cell lysis, followed by slower growth, was observed for the S-layer mutant FM2.5. In contrast, R20291 and FM2.5RW displayed only a transient decrease in growth rate, consistent with natural resistance. Having confirmed a role for the S-layer in lysozyme resistance, we then tested for resistance to the human cathelicidin antimicrobial peptide LL-37 to determine if S-layer-mediated resistance extended to other innate immune effectors (Fig.3.3B). LL-37 is found at mucosal surfaces at concentrations of up to 5 µg/ml in normal conditions with further expression induced in response to infection (*21*). Treatment with 5 µg/ml of LL-37 completely killed exponentially growing cultures of the S-layer mutant FM2.5; whereas, the same treatment caused only a slight reduction in growth rates for R20291 and FM2.5RW. Taken together, these data demonstrate a role for the S-layer in resistance to both lysozyme and LL-37.

3.2.4 S-layer null mutants display severe sporulation defects

It was noted that the S-layer mutant strains survived poorly in the standard charcoal medium used to transport *C. difficile* strains. The ability of *C. difficile* to survive in the environment and be transmitted to new hosts is reliant on the bacterium's ability to produce a

heat and chemical-resistant spore (*22*). We analyzed sporulation efficiency by measuring the numbers of heat-resistant spores as a percentage of total viable CFUs over 5 days (Fig.3S4A). Spore production by wild type R20291 and FM2.5RW cultures was reproducible and equivalent and represented 73 and 85% of total viable counts on day 5, respectively (Fig.3.3C). In contrast, spore production by FM2.5 was significantly lower (p = <0.00001). Spores only represented 4.3% of total viable counts on day 5, which is a 17-20-fold reduction compared to R20291 and FM2.5RW. These observed differences in FM2.5 spore formation were not due to an inability to germinate efficiently (Fig.3.3D). When the bile salt germinant taurocholate was added to purified spore preparations, the speed and efficiency of germination initiation for FM2.5 spores was indistinguishable from that of R20291 and FM2.5RW. Analysis of bacterial cultures by phase contrast microscopy also pointed to a reduction in sporulation efficiency (Fig.3S4C and D), as 5.2-fold fewer phase bright spores were observed in cultures of FM2.5 (FM2.5, 8.7% vs R20291, 44.9%; FM2.5RW, 45.4%). Interestingly, we noted a discrepancy between the magnitudes of the sporulation defect determined microscopically (5.2-fold less than R20291) compared with direct counting of viable spore CFUs (20-fold less than R20291). This suggests that many of the microscopically counted spores were nonviable following the 65°C heat treatment required to differentiate spores from vegetative cell CFUs. To test the FM2.5 spores for possible stress resistance defects, we exposed the cultures to a harsher heat treatment (75°C for 30 min). The 75°C heat treatment further reduced FM2.5 spore viability 33-fold compared with the standard 65°C (Fig.3.3C). The same treatment only reduced spore viability by 1.5-fold for R20291 and 2fold for FM2.5RW. Collectively, these data indicate that the production and quality of the infectious spores are severely impaired by the loss of the S-layer.

Transmission electron microscopy was employed to identify potential morphological changes associated with these defects (Fig.3S5). In FM2.5 cultures we observed spores with disorganized material loosely attached to the electron-dense core that lacked discernible, wellorganized protein coat layers (*23*) (Fig.3S5A). To determine if these unusual spore

morphologies were responsible for the observed thermal sensitivity, we repeated these analyses with spores purified on a Histodenz gradient. Following purification, spores of FM2.5 were morphologically indistinguishable from those of R20291 and FM2.5RW (Fig.3S5C). Surprisingly, purified FM2.5 spores still displayed increased thermal sensitivity. A 75°C heat treatment reduced FM2.5 spore viability 37.1-fold compared to 10- and 18.1-fold for R20291 and FM2.5RW respectively (Fig.3S5B). Several distinct biochemical and structural features of the spore, including the concentric cortex peptidoglycan and protein layers and core dehydration, have been independently linked to heat resistance (*24-26*). Minor defects in any of these features could explain the observed thermal sensitivity of FM2.5 spores.
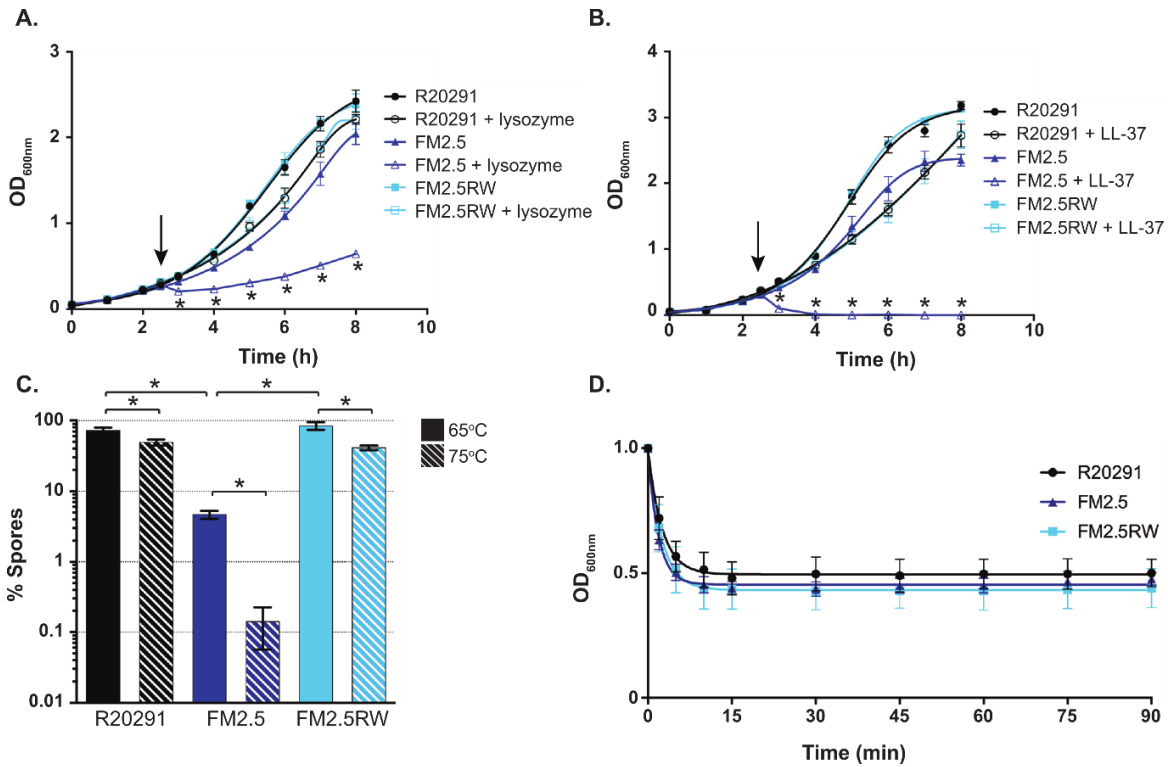
Fig.3.3: Phenotypic characterization of FM2.5. (A and B) Cultures of R20291, FM2.5 and FM2.5RW were challenged with lysozyme (500 µg/ml; **A**) or LL-37 (5 µg/ml; **B**) in exponential phase after 2.5 h (indicated with arrows). Untreated control cultures were grown in parallel. Experiments were carried out in triplicate on biological duplicates. Means and standard deviations are shown. **(C)** Sporulation of R20291, FM2.5, and FM2.5RW after 5 days. Spore CFUs were determined following a standard 65°C heat treatment for 30 minutes or a harsher 75°C heat treatment for 30 minutes. Heat-resistant spore CFUs are expressed as a percentage of total viable CFUs (spores and vegetative cells). Experiments were carried out in duplicate on biological duplicates. Mean and standard deviation are shown. * = P<0.01, determined using two-tailed *t*-tests with Welch's correction. **(D)** Germination of R20291, FM2.5, and FM2.5RW spores. Synchronous germination of purified spores was induced with the bile salt taurocholate. Germination initiation was monitored by measuring the resulting decrease in optical density at OD_{600nm}.

3.2.5 S-layer null mutants are completely avirulent despite persistent gut colonization

We tested the ability of the S-layer mutant FM2.5 to cause disease in the Golden Syrian hamster model of acute CDI. As expected, all animals inoculated with FM2.5RW behaved similarly to animals inoculated with the wild type R20291 strain and succumbed to *C. difficile* infection within 102 h of infection, with mean times to cull for R20291 and FM2.5RW of 67 h 36 min and 57 h 19 min, respectively (p = 0.52, Fig.3.4A). Both groups of hamsters showed typical signs of disease including wet tail and drop in body temperature at experimental endpoint. In contrast, all animals inoculated with FM2.5 displayed no signs of disease and survived for the duration of the 14-day study (p = 0.0018; FM2.5 vs R20291, Fig.3.4A). Very few described mutations located outside of the PaLoc locus have resulted in complete avirulence in this model (*27*). Surprisingly, the lack of virulence was not due to a colonization defect. FM2.5 was capable of persistent colonization; CFUs in the caecum and colon at the end of study (14 days) were not statistically different to those observed for R20291 and FM2.5RW in the same tissues taken at experimental endpoint some 10 days earlier (Fig.3.4B). Toxin measurements from gut contents 14 days after inoculation with FM2.5 showed dramatic reductions in both toxin A or

B activity compared to samples taken from hamsters that succumbed to infection with either the R20291 or FM2.5RW strains (Fig.3.4C and D).

3.2.6 S-layer null mutants lack toxin production in vitro

Given the avirulence and low toxin activity observed in animals, we assayed the FM2.5 and control strains for toxin production *in vitro* (Fig.3.4E). Toxin B was used as an indicator for both toxins since they are coordinately expressed and released (*6*). As expected, both R20291 and FM2.5RW produced toxin upon entry into stationary phase. A small amount of Toxin B was detected intracellularly at 24 hours; thereafter, Toxin B was exclusively detected in the culture

supernatant. Cultures of FM2.5 produced less Toxin B at all time-points, consistent with *in vivo*

observations.
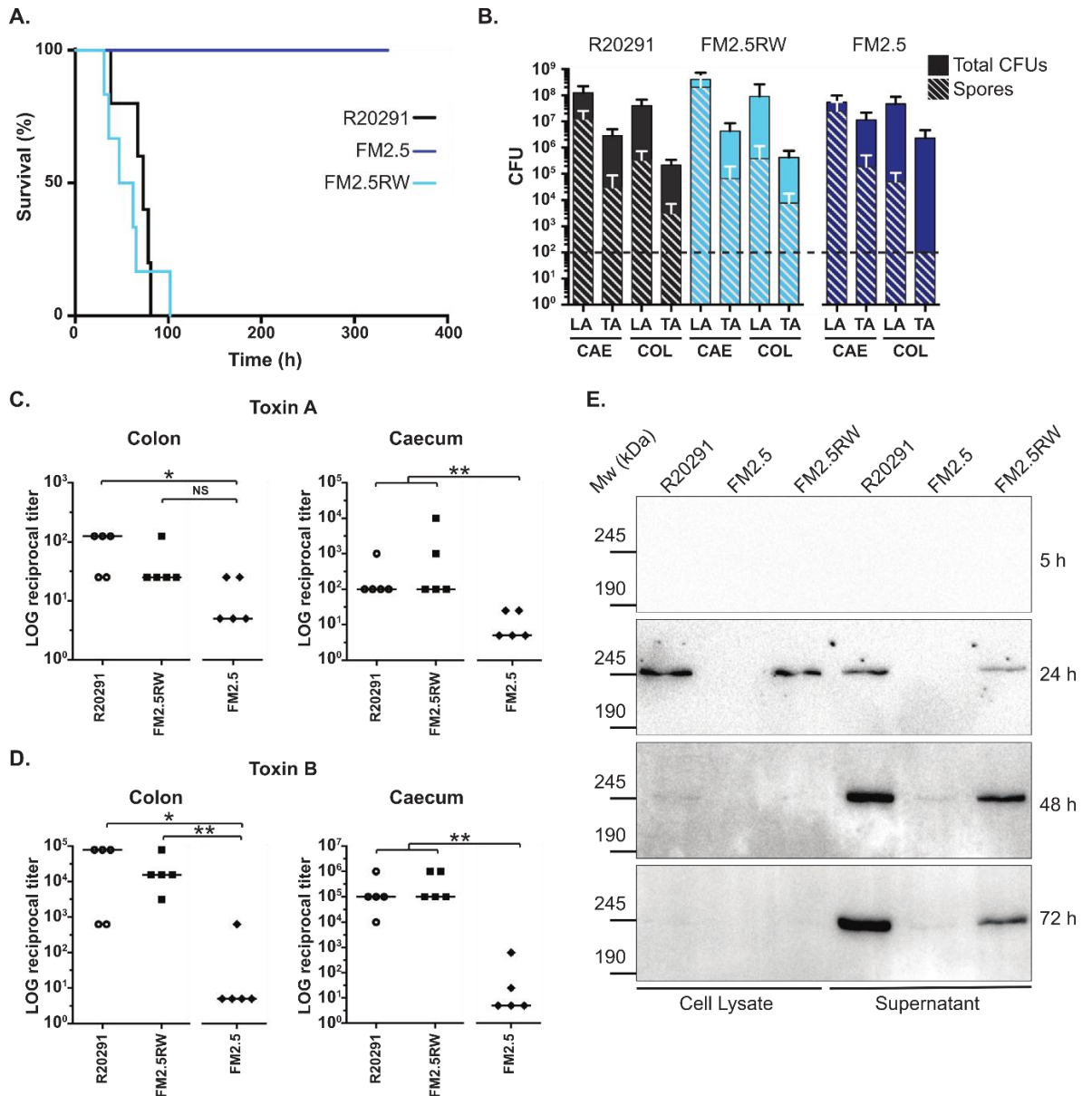
Fig.3.4: In vivo analysis of *slpA* mutant in the Syrian Golden hamster. (A) Times to experimental endpoint of animals infected with R20291 (black line), FM2.5 (dark blue line) and FM2.5RW (light blue line) respectively. Each line represents 6 animals. **(B)** Total CFUs and spore CFUs (following heat treatment at 56°C for 20 min) were determined for Lumen- (LA) and tissue associated (TA) bacteria recovered from caecum (CAE) and colon (COL) of infected animals and quantified at experimental endpoint (R20291 and FM2.5RW) or at 14 days post-infection (FM2.5). Shown are the mean and standard error. The horizontal dotted line indicates the limit of detection. None of the observed differences, including those for the TA-spore CFUs from the colon, are statistically significant. **(C and D)** Relative toxin activity of filtered gut samples on HT-29 (toxin A) and Vero cells (toxin B) respectively. Values represent the reciprocal of the first dilution in which cell morphology was indistinguishable from untreated wells. Samples were taken at experimental endpoint (R20291 and FM2.5RW) or at 14 days post-infection (FM2.5). (* = P <0.05, ** = P <0.01, NS = not significant, determined using a two-tailed nonparametric Mann-Whitney test. **(E)** *In vitro*

cell lysate and culture supernatant samples from R20291, FM2.5 and FM2.5RW were normalized to an equivalent optical density and separated on 6% SDS polyacrylamide gels. Toxin B was detected by Western immunoblot using an anti-toxin B monoclonal antibody. Samples were taken at the indicated time points.

3.3 Discussion

The bactericidal properties of Avidocin-CDs bode well for the clinical application of this new class of antimicrobial agents for combating CDI. Previous studies demonstrated that killing by the prototypic Avidocin-CD, Av-CD291.2, was highly specific for BI/NAP1/027-type strains and did not detectably alter the resident gut microbiota in mice (*12*). For production purposes fermentation of genetically modified *B. subtilis* expressing these agents is similar to many established industrial processes and scalable to thousands of liters. We selected, isolated and characterized rare *C. difficile* mutants resistant to Av-CD291.2 to better understand its mechanism of action. In the process, we discovered the Avidocin-CD binding receptor, the *C. difficile* S-layer protein SlpA. Indeed, our 10 new Avidocin-CDs were all found to target variants of SlpA associated with different SLCTs. Given that the great majority of the sequence variation in SlpA is found within the surface exposed LMW subunit (*9*), this relationship between SLCT and Avidocin-CD sensitivity strongly suggests that the LMW SLP is the binding site for all the studied Avidocin-CDs. Further, it indirectly identifies this portion of the SlpA as the binding receptor for many *C. difficile Myophages* since each Avidocin-CDs is constructed with a different bacteriophage-derived RBP. These findings help explain the limited host ranges observed for many *C. difficile Myophages* (*28, 29*). It also suggests that antigenic variation observed between SLCTs may not be due to immune escape as previously proposed (*9*) but rather due to a molecular arms race between bacteria and bacteriophages. *C. difficile* is under selective pressure to change the bacteriophage receptor, which in turn puts selective pressure on the bacteriophages to evolve new RBPs.

Interestingly, patterns of strain sensitivity to each Avidocin-CD indicate killing was much broader than the parental bacteriophage's host range (Fig.3S2 and 3S7). A possible explanation for this observation is that a typical bacteriophage infection cycle is a 7-step process: attachment, genome injection, replication, transcription, translation, assembly and lysis. Bacteria can become resistant to a bacteriophage by blocking any stage of the infection cycle.

For example, CRISPR-Cas or restriction-modification systems that prevent phage DNA from replicating (*30*). In contrast, Avidocin-CD killing is a 2-step process: attachment to the target bacterium followed by killing via creation of a small pore that dissipates the target bacterium's membrane potential, importantly, without lysis or release of macromolecules such as exotoxin from the cytoplasm (Fig.3S6). As observed with other Avidocins and R-type bacteriocins (*16, 31*), selected Avidocin-CD resistant mutants survived due to mutations that caused the loss or modification of the SlpA binding receptor and not due to mutations that directly disrupted the killing mechanism, such as improper pore formation, or caused proteolytic cleavage of the agent. The bias towards receptor mutations suggests the Avidocin-CD killing mechanism is simple and robust.

Modification of the binding receptor to avoid Avidocin-CD killing either through missense mutations or S-layer switching, as evidenced by the random association between clades and SLCTs (*9*), could also theoretically lead to resistance. While sequence variations between *slpA* alleles within each SLCT indicate missense mutations do occur (*9*), our findings suggest resistance due to this type of modification is unlikely since the bacteriophage RBPs used to construct the Avidocin-CDs have already evolved to counter this mode of potential escape. For instance, both Av-CD684.1 and Av-CD685.1 kill every SLCT-7 strain tested despite sequence identities between SLCT-7 SlpAs as low as 81% (Fig.3S8). As for the emergence of resistance via horizontal transfer of the S-layer cassette, the administration of a cocktail of Avidocin-CDs that kill all the common SLCTs would make successful resistance via S-layer switching extremely unlikely. It appears that the only likely means of resistance to all the Avidocin-CDs is through complete loss of SlpA, as observed for resistance to Av-CD291.2. As a consequence the observed phenotypes for FM2.5 are germane to all likely Avidocin-CD-resistant mutants.

For a precision medicine agent to be successful, knowing the molecular target, in this case the binding receptor, is vital in designing accurate diagnostics to guide treatment decisions. If Avidocin-CDs were to be administered individually, a diagnostic determining SLCT of infecting

C. difficile would be highly accurate for informing Avidocin-CD treatment decisions. However, an increase in the prevalence of hybrid cassettes, as found in ribotype 078 strains (*9*), would decrease the accuracy for this typing method. Strain ribotyping could also be employed to avoid development of new diagnostics, but predicting sensitivity to a particular Avidocin-CD may be challenging as ribotypes are not consistently linked to a single SLCT (*i.e.* 012, 014, and 015, as noted above). SlpA-typing would provide the most accurate diagnostic as well as prove more illuminating than ribotyping since SlpA is directly related to the physiology of *C. difficile*, whereas, ribotyping detects physiologically inconsequential ribosomal RNA gene polymorphisms. It may also be possible to administer a cocktail of 5-6 Avidocin-CDs that target 12 of the 14 *C. difficile* SLCTs. If such a cocktail of Avidocin-CDs were to be administered, a point of care diagnostic would only need to detect the presence of *C. difficile* to guide treatment decisions.

The strong selective pressure afforded by Av-CD291.2 allowed isolation of the first spontaneous *C. difficile* S-layer null mutants. In addition to enabling identification of the Avidocin-CD cell surface receptor, these mutants also provide an unprecedented opportunity to study S-layer function (Fig.3.5). Given the ubiquity of S-layers in both Bacteria and Archaea, including many pathogenic species, surprisingly little is known about their function. It has been suggested that the S-layer could act as a molecular sieve to exclude certain large biomolecules from the cell envelope (*18*); however, this has not been confirmed in live cells. We have demonstrated that an intact S-layer is required for resistance to two components of the innate immunity system, lysozyme and the antimicrobial peptide LL-37. Assembled S-layers are highly symmetrical with regular repeating pores. The size of the pores in the *C. difficile* S-layer is not yet known but in other species pore sizes of between 2 and 6 nm have been reported. A pore size of 2 nm could conceivably exclude the16 kDa globular protein lysozyme but a small peptide such as LL-37 would experience no such steric hindrance. It is possible that charged surfaces on the assembled S-layer serve to sequester the cationic peptide away from the cell envelope in a

manner analogous to capsular polysaccharides (*32*). Our data has identified other pleiotropic functions for the *C. difficile* S-layer. It is clear that the S-layer is the cell surface receptor for all of the Avidocin-CDs described here and, by extension, the receptor for the bacteriophages from which the RBP-encoding genes were cloned. Although a Bacillus bacteriophage has been found to bind S-layer protein Sap (*33*), this is the first time a receptor for a *C. difficile* bacteriophage has been identified. Furthermore, our data imply that S-layer recognition is a common feature of bacteriophages that infect this species.

Surprisingly, the S-layer mutant also displayed severe sporulation defects, with fewer and morphologically defective spores produced. Despite a number of well-studied spore-forming organisms producing S-layers, including *B. anthracis* (*33*), S-layer biogenesis has not been reported previously to affect sporulation. There is currently no evidence to suggest that the SLPs are a structural part of the mature spore and the mechanisms by which the S-layer can influence sporulation are currently unknown. However, this observation has serious ramifications for the ability of an Avidocin-CD-resistant, S-layer defective mutant to survive and be transmitted to other hosts. The spore is an absolute requirement for *C. difficile* survival in the aerobic environment and is critical for transmission (*22*). In addition to sporulation defects, the S-layer mutant was also found to produce less toxin *in vitro*. Interestingly, there have been previous suggestions of feedback between sporulation and the complex regulatory network controlling toxin production (*22*). Although the mechanism by which these processes are affected in SlpA null mutants is far from clear, it is possible that the S-layer feeds into a point of crosstalk between regulation of virulence and transmission. Given the poor toxin production and other diverse phenotypes identified for the S-layer mutant, it is probably not surprising that the mutant was entirely avirulent in the hamster model of acute infection. What did come as a surprise, however, was the ability of the S-layer mutant to stably colonize and persist in the hamster gut for the 14-day duration of the experiment. Previous reports have pointed to a role for the S-layer in epithelial cell adhesion; however, these earlier studies were performed without

access to an *slpA*-defective strain (*34*). The S-layer defective mutants and isogenic controls expressing the SLCT-specific *slpA* alleles provide the ideal controls to test these conflicting findings and better define the effect SlpA type has on these functions.

There are several study limitations that should be considered when evaluating the data presented. The activity of each new Avidocin-CD was observed *in vitro* and needs to be confirmed in an animal treatment model, as done for Av-CD291.2. Similarly, the exquisite specificity of each agent for distinct *C. difficile* SLCTs also needs be tested *in vivo* to confirm that treatment with these agents will not alter the diversity of the gut microbiota. The avirulent phenotype and long-term persistence of the Avidocin-CD-resistant *slpA* mutants in the hamster model should also be confirmed in other animal species. Finally, the implications of the mutants' observed sporulation defects on transmission and the spread of resistance remain to be tested *in vivo*. In all cases, we do not anticipate different outcomes from those described here.

After accounting for these limitations, analysis of the data clearly demonstrates that acquisition of Avidocin-CD resistance results in loss of toxin production and complete loss of virulence. Virulence factors make attractive targets for new antimicrobials as they tend to be species-specific, and emergence of resistance is likely to reduce virulence. The ability of the S-layer mutant to stably colonize the gut in the absence of clinical disease reveals that the *in vivo* lifestyle of the organism is independent of toxin production and virulence. The prevalence of non-toxigenic, avirulent *C. difficile* strains in the general population (*35*) supports this hypothesis. As a result, we predict that there will be no competing selective pressure to restore virulence in the context of Avidocin-CD resistance.

In summary, we have developed and characterized multiple new Avidocin-CDs, providing crucial insights into their potential advantages in the clinic. The precise killing activity of Avidocin-CDs makes them attractive agents for both treatment and prevention since they can be administered to patients without altering the diversity of the complex gut microbiota. In

addition, when resistance does emerge, Avidocin-CDs force the pathogen to sacrifice virulence for viability - making the potential clinical impact of resistance inconsequential.

A.

Wild Type *C. difficile*:

Avidocin-CD Killing Bacteriophage Adsorption

Lysozyme LL-37

S-layer

Peptidoglycan & Secondary Cell Wall Polymers

Cytoplasmic Membrane

Toxin Production

Endospore → Sporulation

Cytoplasm

Virulence

? → Host Cell Adhesion

TLR4 Binding

? → Cytokine Activation

B.

S-layer Mutant

Resistance

Lysozyme LL-37

Peptidoglycan & Secondary Cell Wall Polymers

Cytoplasmic Membrane

Toxin Production

Defective endospore → Sporulation

Cytoplasm

Virulence

? → Host Cell Adhesion

TLR4 Binding

C.

Unbound Avidocin-CD

Contractile Sheath

Tail Fibers

RBPs

Binding to S-layer

Bound Avidocin-CD

Loss of cations, ATP and H$^+$

Nanotube Core

Fig.3.5: Schematic diagram depicting the phenotypes of Wild Type (A) and S-layer null mutant (B) cells in *C. difficile* biology and pathophysiology. Processes labelled in black indicate new functions discerned in this manuscript. Processes labelled in grey indicate previously identified functions. Dotted lines indicate relationships that need to be studied further. Question marks indicate previously described connections incompatible with observations made with the S-layer null mutant and warranting further investigation. **(C)** Schematic of Avidocin-CD structure and function. Binding to *C. difficile* by the Avidocin-CD receptor binding proteins (RBPs) triggers the sheath to contract and force the hollow nanotube core across the cell envelope. The resulting pore allows small metabolites, such as protons, ATP and cations, to escape from the cell cytoplasm, which, in turn, disrupts the cell's membrane potential and kills the cell.

3.4 Materials and Methods

3.4.1 Study design

The objective of this study was to characterize a panel of Avidocin-CDs, anti-bacterial protein complexes constructed to specifically target and kill *C. difficile*. During these experiments, the isolation of Avidocin-CD resistant *slpA* mutants allowed detailed analysis of S-layer function for the first time. Starting from the prototypic Av-CD291.2 (*12*) we constructed a panel of new Avidocin-CDs using bacteriophage RBPs we identified. Each new Avidocin-CD displayed a unique spectrum of killing activity with a strong correlation to SLCT. Sensitization of two insensitive *C. difficile* strains by heterologous expression of a cognate SLCT SlpA alone allowed identification of the S-layer as the receptor for all described Avidocin-CDs. Analysis of the *slpA* mutant identified previously unsuspected *in vitro* roles for the S-layer in resistance to the immune effectors lysozyme and LL-37, and in the production of mature heat-resistant spores. Finally, use of the Golden Syrian hamster model of acute infection demonstrated that the *slpA* mutant was entirely avirulent despite persistent infection. Greatly reduced toxin activity was detected in intestinal contents from animals colonized with the *slpA* mutant and this observation was supported by identification of a toxin production defect *in vitro*..

3.4.2 Strains, bacteriophage and culture conditions.

Bacterial strains used in this study are described in Figure 3S2. DNA oligonucleotides are described in Table 3S2. *E. coli* strains were routinely grown in LB broth and on LB agar (VWR).
C. difficile strains were routinely grown under anaerobic conditions on BHI or BHI-S agar and in TY broth (*36*) except where otherwise stated. Cultures were supplemented with chloramphenicol (15 μg/ml), thiamphenicol (15 μg/ml) or anhydrotetracycline (20 ng/ml) as required. *C. difficile* SLCTs were determined by analyzing the nucleotide sequence of the *slpA* gene. When necessary, the *slpA* gene was sequenced using oligonucleotide primers previously described (*37, 38*) or primers 023-F and 023_010-R or 014+++-F and 014_002+-R (nucleotide

sequences – Table 3S2). The variable region of strain 19142 (ribotype 046) *slpA* gene did not

display high sequence identity with other *slpA* alleles, and has been designated SLCT-13. A partial

strain 19142 *slpA* sequence was deposited in Genbank (accession: KX610658).

3.4.4 Cloning of *slpA* alleles and construction of revertants.

The *slpA* genes from a panel of *C. difficile* strains with distinct SLCTs were amplified using the

primers listed in Table 3S2 and cloned between *Bam*HI and *Sac*I sites in the inducible expression

plasmid pRPF185 (*40*). Plasmids were transferred into *C. difficile* strain FM2.5 by conjugation

from the *E. coli* donor CA434. For construction of *slpA* revertants, a watermark consisting of two

adjacent synonymous mutations, was introduced into the SLCT-4 *slpA* using primers RF135 and

RF136. Single crossover recombinant colonies were identified by an enhanced growth

phenotype, and recombination was confirmed by PCR using primers RF33 and NF1323. Stable

double crossover revertants were isolated following growth without selection for approximately

20 generations and confirmed by screening for plasmid loss, sequencing of the *slpA* gene, and

confirmation of SLP expression by SDS-PAGE analysis (described below).

3.4.5 Construction and expression of Avidocin-CDs.

New Avidocin-CDs were constructed with bacteriophage RBPs identified from newly isolated

bacteriophages and prophage insertions. The RBP genes along with the adjacent 3'end of the

Baseplate attachment region (Bpar) and chaperone genes were cloned and inserted into an

integration vector containing the Diffocin-4 gene cluster via a combination of restriction enzyme

digests and Gibson assembly. The fusion sites used to replace the bacteriocin genes and

integrate the bacteriophage derived genes are shown (Fig.3S9). Avidocin constructs using the

downstream fusion site 1 are denoted with a ".1" suffix (i.e. Av-CD681.1), while constructs using

the downstream fusion site 2 are denoted with a ".2" (i.e. Av-CD027.2). The resulting integration

vectors were sequence verified and transformed into the *B. subtilis* production strain as

previously described (*12*). The nucleotide sequences for the Bpar fusion, RBP and downstream

chaperones have been deposited in Genbank along with source information. Accession numbers are provided in Table 3S3. Expression of each Avidocin-CD was induced with IPTG as previously described (*12*). Preparation of each Avidocin-CD and Diffocin-4 relied on a combination of Polyethlene-glycol-8000 precipitations and ultracentrifugation as described previously (*12*).

3.4.6 Isolation of Av-CD291.2-resistant clones.

A greater than 100-fold excess of active Av-CD291.2 was added to exponentially growing *C. difficile* strain R20291. After 75 min bacteria were harvested, resuspended in phosphate buffered saline (PBS), and spread on non-selective Brucella agar plates maintained under anaerobic conditions at 37°C overnight. CFU counts before and after addition of Av-CD291.2 were compared to determine resistance frequency. Resistant colonies were streaked on *C. difficile* selective agar and Av-CD291.2-resistance was confirmed. Av-CD291.2-resistant strains FM2.5 and FM2.6 was sequenced by Illumina Miseq and identified *slpA* mutations were confirmed by Sanger sequencing.

3.4.7 Bioassays to determine Avidocin-CD killing activity.

Avidocin-CD bactericidal activity was assayed by a semi-quantitative spot method as previously described (*12, 13*). For broth-based killing assays, *C. difficile* strains were grown overnight in TY broth and then sub-cultured to an OD_{600nm} of 0.05 in 1 ml fresh TY supplemented with 1mM $CaCl_2$. Av-CD291.2 (50 µl) was added to each culture after 2.5 h. Growth was monitored by measuring the OD_{600nm} hourly.

3.4.8 Extraction of S-layer and associated proteins.

S-layer proteins were extracted using low pH glycine as previously described (*8*) and analyzed by SDS-PAGE using standard methods.

3.4.9 Quantitative analysis of sporulation and germination.

Overnight bacterial cultures were diluted to 0.01 OD_{600nm} in BHI-S broth, incubated for 8 hours, diluted again to 0.0001 OD_{600nm} and grown overnight. The resulting stationary phase cultures (T=0) were then incubated for 5 days. The relative proportion of spores and vegetative cells was determined at 24 h intervals by counting total and heat-resistant (65°C for 30 min) CFUs on BHI-S agar supplemented with 0.1 % taurocholate. Assays were repeated in triplicate with biological triplicates. In order to test a range of different stress conditions, additional samples were taken on day 5 and treated at either 75°C for 30 min, 50% ethanol for 30 min or incubated aerobically for 18 h before enumerating viable spores as before. For the germination studies, spores were purified as previously described (*19*) and Ca^{2+}-DPA release and core rehydration was monitored by measuring changes in OD_{600nm} upon addition of 0.5% taurocholate.

For microscopic examination, bacterial samples were washed with PBS and fixed in 3.2% paraformaldehyde. For phase contrast microscopy, samples were mounted in 80% glycerol and visualised using a Nikon Ti Eclipse inverted microscope. For TEM, samples were additionally fixed in 3% glutaraldehyde in 0.1 M cacodylate buffer followed by 1% OsO_4, then dehydrated in ethanol and embedded in araldite resin. Embedded samples were sectioned at 85 nm on a Leica UC6 ultramicrotome, transferred onto coated copper grids, stained with uranyl acetate and lead citrate and visualized using a FEI Tecnai BioTWIN TEM at 80 kV fitted with a Gatan MS600CW camera.

3.4.10 Analysis of resistance to lysozyme and LL-37.

Broth based killing assays were carried out as described above but with lysozyme (500 µg/ml) or LL-37 (5 µg/ml) added after 2.5 h growth. Cell density was monitored by measuring the OD_{600nm} hourly. Assays were carried out in triplicate on biological duplicates.

3.4.11 Animal experiments.

The Golden Syrian hamster model was performed as previously described (*39*). All procedures were performed in strict accordance with the Animals (Scientific Procedures) Act 1986 with specific approval granted by the Home Office, U.K. (PPL60/4218).

Briefly, 3 weeks before bacterial challenge, telemetry chips were surgically inserted by laparotomy into the peritoneal cavity of female Golden Syrian hamsters. Animals were allowed to recover from surgery before being placed in cages above transponder receiver plates that allow direct monitoring of body temperature. All animals were orally dosed with clindamycin phosphate (30 mg/kg) five days prior to challenge with 10^4 spores of either *C. difficile* R20291, FM2.5 or FM2.5RW as previously described (*39*). Six animals per group were inoculated and animals monitored carefully for onset of symptoms. Animals were culled when core body temperature dropped below 35°C.

To quantify colonization at experimental endpoint, the caecum (CAE) and colon (COL) of each animal were removed, opened longitudinally and washed thoroughly in PBS to remove unbound bacteria; these organisms were designated luminal associated organisms (LA). The washed tissue was then homogenized and subsequently recovered material contained more intimately associated organisms; designated tissue associated (TA). Total viable counts were determined by plating serial dilutions on ChromID (Biomeuriex). Spores were enumerated following heat treatment at 56°C for 20 min. Colonies from all plates were subjected to PCR amplification of both the *slpA* gene specifically, using primers RF110 and RF111, and seven repeat regions to confirm strain identity by MVLA analysis 2 to confirm strain identity.

3.4.12 Quantification of toxin expression.

Quantification of toxin activity was performed using a Vero cell based cytotoxicity assay. Dilutions of filtered (0.45 µm) post-challenge gut contents were added to confluent monolayers for 24h and the end-point titer was defined as the first dilution at which Vero cell morphology

was indistinguishable from untreated wells. Each experiment was performed in duplicate on two separate occasions. For measurements of toxin production *in vitro*, whole cell lysates or culture supernatants were separated by SDS-PAGE, transferred to PVDF membrane and toxin B detected by Western immunoblot using a specific mouse monoclonal antibody (MA1-7413, Thermo Fisher).

3.4.13 Statistical analyses.

Data were analyzed using GraphPad Prism software (GraphPad Software Inc.). Toxin production was compared using a two-tailed nonparametric Mann-Whitney test, and animal survival curves were analyzed using a Log-rank (Mantel-Cox) test. All other statistical analyses were performed using two-tailed *t*-tests with Welch's correction.

3.5 References

1. J. R. Marchesi, D. H. Adams, F. Fava, G. D. Hermes, G. M. Hirschfield, G. Hold, M. N. Quraishi, J. Kinross, H. Smidt, K. M. Tuohy, L. V. Thomas, E. G. Zoetendal, A. Hart, The gut microbiota and host health: a new clinical frontier. *Gut* **65**, 330-339 (2016).

2. D. A. Rasko, V. Sperandio, Anti-virulence strategies to combat bacteria-mediated disease. *Nat. Rev. Drug Discov.* **9**, 117-128 (2010).

3. R. C. Allen, R. Popat, S. P. Diggle, S. P. Brown, Targeting virulence: can we make evolution-proof drugs? *Nat. Rev. Microbiol.* **12**, 300-308 (2014).

4. F. C. Lessa , Y. Mu , W. M. Bamberg , Z. G. Beldavs , G. K. Dumyati , J. R. Dunn , M. Farley , S. M. Holzbauer , J. I. Meek , E. C. Phipps , L. E. Wilson , L. G. Winston , J. Cohen , B. M. Limbago , S. K. Fridkin , D. N. Gerding , L. C. McDonald Burden of *Clostridium difficile* Infection in the United States. *New Engl. J. Med.* **372**, 825-834 (2015).

5. "ANTIBIOTIC RESISTANCE THREATS in the United States, 2013," *Threat Report 2013* (Centers for Disease Control and Prevention, Atlanta, 2013).

6. W. K. Smits, D. Lyras, D. B. Lacy, M. H. Wilcox, E. J. Kuijper, *Clostridium difficile* infection. *Nat Rev Dis Primers* **2**, 16020 (2016).

7. R. P. Fagan, N. F. Fairweather, Biogenesis and functions of bacterial S-layers. *Nat. Rev. Microbiol.* **12**, 211-222 (2014).

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

9. K. E. Dingle, X. Didelot, M. A. Ansari, D. W. Eyre, A. Vaughan, D. Griffiths, C. L. Ip, E. M. Batty, T. Golubchik, R. Bowden, K. A. Jolley, D. W. Hood, W. N. Fawley, A. S. Walker, T. E. Peto, M. H. Wilcox, D. W. Crook, Recombinational switching of the *Clostridium difficile* S-layer and a novel glycosylation gene cluster revealed by largescale whole-genome sequencing. *J. Infect. Dis.* **207**, 675-686 (2013).

10. S. E. Willing, T. Candela, H. A. Shaw, Z. Seager, S. Mesnage, R. P. Fagan, N. F. Fairweather, *Clostridium difficile* surface proteins are anchored to the cell wall using CWB2 motifs that recognise the anionic polymer PSII. *Mol. Microbiol.* **96**, 596-608 (2015).

11. A. Ryan, M. Lynch, S. M. Smith, S. Amu, H. J. Nel, C. E. McCoy, J. K. Dowling, E. Draper, V. O'Reilly, C. McCarthy, J. O'Brien, D. Ni Eidhin, M. J. O'Connell, B. Keogh, C. O. Morton, T. R. Rogers, P. G. Fallon, L. A. O'Neill, D. Kelleher, C. E. Loscher, A role for TLR4 in *Clostridium difficile* infection and the recognition of surface layer proteins. *PLoS Pathog* **7**, e1002076 (2011).

12. D. Gebhart, S. Lok, S. Clare, M. Tomas, M. Stares, D. Scholl, C. J. Donskey, T. D. Lawley, G. R. Govoni, A modified R-type bacteriocin specifically targeting *Clostridium difficile* prevents colonization of mice without affecting gut microbiota diversity. *mBio* **6**, e02368 (2015).

13. D. Gebhart, S. R. Williams, K. A. Bishop-Lilly, G. R. Govoni, K. M. Willner, A. Butani, S. Sozhamannan, D. Martin, L. C. Fortier, D. Scholl, Novel high-molecular-weight, Rtype bacteriocins of *Clostridium difficile*. *J. Bacteriol.* **194**, 6240-6247 (2012).

14. P. Ge, D. Scholl, P. G. Leiman, X. Yu, J. F. Miller, Z. H. Zhou, Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nat. Struct. Mol. Biol.* **22**, 377-382 (2015).

15. D. Scholl, M. Cooley, S. R. Williams, D. Gebhart, D. Martin, A. Bates, R. Mandrell, An engineered R-type pyocin is a highly specific and sensitive bactericidal agent for the food-borne pathogen *Escherichia coli* O157:H7. *Antimicrob. Agents Chemother.* **53**, 3074-3080 (2009).

16. D. Scholl, D. Gebhart, S. R. Williams, A. Bates, R. Mandrell, Genome sequence of *E. coli* O104:H4 leads to rapid development of a targeted antimicrobial agent against this emerging pathogen. *PLoS One* **7**, e33637 (2012).

17. S. R. Williams, D. Gebhart, D. W. Martin, D. Scholl, Retargeting R-type pyocins to generate novel bactericidal protein complexes. *Appl. Environ. Microbiol.* **74**, 3868-3876 (2008).

18. M. Sara, U. B. Sleytr, Molecular sieving through S layers of *Bacillus stearothermophilus* strains. *J. Bacteriol.* **169**, 4092-4098 (1987).

19. M. Dembek, L. Barquist, C. J. Boinett, A. K. Cain, M. Mayho, T. D. Lawley, N. F. Fairweather, R. P. Fagan, High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *mBio* **6**, e02383 (2015).

20. T. D. Ho, K. B. Williams, Y. Chen, R. F. Helm, D. L. Popham, C. D. Ellermeier, *Clostridium difficile* extracytoplasmic function sigma factor sigmaV regulates lysozyme resistance and is necessary for pathogenesis in the hamster model of infection. *Infect. Immun.* **82**, 2345-2355 (2014).

21. D. M. Bowdish, D. J. Davidson, Y. E. Lau, K. Lee, M. G. Scott, R. E. Hancock, Impact of LL-37 on anti-infective immunity. *J Leukoc Biol* **77**, 451-459 (2005).

22. L. J. Deakin, S. Clare, R. P. Fagan, L. F. Dawson, D. J. Pickard, M. R. West, B. W. Wren, N. F. Fairweather, G. Dougan, T. D. Lawley, The *Clostridium difficile spo0A* gene is a persistence and transmission factor. *Infect. Immun.* **80**, 2704-2711 (2012).

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

24. J. Barra-Carrasco, V. Olguin-Araneda, A. Plaza-Garrido, C. Miranda-Cardenas, G. CofreAraneda, M. Pizarro-Guajardo, M. R. Sarker, D. Paredes-Sabja, The *Clostridium difficile* exosporium cysteine (CdeC)-rich protein is required for exosporium morphogenesis and coat assembly. *J. Bacteriol.* **195**, 3863-3875 (2013).

25. P. Setlow, Spores of *Bacillus subtilis*: their resistance to and killing by radiation, heat and chemicals. *J. Appl. Microbiol.* **101**, 514-525 (2006).

26. P. C. Strong, K. M. Fulton, A. Aubry, S. Foote, S. M. Twine, S. M. Logan, Identification and characterization of glycoproteins on the spore surface of *Clostridium difficile*. *J. Bacteriol.* **196**, 2627-2637 (2014).

27. E. M. Ransom, K. B. Williams, D. S. Weiss, C. D. Ellermeier, Identification and characterization of a gene cluster required for proper rod shape, cell division, and pathogenesis in *Clostridium difficile*. *J. Bacteriol.* **196**, 2290-2300 (2014).

28. J. Y. Nale, J. Spencer, K. R. Hargreaves, A. M. Buckley, P. Trzepinski, G. R. Douce, M. R. Clokie, Bacteriophage combinations significantly reduce *Clostridium difficile* growth *in vitro* and proliferation *in vivo*. *Antimicrob. Agents Chemother.* **60**, 968-981 (2016).

29. O. Sekulovic, J. R. Garneau, A. Neron, L. C. Fortier, Characterization of temperate phages infecting *Clostridium difficile* isolates of human and animal origins. *Appl. Environ. Microbiol.* **80**, 2555-2563 (2014).

30. J. E. Samson, A. H. Magadan, M. Sabri, S. Moineau, Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol.* **11**, 675-687 (2013).

31. J. M. Ritchie, J. L. Greenwich, B. M. Davis, R. T. Bronson, D. Gebhart, S. R. Williams, D. Martin, D. Scholl, M. K. Waldor, An *Escherichia coli* O157-specific engineered pyocin prevents and ameliorates infection by *E. coli* O157:H7 in an animal model of diarrheal disease. *Antimicrob. Agents Chemother.* **55**, 5469-5474 (2011).

32. E. Llobet, J. M. Tomas, J. A. Bengoechea, Capsule polysaccharide is a bacterial decoy for antimicrobial peptides. *Microbiology* **154**, 3877-3886 (2008).

33. R. D. Plaut, J. W. Beaber, J. Zemansky, A. P. Kaur, M. George, B. Biswas, M. Henry, K. A. Bishop-Lilly, V. Mokashi, R. M. Hannah, R. K. Pope, T. D. Read, S. Stibitz, R. Calendar, S. Sozhamannan, Genetic evidence for the involvement of the S-layer protein gene *sap* and the sporulation genes *spo0A*, *spo0B*, and *spo0F* in Phage AP50c infection of *Bacillus anthracis*. *J. Bacteriol.* **196**, 1143-1154 (2014).

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

35. J. K. Shim, S. Johnson, M. H. Samore, D. Z. Bliss, D. N. Gerding, Primary symptomless colonisation by *Clostridium difficile* and decreased risk of subsequent diarrhoea. *Lancet* **351**, 633-636 (1998).

36. B. Dupuy, A. L. Sonenshein, Regulated transcription of *Clostridium difficile* toxin genes. *Mol. Microbiol.* **27**, 107-120 (1998).

37. H. Kato, Typing by sequencing the *slpA* gene of *Clostridium difficile* strains causing multiple outbreaks in Japan. *J. Med. Microbiol.* **54**, 167-171 (2005).

38. T. Karjalainen, N. Saumier, M. C. Barc, M. Delmee, A. Collignon, *Clostridium difficile* genotyping based on *slpA* variable region in S-layer gene sequence: an alternative to serotyping. *J. Clin. Microbiol.* **40**, 2452-2458 (2002).

39. A. M. Buckley, J. Spencer, D. Candlish, J. J. Irvine, G. R. Douce, Infection of hamsters with the UK *Clostridium difficile* ribotype 027 outbreak strain R20291. *J. Med. Microbiol.* **60**, 1174-1180 (2011).

40. R. P. Fagan, N. F. Fairweather, *Clostridium difficile* has two parallel and essential Sec secretion systems. *J. Biol. Chem.* **286**, 27483-27493 (2011).

41. C. S. Wong, S. Jelacic, R. L. Habeeb, S. L. Watkins, P. I. Tarr, The risk of the hemolyticuremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. *N Engl J Med* **342**, 1930-1936 (2000).

42. J. Peltier, H. A. Shaw, E. C. Couchman, L. F. Dawson, L. Yu, J. S. Choudhary, V. Kaever, B. W. Wren, N. F. Fairweather, Cyclic diGMP regulates production of sortase substrates of *Clostridium difficile* and their surface exposure through ZmpI proteasemediated cleavage. *J. Biol. Chem.* **290**, 24453-24469 (2015).

3.6 Acknowledgments:

3.7 Author contributions:

D.S., G.R.D., G.R.G. and R.P.F. designed and coordinated the study. J.A.K, D.G., A.M.B., S.L., G.R.G. conducted the experiments. G.R.D., G.R.G. and R.P.F. wrote the manuscript with contributions from all co-authors.

3.8 Competing interests:

D.G., S.L., D.S., and G.R.G. are current or past employees of and own stock in AvidBiotics Corp. R.P.F. received a research grant from AvidBiotics Corp. AvidBiotics Corp. hold the following patents: US8206971 (Modified bacteriocins and methods for their use), US8673291

(Diffocins and methods of use therof), US9115354 (Diffocins and methods of use therof), and EP2576604 (Diffocins and methods of use therof)

3.9 Data and materials availability:

Nucleotide sequences have been deposited in Genbank with accession identifiers: KX610658, KX557294, KX592438, KX592434, KX592441. KX592442, KX592443, KX592444, KX592439, KX592435, KX592437, KX592436, KX592440. Avidocin-CDs are available from

AvidBiotics Corp. subject to a material transfer agreement

3.10 supplemental tables and figures
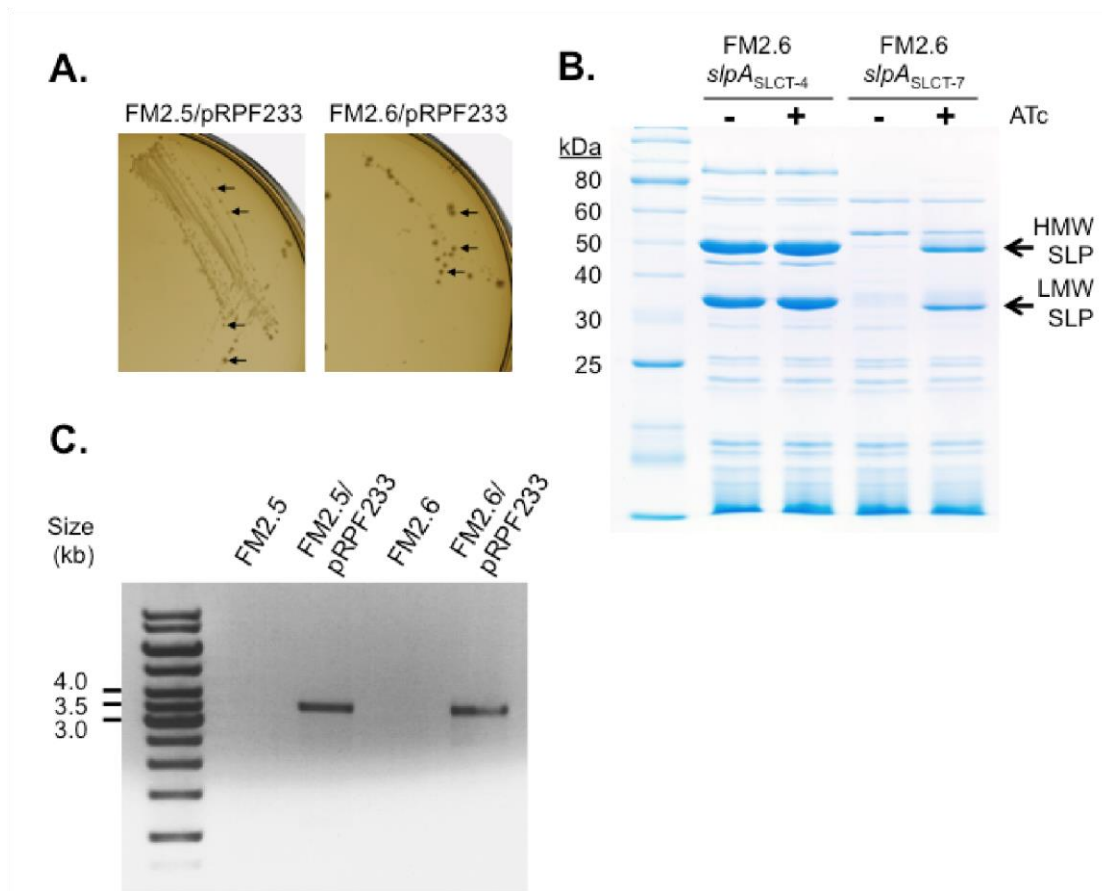
Fig.3S1: Restoration of wild type *slpA* to the chromosome of FM2.5 and FM2.6. Plasmid pRPF233, carrying a wild type copy of the R20291 *slpA* gene under the control of a tetracycline-inducible promoter, was introduced into FM2.5 and FM2.6 by conjugation from an *E. coli* donor strain CA434. **(A)** When grown on solid media, supplemented with thiamphenicol to select for the plasmid, both FM2.5/pRPF233 and FM2.6/pRPF233 produced two distinct colony morphologies; small colonies that were indistinguishable from those typical of FM2.5 and FM2.6 and larger colonies that appeared to grow significantly faster than the *slpA* mutants. **(B)** When the S-layer profile of the larger colonies was analyzed by SDS-PAGE constitutive SlpA expression, independent of anhydrotetracycline (ATc) induction, was observed. However, when the *slpA* gene from strain 630 (SLCT-7), which shares 69.4% nucleotide identity with *slpA*$_{R20291}$ (SLCT-4), was introduced expression of SlpA remained inducible. **(C)** Analysis of these large colonies by PCR using one chromosome-specific primer (RF33, annealing 1,250 bp downstream of *slpA* in the *secA2* gene) and one plasmid-specific primer (NF1323, annealing in the Ptet promoter 122 bp upstream of *slpA* on pRPF233) confirmed that pRPF233 had integrated into the chromosome at the *slpA* locus by a single crossover recombination event. Stable *slpA* revertants, FM2.5RW and FM2.6RW, were later generated from these single crossover mutants by culturing the bacteria without selection and then screening for plasmid loss and retention of the fast growth colony phenotype.

Strain Designation	Source	Ribotype	SlpA-type	Diffocin-4	Av-CD291.2	Av-CD027.2	Av-CD681.1	Av-CD682.1	Av-CD684.1	Av-CD685.1	Av-CD305.1	Av-CD123.2	Av-CD242.2	Av-CD147.1	Av-CD630.1
Strain Source (SLCT)				CD4 (12)	R20291 (4)	panel (4)	M68 (7)	M68 (7)	M68 (7)	M68 (7)	CD305 (11)	19123 (6)	CD62 (12)	19147 (1)	CD630 (7)
RBP type				RTB I	PI II-B	phage II-B	PI II-A	PI II-B	PI/Phage II-A	PI II-A	PI II-A	phage III	phage II-A	phage II-A	PI II-A
Ox1424 (ST12)	Dingle	003	1		-	-	-		-	-	-	-	+	+	+
19108	Citron	053	1	+	-	-	-		-	-	-		+	+	+
19147	Citron	053	1	+	-	-	-		-	-			+	+	+
19110	Citron	053	1	+	-	-	-		-	-		-	+	+	+
19098	Citron	057	1	-	-	-	-	-	-	-		-	+	+	+
Totals (Killed/Tested)				3/4	0/5	0/5	0/5	0/1	0/5	0/5	0/2	0/3	5/5	5/5	5/5
Ox858 (ST16)	Dingle	029	2		-	+	-		-	-	-	-	-	-	
Ox2404 (ST16)	Dingle	029	2		-	+	-		-	-	-	-	-	-	
19117	Citron	103	2	-	-	-	-		-	+		-	-	-	-
CD843	Riley	103	2	-	-	-	-	-	-	+		-	-	-	-
Totals (Killed/Tested)				0/1	0/4	2/4	0/4	0/4	0/3	2/4	0/2	0/4	0/4	0/4	0/2
Ox1121 (ST35)	Dingle	284	3												
Totals (Killed/Tested)					0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
19103	Citron	001	4	+	+	+	-		-	-	-	-	-	-	
19135	Citron	001	4	+	+	+	-	+	-	-	-	-	-		
LIV24	Lawley	001	4	+	+	+	-	+	-	-	-	-	-		
BI-9	Lawley	001	4	+	+	+	-	+	-	-	-	-	-		
19137	Citron	015	4	+	+	+	-	+	-	-		-	-	-	
19129	Citron	027	4	+	+	+	-		-	-		-	-	-	
19139	Citron	027	4	+	+	+	-		-	-		-	-	-	
19126	Citron	027	4	-	+	+	-	+	-	-		-	-	-	
20068	Citron	027	4	-	+	+	-	+	-	-		-	-	-	
20315	Citron	027	4	-	+	+									
CD196	Lawley	027	4	-	+	+					-				
R20291	Lawley	027	4	-	+	+	-	+	-	-	-	-	-	-	-
Ox160 (ST1)	Dingle	027	4	+	+	+	-		-	-	-	-	-	-	
ATCC43255	ATCC	087	4	+	+	+	-	+	-	-					-
Totals (Killed/Tested)				8/13	14/14	14/14	0/12	8/8	0/12	0/12	0/4	0/8	0/8	0/8	0/1
Ox1437a (ST7)	Dingle	026	5		-	-	-	+	-	-	-	-	-	-	
Totals (Killed/Tested)					0/1	0/1	0/1		0/1	0/1	0/1	0/1	0/1	0/1	0/1

Strain	Source	No.	CC													
19123	Citron	014	6	-	-	+	+	+	-	-	-	-	-	-	-	-
Ox1896 (ST17)	Dingle	018	6	-	-	+	+	+	-	-	-	-	-	-	-	-
CD4	Fortier	024	6	w	-	+	+	+	-	-	-	-	-	-	-	-
M120	Lawley	078	6H	-	-	+	+	-	-	-	-	-	-	-	-	-
8119	Cepheid	078	6H			+	+									
8401	Cepheid	078	6H			+	+									
8905	Cepheid	078	6H			+	+									
10399	Cepheid	078	6H			+	+									
10411	Cepheid	078	6H			+	+									
Ox575 (ST11)	Dingle	078	6H	-	-	+	+	+	-	-	-	-	-	-	-	-
Totals (Killed/Tested)		012	6	0/2	0/4	10/10	10/10	2/3	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5
CD630	Lawley	012	7	-	-	-	-	-	+	+	-	+	-	-	-	-
M68	Lawley	017	7	+	-		-	-	+	+	+	+	+	+	+	+
CF5	Lawley	017	7	+	-	-	-	-	+	+	+	+	+	+	-	-
19119	Citron	019	7	+	-	-	-	-	+	+	+	+	+	+	-	-
19121	Citron	019	7	+	-	-	-	-	+	+	+	+	+	+	-	-
19134	Citron	137	7	-	-	-	-	-	+	+	+	+	-	-	-	-
Ox1145 (ST4)	Dingle	137	7	-	-	-	-	-	+	+	+	+	-	-	-	-
Totals (Killed/Tested)		137	7	4/6	0/7	0/4	0/7	0/6	7/7	7/7	7/7	7/7	0/4	0/4	0/3	0/4
Ox1396 (ST54)	Dingle	012	8	-	+	-	+	-	+	+	+	+	-	-	+	-
Totals (Killed/Tested)			8	1/5	1/1	0/1	1/1	0/1	1/1	1/1	1/1	1/1	0/1	0/1	1/1	0/1
19099	Citron	002	9	-	-		+	-	-	-	-	-	+	+	-	-
19131	Citron	002	9	-	-		+	-	-	-	-	-	+	+	-	-
TL178	Lawley	002	9	-	-		+	-	-	-	-	-	+	+	-	-
Ox1192c (ST8)	Dingle	002	9	+	-	-	+	-	-	-	-	-	+	+	-	-
TL174	Lawley	015	9	-	-		+	-	-	-	-	-	-	-	-	-
19145	Citron	153	9	-	-		+	-	-	-	-	-	-	-	-	-
Totals (Killed/Tested)			9	1/5	0/4	0/1	6/6	0/5	0/6	0/6	0/4	0/4	4/4	4/4	0/2	0/2
TL176	Lawley	014	10	-	-		-	-	-	-	-	-	+	+	-	
19102	Citron	106	10	-			-	-	-	-	-	-	+	+		
LIV22	Lawley	106	10	-			-	-	-	-	-	-	+	+		
Ox1533 (ST13)	Dingle	129	10		-		-	-	-	-	-	-	+	+	-	-
Totals (Killed/Tested)			10	0/3	0/2	0/1	0/4	0/3	0/4	0/4	0/3	4/4	0/3	0/1	0/1	0/1
CD305	Lawley	023	11	-	-	-	-		-	-	+	+	-	-	-	-
Ox1523 (ST5)	Dingle	023	11		-	-	-		-	-	+	+	-	-	-	-
19155	Citron	080	11	-	-	-	-	-	-	-	+	+	-	-	-	-
Totals (Killed/Tested)			11	0/2	0/2	0/1	0/2	0/1	0/3	0/3	3/3	3/3	0/1	0/1	0/1	0/1
19146	Citron	010	12	-	-	-	-	-	-	-	-	-	-	+	-	-
CD62	Riley	010	12	-	-	-	-	-	-	-	-	-	-	+	-	-
CD242	Riley	010	12		-	-	-	-	-	-	-	-	-	+	-	-
Ox1342 (ST15)	Dingle	012	12		-	-	-	-	-	-	-	-	-	+	-	-
F200	Young	12w	12			-	-	-	-	-	-	-	-	+	+	-
Totals (Killed/Tested)			12	0/2	0/3	0/3	0/4	0/3	0/4	0/4	0/3	4/4	0/3	5/5	0/3	0/3
19142	Citron	046	13	-	+	-	-	-	-	-	-	-	-	-	-	-
Totals (Killed/Tested)			13	0/1	1/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1

Fig.3S2: *C. difficile* **strain sensitivity patterns to Avidocin-CDs and Diffocin-4.** Preparations of each Avidocin-CD and Diffocin-4 were serially diluted and spotted on bacterial lawns for each strain. If a zone of clearance correlating with killing was detected, the strain was deemed sensitive to the agent tested and is demarked by a "+" in the figure. If no zone of clearance was observed, the strain was deemed insensitive and is demarked with "-" in the figure. Faint zones of clearance are demarked with a "w". Not all strains were tested against each Avidocin-CD preparation. The number of strains killed per the total number of strains assayed is listed for each SLCT. Totals in red indicate less than 100% sensitivity. Boxes highlighted in red indicate testing on RBP strain source. The strain designation, source, ribotype and SLCT are listed. Source designations: Dingle = Kate Dingle, University of Oxford; Riley = Tom Riley, University of Western Australia; Lawley = Trevor Lawley, Wellcome Trust Sanger Institute; Tenover = Fred Tenover, Cepheid; Young = Vincent Young, University of Michigan; R.M.A.R.
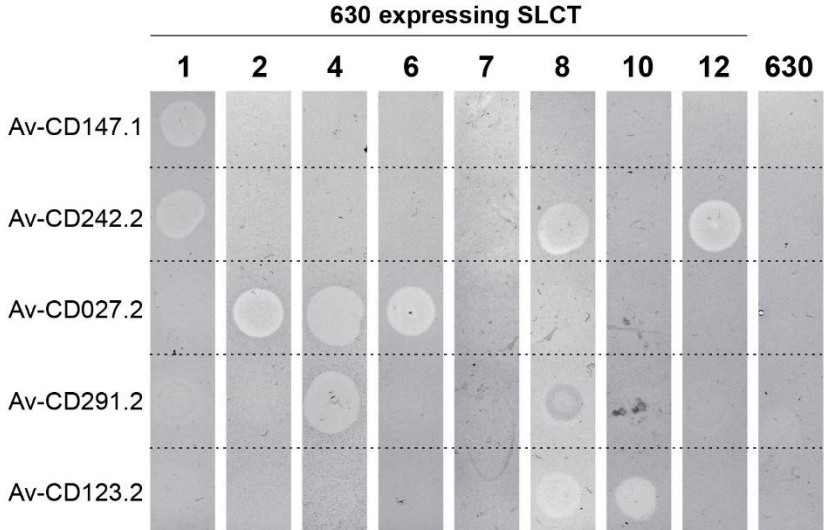= R.M. Alden Research Laboratory.

A.
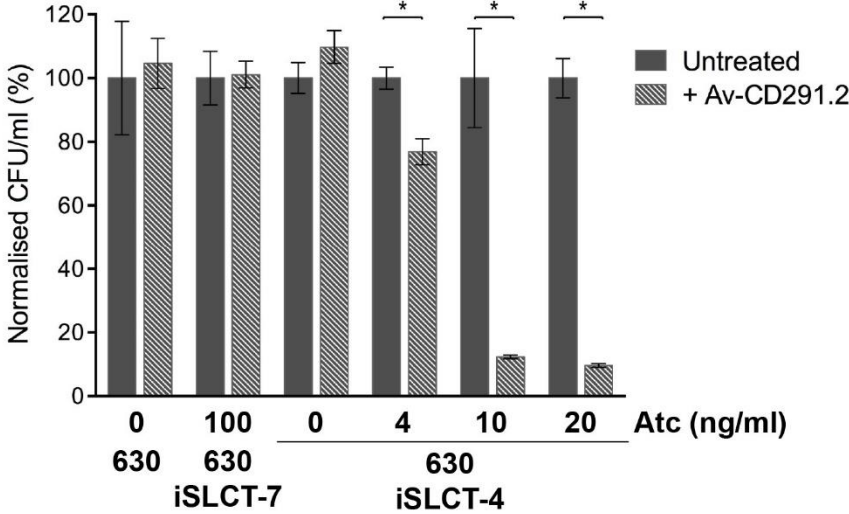
B.
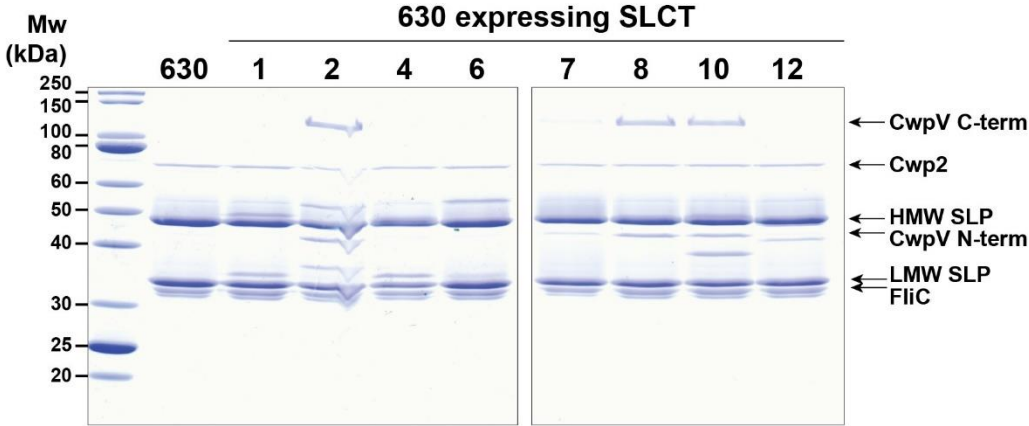
C.

Fig.3S3: Avidocin-CD sensitivity correlates with SLCT. **(A)** Av-CD291.2 killing efficiency is dependent on the level of SLCT-4 SlpA induction. Strain 630 (SLCT-7) or 630 carrying a plasmid-borne inducible copy of SLCT-7 SlpA or SLCT-4 SlpA were subcultured to an OD_{600nm} of 0.05, grown for two hours, induced with the indicated concentration of ATc and then challenged with Av-CD291.2 following two hours of induction. Identical untreated cultures were included as controls. Shown are means and standard deviations of biological duplicates assayed in triplicate. p =<0.01, determined using two-tailed *t*-tests with Welch's correction. **(B)** SDS-PAGE analysis of S-layer extracts from 630 (SLCT-7) and a panel of eight 630 derivative strains expressing SLCT-1, 2, 4, 6, 7, 8, 10 and 12 under the control of a tetracycline-inducible promoter (*40*). Heterologous SlpA expression was induced with 20 ng/ml anhydrotetracycline. The positions of the LMW and HMW SLPs, the minor cell wall proteins CwpV and Cwp2, and the flagellar subunit FliC are indicated. **(C)** Spot bioassays with 5 Avidocin-CDs on the *C. difficile* strains used in panel B following induction with anhydrotetracycline (20 ng/ml). The zone of clearance caused by each Avidocin-CD is shown along with the SLCT.
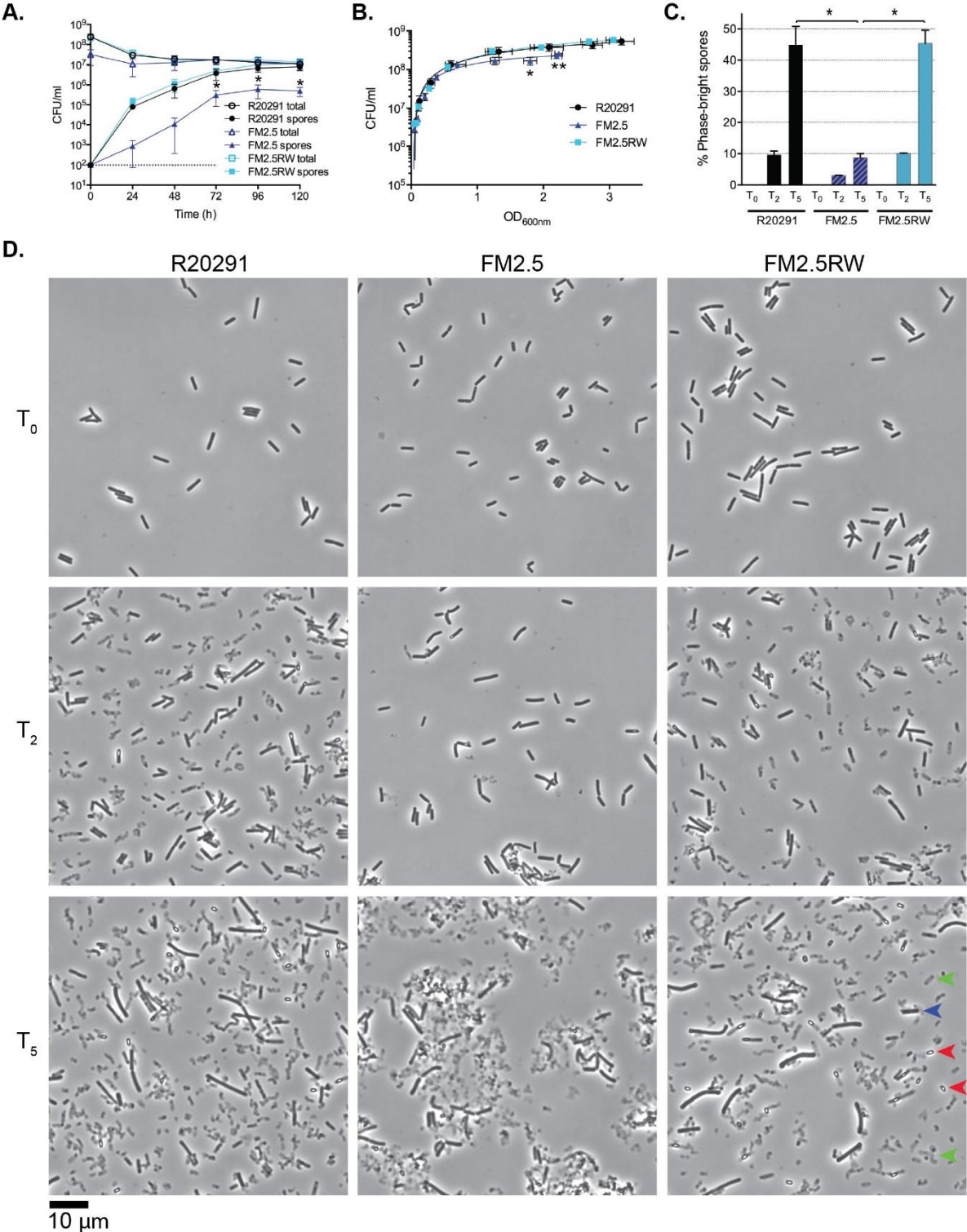
A.

B.

C.

D.

Fig.3S4: Characterization of growth and sporulation. (A) Sporulation of R20291, FM2.5, and FM2.5RW over 5 days. Total CFUs and heat-resistant spore CFUs (65°C for 30 min) were determined at 24 h intervals. Experiments were carried out in duplicate on biological duplicates. Mean and standard deviation are shown. The dotted line indicates the limit of detection. **(B)** Overnight cultures of strains R20291, FM2.5, and FM2.5RW were subcultured to an OD_{600nm} of 0.05 and growth was monitored hourly for 9 hours by measuring the OD_{600nm} and direct counting of CFUs. Shown are the means and standard deviations of biological duplicates with technical triplicates. Growth of all 3 strains is identical until an OD_{600nm} of 1.25. FM2.5 prematurely enters stationary phase, reaching a maximum OD_{600nm} of 2.2 compared with 3.2 for R20291. **(C)** Sporulation of R20291, FM2.5, and FM2.5RW was monitored by direct counting vegetative cells and phase bright spores by phase contrast microscopy at T=0, 48 and 120 of a standard 5 day sporulation assay. 20 microscope fields were counted for each strain at each timepoint. The number of spores is expressed as a percentage of total with the mean and standard deviations shown. * = p<0.01, determined using using two-tailed *t*-tests with Welch's correction. **(D)** Example microscope fields representative of those counted for panel C. For clarity only 40% of each field is shown. In the bottom right image representative spores are marked with red arrow heads, a representative vegetative cell with a blue arrow head and representative cell debris with green arrow heads.

A.
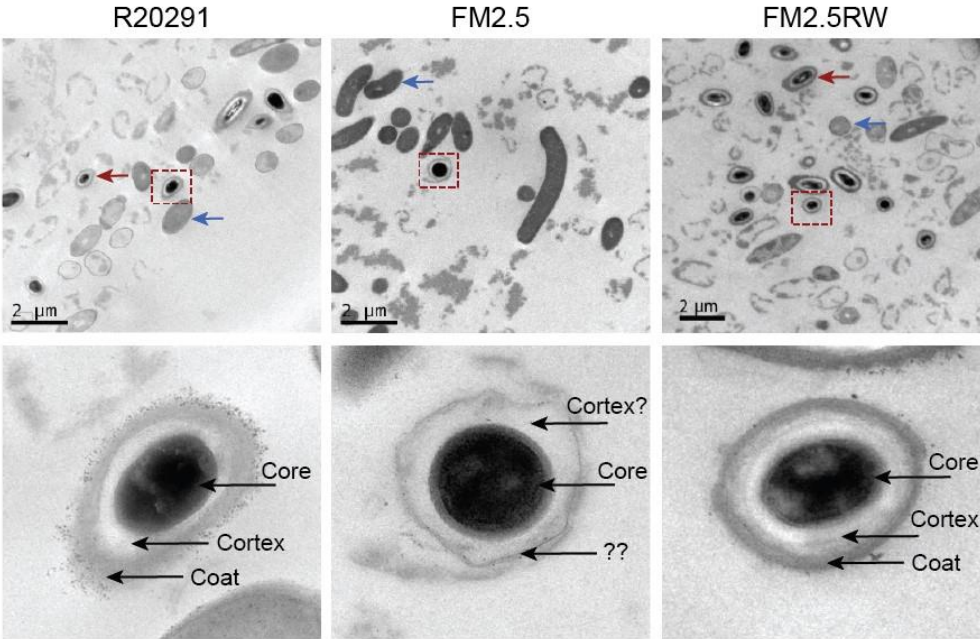
B.
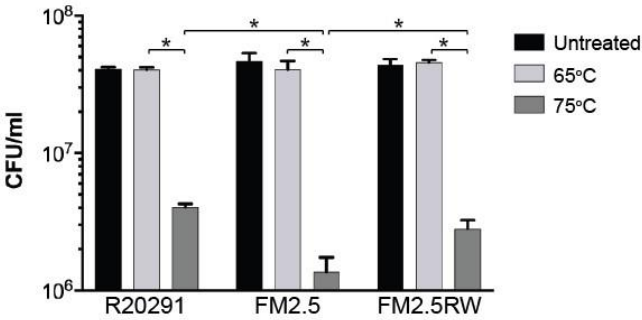
C.
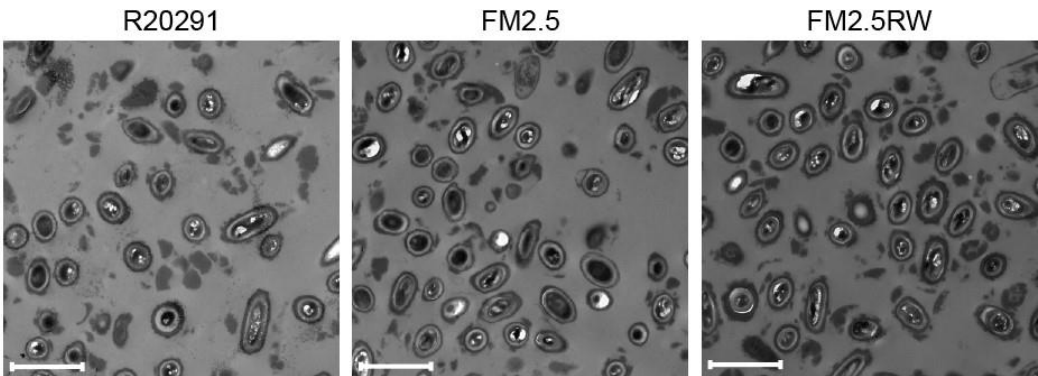
Fig.3S5: Spore morphology and thermal sensitivity. (A) Electron micrographs of cultures on day 5 of the sporulation assay shown in Figure 3S4A. Spore-containing representative fields are shown. Example spores (red arrows) and vegetative cells (blue arrows) are indicated. Higher magnification views of individual spores are shown below (boxed in top panels). The core, cortex and coat layers are indicated. **(B)** Spores were purified on a 20-50% Histodenz gradient and CFUs were quantified prior to (Untreated) and following heat stress at 65°C or 75°C for 30 min. Shown are the mean and standard errors of biological triplicates assayed in triplicate. * = p<0.01, determined using using two-tailed *t*-tests with Welch's correction. **(C)** Electron micrographs of spores used in panel B. The scale bar on each image indicates 2 μm.

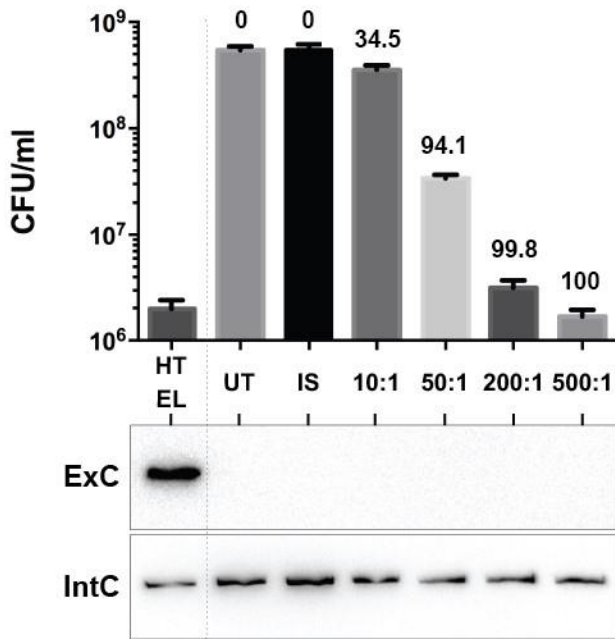**Fig.3S6: Bactericidal activity by Avidocin-CD does not result in intracellular toxin release.** For toxin producing pathogens, antibiotics that lyse bacteria or increase toxin production can exacerbate disease severity and lessen the effectiveness of the treatment by inadvertently releasing toxins (*41*). We examined culture supernatants following treatment with Avidocin-CDs for the presence of extracellular Toxin B to determine if killing by an Avidocin-CD could result in unwanted toxin release. R20291 (RT027) bacteria were incubated for an hour with either Av-CD291.2 at the indicated ratio of agent to cells, Av-CD684.1 (which does not kill strain R20291) at a 500:1 ratio, or left untreated. After incubation, surviving bacteria (CFU/mL) were enumerated. The percentage of CFU following each treatment relative to untreated control is shown. The number of spores present in the untreated sample was determined following heat treatment at 65 °C for 30 min to kill vegetative cells (lane HK). Av-CD291.2 killing occurred in a dose-dependent manner, while Av-CD684.1 treated cells behaved like untreated controls. Following Avidocin-CD treatment released toxin B in culture supernatants (ExC) was detected by Western immunoblot using a specific monoclonal antibody. As a positive control for toxin release, R20291 was treated with a bacteriophage (CD27L) endolysin, which effectively lyses *C. difficile* (*42*). Following Avidocin-CD treatment, the amount of remaining intracellular toxin B was determined by lysing cells with CD27L endolysin and detection by Western immunoblot as before (IntC). A fresh sample of untreated R20291 was lysed with CD27L endolysin to show normal intracellular toxin quantities (lane EL). Despite effective killing, intracellular Toxin B remained constant with no detectable release observed in the culture supernatants (ExC) at any concentration, including samples that exhibited total killing of all vegetative bacterial cells due to a large excess (500:1) of Avidocin-CD to bacteria. These results confirm that Avidocin-CDs are bactericidal but neither lyse the target cell nor release harmful intracellular stores of toxin B. As a result, the use of Avidocin-CD to treat or prevent CDI is unlikely to exacerbate disease symptoms.
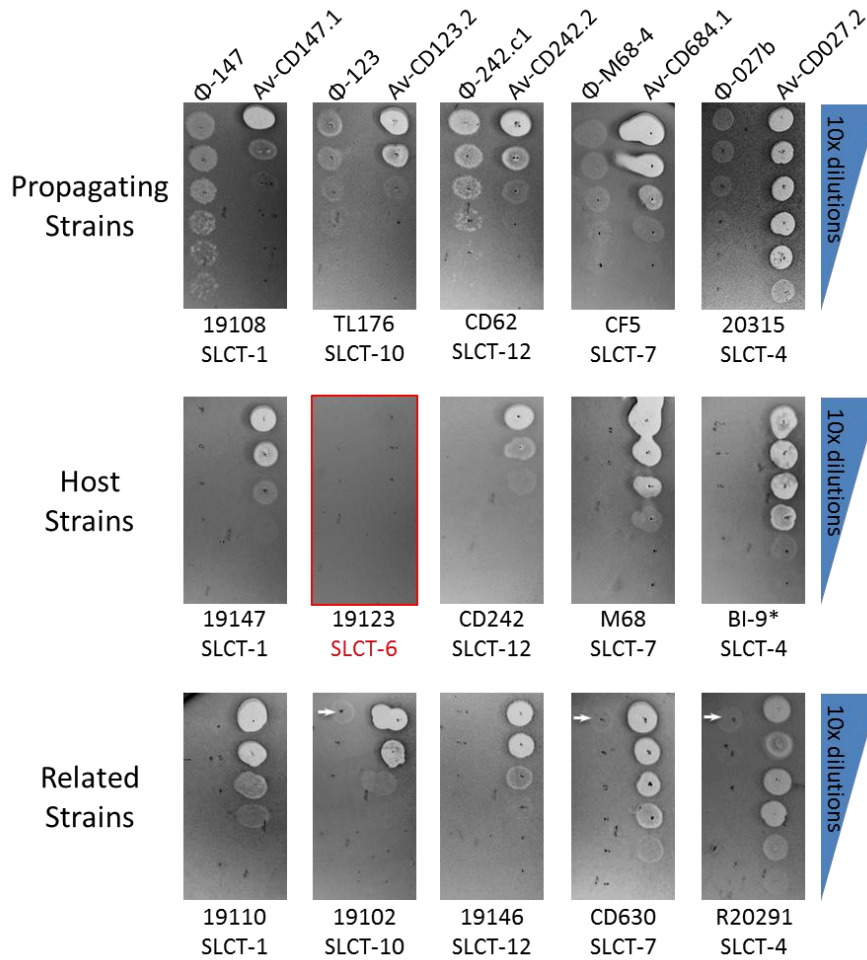
Fig.3S7: Comparison of *C. difficile* bacteriophage host range vs. Avidocin-CD sensitivity. Preparations of each bacteriophage and corresponding Avidocin-CD were serially diluted and spotted on bacterial lawns corresponding to the bacteriophage propagating strains, bacteriophage host strains and strains with same SLCT as propagating strain. Bacteriophage plaques are small and numerous. Avidocin-CD killing results in a large zone of clearance. Only one host strain, denoted in red, had a different SLCT than propagating strain. At high bacteriophage titers, bacteriophage are observed to cause "lysis from without" and is denoted by small white arrows.

```
SlpA-CD630    MNKKNIAIAMSGLTVLASAAPVFAATTGTQGYTVVKNDWKKAVKQLQDGLKDNSIGKITV    60
SlpA_M68      MNKKNLAMAMAAVTVVGSAAPIFADS-TTPGYTVVKNDWKKAVKQLQDGLKNKTISTIKV    59
              *****:*:**:.:**:.****:**  :   * *********************::*..*.*

SlpA-CD630    SFNDGVVGEVAPKSA-NKKADRDAAAEKLYNLVNTQLDKLGDGDYVDFSVDYNLENKIIT    119
SlpA_M68      SFNGNSVGEVTPASSGAKKADRDAAAEKLYNLVNTQLDKLGDGDYVDFEVTYNLATQIIT    119
              ***    ****:* *:  ****************************.* *** .:*** 

SlpA-CD630    NQADAEAIVTKLNSLNEKTLIDIATKDTFGMVSKTQDSEGKNVAATKALKVKDVATFGLK    179
SlpA_M68      K-AEAEAVLTKLQQYNDKVLINSATDTVKGMVSDTQVDSK-N-VAANPLKVSDMYTIPSA    176
              : *:***::***:. *:*.**: **. . ****.** ..  * .*:: ***.*: *:

SlpA-CD630    SGGSEDTGYVVEMKA-GAVEDKYGKVGDSTAGIAINLPST-GLEYAGKGTTIDFNKTLKV    237
SlpA_M68      ITGSDDSGYSIAKPTEKTTSLLYGTVGDATAGKAITVDTASNEAFAGNGKVIDYNKSFKA    236
               **:*:** :   :  :.. **.***:*** **.: ::    :**:*..**:**::*.

SlpA-CD630    DVTGGSTPSAVAVSGFVTKDDTDLAKSGTINVRVINAKEESIDIDASSYTSAENLAKRYV    297
SlpA_M68      TVQGDGT---VKTSGVVLKDASDMAATGTIKVRVTSAKEESIDVDSSSYISAENLAKKYV    293
              * * .*    * .**.* ** :*:* :***:*** .*******:*:*** *******:**

SlpA-CD630    FDPDEISEAYKAIVALQNDGIESNLVQLVNGKYQVIFYPEGKRLETKSANDTIASQDTPA    357
SlpA_M68      FNPKEVSEAYNAIVALQNDGIESDLVQLVNGKYQVIFYPEGKRLETKSA-DIIADADSPA    352
              *:*.*:****:************:***********************:  * **. *:**

SlpA-CD630    KVVIKANKLKDLKDYVDDLKTYNNTYSNVVTVAGEDRIETAIELSSKYYNSDDKNAITDK    417
SlpA_M68      KITIKANKLKDLKDYVDDLKTYNNTYSNVVTVAGEDRIETAIELSSKYYNSDDKNAITDD    412
              *:.*********************************************************.

SlpA-CD630    AVNDIVLVGSTSIVDGLVASPLASEKTAPLLLTSKDKLDSSVKSEIKRVMNLKSDTGINT    477
SlpA_M68      AVNNIVLVGSTSIVDGLVASPLASEKTAPLLLTSKDKLDSSVKSEIKRVMNLKSDTGINT    472
              ***:********************************************************

SlpA-CD630    SKKVYLAGGVNSISKDVENELKNMGLKVTRLSGEDRYETSLAIADEIGLDNDKAFVVGGT    537
SlpA_M68      SKKVYLAGGVNSISKDVENELKNMGLKVTRLSGEDRYETSLAIADEIGLDNDKAFVVGGT    532
              ************************************************************

SlpA-CD630    GLADAMSIAPVASQLKDGDATPIVVVDGKAKEISDDAKSFLGTSDVDIIGGKNSVSKEIE    597
SlpA_M68      GLADAMSIAPVASQLKDGDATPIVVVDGKAKEISDDAKSFLGTSDVDIIGGKNSVSKEIE    592
              ************************************************************

SlpA-CD630    ESIDSATGKTPDRISGDDRQATNAEVLKEDDYFTDGEVVNYFVAKDGSTKEDQLVDALAA    657
SlpA_M68      ESIDSATGKTPDRISGDDRQATNAEVLKEDDYFKDGEVVNYFVAKDGSTKEDQLVDALAA    652
              *********************************.**************************

SlpA-CD630    APIAGRFKESPAPIILATDTLSSDQNVAVSKAVPKDGGTNLVQVGKGIASSVINKMKDLL    717
SlpA_M68      APIAGRFKESPAPIILATDTLSSDQNVAVSKAVPKDGGTNLVQVGKGIASSVINKMKDLL    712
              ************************************************************

SlpA-CD630    DM      719
SlpA_M68      DM      714
              **
```

Fig.3S8: Clustal Omega alignment of SlpA sequences from strains 630 and M68. Identical residues are demarked with an asterisk. Residues with strongly similar properties are demarked with colon. Residues with weakly similar properties are demarked with a period. An 81.7% sequence identity was found between the two SLCT-7 SlpA variants.
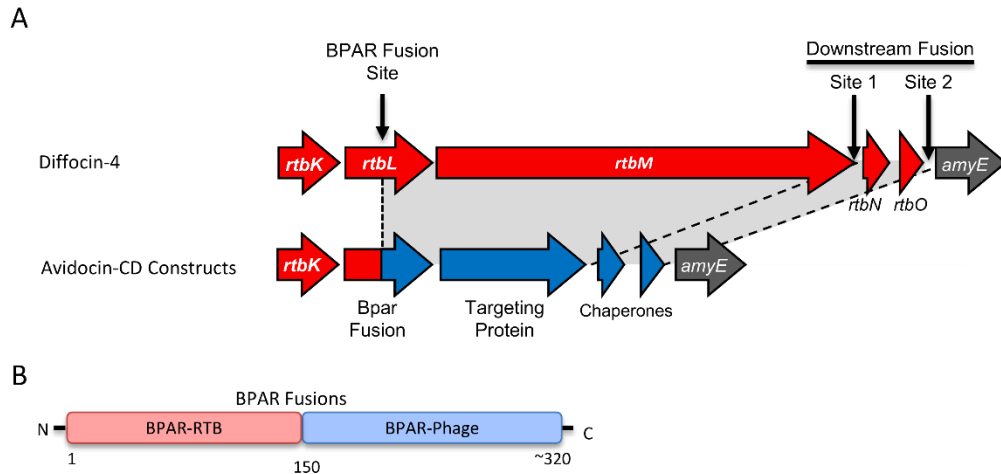
Fig.3S9: Schematic describing Avidocin-CD construction. (**A**) The fusion sites used to integrate the bacteriophage and prophage genes into the Diffocin-4 gene cluster are shown. There is a Bpar fusion site in *rtbL* as well as alternative downstream fusion sites before or after the predicted Diffocin-4 chaperones. The nucleotide sequences for the fusion region created in each Avidocin-CD construct fusion have been deposited in Genbank and given the following accession numbers (Table 3S3). (**B**) Schematic depicting R-type Bacteriocin and Bacteriophage regions for fused BPAR protein.

Table 3S1: Newly identified *C. difficile* phages

Phage	Source	Propagating Strain	PtsM-like RBP
Phi-027b	panel of 027 strains	20315 (SLCT-4)	Yes
Phi-123	19123 (SLCT-6)	TL176 (SLCT-10)	No
Phi-147	19147 (SLCT-1)	19108 (SLCT-1)	Yes
Phi-242.c1	CD242 (SLCT-12)	CD62 (SLCT-12)	Yes
Phi-68.4	M68 (SLCT-7)	CF-5 (SLCT-7)	Yes

Table 3S2: Primers used in this chapter

Primer name	Primer sequence	Use
023-F	ATGAAAAAAAGAAATTTAGCA	Amplification of *slpA* variable region
023_010-R	CTATAGCTGTTTSTATTCTGTC	Amplification of *slpA* variable region
014+++-F	ATGAATAAGAAAARTTTRGCA	Amplification of *slpA* variable region
014_002+-R	CTATWGCAGTCTCTATTCTATC	Amplification of *slpA* variable region
RF1	GATCGGATCCTTACATACTTAATAAATC TTTTAATTTATTTATAACTG	Cloning of SLCT-4 *slpA* from strain R20291 with RF2
RF2	GATCGAGCTCTATAATGTTGGGAGGAAT TTAAGAAATG	Cloning of SLCT-4 *slpA* from strain R20291 with RF1
RF6	GATCGGATCCTTACATATCTAATAAATC TTTTATTTTACTTACAAC	Cloning of SLCT-H2/6 *slpA* from strain Ox575 with RF2

RF7	GATCGGATCCCTACATATCTAATAAGTC TTTTATCTTAGTG	Cloning of SLCT-10 *slpA* from strain Ox1533 with RF2
RF8	GATCGGATCCTTACATATCTAATAAATC TTTTAATTTGCTAACAAC	Cloning of SLCT-1 *slpA* from strain Ox1424 with RF2
RF33	GGCTTCTATAGCTTGGTGAA	Confirmation of *slpA* recombination with NF1323. Chromosome-specific
RF99	GATCGGATCCTTACATATCTAATAAATC TTTTATTTTATTTATTACTG	Cloning of SLCT-6 *slpA* from strain Ox1896 with RF2
RF101	GATCGGATCCTTATAATTCTAATAAATC TTTTATTCTCTTAATAAC	Cloning of SLCT-9 *slpA* from strain Ox1192c with RF134
RF110	GACATAACTGCAGCACTACTTG	Amplification of SLCT-4 *slpA* region containing point mutations and watermark
RF111	CAGGATTAACAGTATTAGCTTCTGC	Amplification of SLCT-4 *slpA* region containing point mutations and watermark
RF132	GATCGGATCCTTACATATTTAATAAATC TTTAAGCTTAGTTAC	Cloning of SLCT-8 *slpA* from strain Ox1396 with RF2
RF133	GATCGGATCCTTACATTCCTAATAAATC TTTTAATTTATTAATAAC	Cloning of SLCT-12 *slpA* from strain Ox1342 with RF134
RF134	GATCGAGCTCTATAATGTTGGGAGGAAT TTAAGGAATG	Cloning of SLCT-9 and 12 *slpA* with RF101 and RF133 respectively

RF135	CAGGTTGATAATAAATTAGACAATTTAG GTGATGG	Introduction of *slpA* watermark with RF136
RF136	TGTAAACAATAATTTACTTGCATCTTCT G	Introduction of *slpA* watermark with RF135
RF231	GATCGGATCCTTACATATCTAATAAATC TTTCATTTTGCTTATTAC	Cloning of SLCT-2 *slpA* from strain Ox858 with RF2
NF1323	CTGGACTTCATGAAAAACTAAAAAAAA TATTG	Confirmation of *slpA* recombination with RF33. Plasmid-specific
NF1414	GATCGAGCTCTATAATGTTGGGAGGAAT TTAAGAAATG	Cloning of SLCT-7 *slpA* from strain 630 with NF1415
NF1415	GATCGGATCCTTACATATCTAATAAATC TTTCATTTTG	Cloning of SLCT-7 *slpA* from strain 630 with NF1414

Table 3S3: Genbank accession identifiers

Construct	Source	Accession Id:
Diffocin-4	CD4	KX557294
Av-CD291.2	R20291 genome	KX592438
Av-CD027.2	Phi-027b	KX592434
Av-CD681.1	M68 genome	KX592441
Av-CD682.1	M68 genome	KX592442
Av-CD684.1	M68 genome/phi-M68.4	KX592443
Av-CD685.1	M68 genome	KX592444
Av-CD305.1	CD305 genome	KX592439
Av-CD123.2	Phi-123	KX592435

Av-CD242.2	Phi-242.c1	KX592437
Av-CD147.1	Phi-147	KX592436
Av-CD630.1	630 genome	KX592440

Chapter IV

Characterisation of FM2.5 and the Roles of the S-layer

4.1 Introduction

The work described in the previous chapter (Kirk and Fagan, 2016) focused on phenotypic characterisation of *C. difficile* mutants that have lost *slpA* function, specifically to investigate the clinical threat of these Avidocin-CD resistant mutants. Further characterisation of these mutants, was required to understand the functions of the S-layer and the requirements for *C. difficile* growth when lacking this structure.

The largest member of the Cell Wall Protein family, CwpV, was overexpressed in the *slpA* deficient mutants. Approximately 5% of cells within a population of *C. difficile* express CwpV, when grown under standard laboratory conditions (Emerson et al., 2009). CwpV expression is under control of a phase variable promoter, with expression being either ON or OFF, depending on the orientation of a genetic switch upstream of the ORF. The 195 bp genetic switch, flanked by 21 bp inverted repeats, is inverted through site-specific recombination, mediated by the recombinase RecV. When this switch region is in the OFF orientation, a rho-independent terminator is predicted to form, resulting in premature termination of transcription (Emerson et al., 2009). CwpV is an auto-aggregation protein, and its expression is associated with a change in colony morphology, bacterial aggregation in liquid broth, and phage resistance (Reynolds et al., 2011; Sekulovic et al., 2015). CwpV has also been identified as a SecA2 substrate (Fagan and Fairweather, 2011) and undergoes auto-proteolysis post-secretion (Dembek et al., 2012).

Phase variation is a common mechanism for regulation of extracellular structures, including but not limited to, pili, cellular capsules, and flagella. This can aid in the evasion of the

host immune system and regulation of adhesion to the host (as reviewed in (van der Woude and Baumler, 2004)). Phase variable expression is not limited to extracellular proteins however. For example, the *flgB* operon of *C. difficile* is under the control of a RecV-dependent phase variable promoter. This operon contains the gene encoding the regulatory protein SigD which, in turn, controls the expression of flagella proteins, the toxins, other regulatory proteins, and proteins involved in metabolism (El Meouche et al., 2013). When the *flgB* switch is in the OFF orientation, cells are attenuated for toxin production (Anjuwon-Foster and Tamayo, 2017). Phase variation of regulatory proteins results in coordinated phase variable expression of genes under control of these proteins. Phase variation therefore has the potential to have polar effects on genes in multiple genetic pathways. To date, four phase variable regions have been described in the *C. difficile* strain R20291 genome, upstream of *cwpV, flgB, CDR20291_1514,* and, *CDR20291_0685* (Anjuwon-Foster and Tamayo, 2017; Stabler et al., 2010). Excluding *CDR20291_0685,* these phase variable regions are flanked by inverted repeats, a characteristic property of site-specific recombination-dependent phase variable regions.

Work presented in this chapter focuses on characterisation of the effects of overexpression of CwpV in FM2.5. In addition, genomic and epigenetic difference between FM2.5 and R20291 were investigated, leading to the identification of an addition phase variable promoter and potential suppressor mutations in FM2.5. Characterisation and modification of the native *slpA* locus is necessary for more in depth analysis of S-layer function. Attempts to modify this locus and identify the *slpA* promoter are also described.

4.2 Results

4.2.1 CwpV Expression in FM2.5

Although CwpV is normally only expressed in approximately 5% of wild-type cells in a population (Emerson et al., 2009), CwpV is more frequently found in S-layer extracts from FM2.5 (Fig.4.1A). This was also observed for the independent mutant FM2.6 (data not shown). Due to this observation, the effects of loss of *slpA* function on CwpV expression, and of CwpV expression in FM2.5 were investigated.

4.2.1.1 CwpV is Expressed More Frequently in FM2.5

As CwpV expression is phase variable, it was hypothesised that increased CwpV expression in FM2.5 may be due to an increased frequency of cells in a population containing the *cwpV* switch in the ON orientation. To examine the frequency of CwpV expression in FM2.5 and to confirm if this was caused by loss of *slpA* function, quantification of CwpV expression in FM2.5 and FM2.5RW was necessary.

gDNA was purified from O/N cultures of eight biological replicates of R20291, FM2.5, and FM2.5RW. The orientation of the *cwpV* switch was determined by asymmetric restriction digest analysis of PCR amplified DNA fragments (Fig.4.1C and D). There was no statistically significant difference between the recorded frequencies of promoter orientation between R20291 and FM2.5RW (median of 0% ON). However, there was a significant difference between R20291 and FM2.5 (median of 92% ON, $P < 0.005$). These data confirm the previous observation that *cwpV* expression is upregulated in FM2.5, and this upregulation is dependent on loss of *slpA* function.
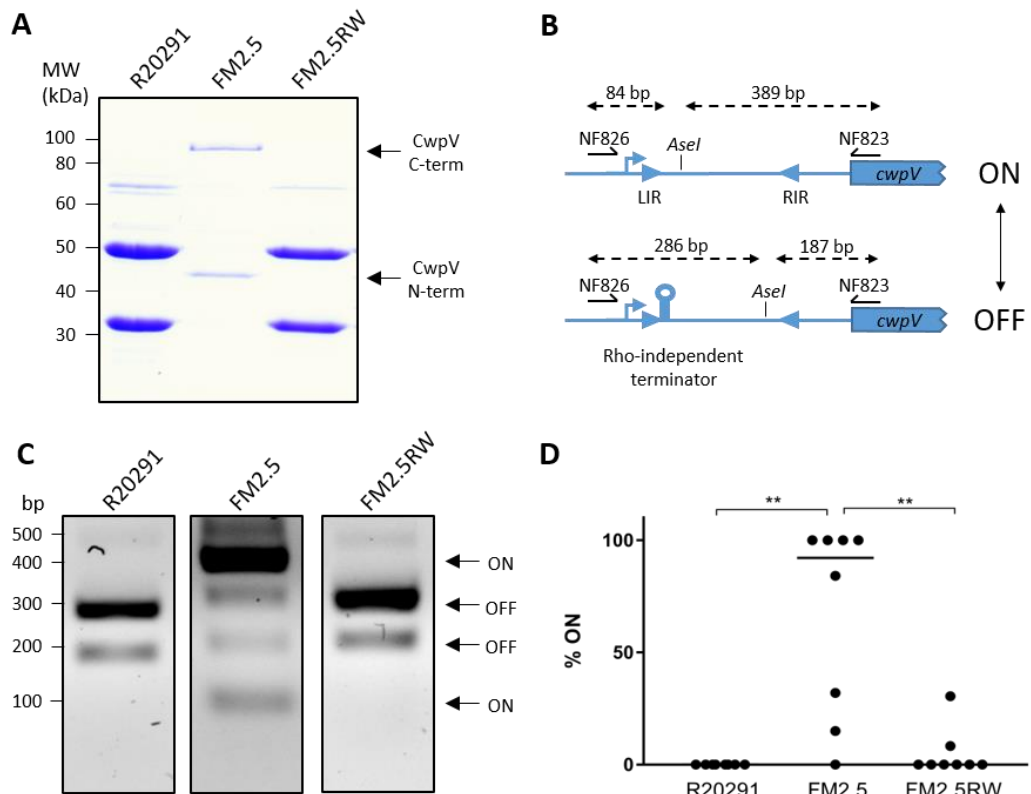
Fig.4.1: Analysis of CwpV expression in FM2.5. (A) Representative 12% SDS-PAGE showing S-layer extracts from R20291, FM2.5 and FM2.5RW. CwpV is cleaved post-secretion and the N-terminal and C-terminal subunits are labelled. CwpV is present in S-layer extracts isolated from FM2.5 more frequently than those from R20291 or FM2.5RW. **(B)** Schematic representation of the promoter region of *cwpV* in the ON and OFF orientations. Expression of CwpV is phase variable due to the presence of inverted repeats (LIR and RIR) flanking an invertable region downstream of the promoter. A Rho-independent terminator is predicted to form when the promoter region is in the OFF orientation leading to premature transcriptional termination (Emerson et al., 2009). Primers NF863 and NF823 flank the invertable region, which contains an *Ase*I restriction site. *Ase*I digestion of DNA amplified using these primers results in DNA fragments of different sizes depending on the orientation of the switch region. **(C)** 2% agarose gels showing representative *Ase*I RFLP analysis of DNA fragments amplified from gDNA isolated from R20291, FM2.5, and FM2.5RW. This results in 84 bp and 389 bp fragments when in the "ON" orientation, and 187 bp and 286 bp fragments when in the "OFF" orientation. **(D)** Quantitative analysis of the percentage of bacteria containing the *cwpV* switch in the ON orientation from populations of R20291, FM2.5 and FM2.5RW. This was calculated via densitometric analysis of relevant bands from agarose gels containing *Ase*I digested DNA. Asymmetric digest analysis was performed using eight biological replicates. Each circle on the graph represents one biological repeat and the medians are shown. ** *P* < 0.005 determined using non-parametric, two tailed Mann-Whitney U tests.

4.2.1.2 Generation of cwpV Locked ON and OFF Strains

To test the effects of CwpV expression in FM2.5, it was first necessary to lock the *cwpV* switch in both the ON and OFF orientations. This was achieved through inactivation of the gene encoding the recombinase responsible for inversion of the *cwpV* switch, *recV* (Emerson et al., 2009). *recV* mutants, created in FM2.5 and FM2.5RW using ClosTron mutagenesis, were designated FM2.5 *recV*::*erm* and FM2.5RW *recV*::*erm* (Fig.4.2A).

To lock the *cwpV* switch in either the ON and OFF orientation, *recV* inactivated strains were restreaked on BHI and the orientation of the switch was determined via PCR (Fig.4.2B). This was repeated until strains where the switch region was locked in one orientation were isolated. FM2.5 *recV*::*erm* was exclusively isolated with the genetic switch in the ON orientation and FM2.5RW *recV*::*erm* was isolated with the switch in the OFF orientation. These strains were denoted as FM2.5 *recV*::*erm cwpV* ON (FM2.5 *cwpV* ON) and FM2.5RW *recV*::*erm cwpV* OFF (FM2.5RW *cwpV* OFF). Isolation of *recV* inactivated mutants with the *cwpV* switch in the opposite orientations was first attempted using the previously described method (Reynolds et al., 2011). RecV was transiently expressed in the mutant strains to complement the phase switching defect. However, we failed to isolate mutants with the switch in the opposite orientations. pMTLSC7215 vectors were then created in which the *cwpV* switch, flanked by sequences of homology to the chromosome, were in the OFF and ON orientations. These plasmids were used to performed allele exchange mutagenesis in the *recV* inactivated strains (Fig.5.2C). This method resulted in the isolation of FM2.5 *recV::erm cwpV* OFF (FM2.5 *cwpV* OFF) and FM2.5RW *recV::erm cwpV* ON (FM2.5RW *cwpV* ON).

R20291 *recV::erm* mutants (R20291 *cwpV* ON and R20291 *cwpV* OFF) were provided by Neil Fairweather (Imperial College London). All six strains displayed the expected pattern of CwpV expression (Fig.4.2D).
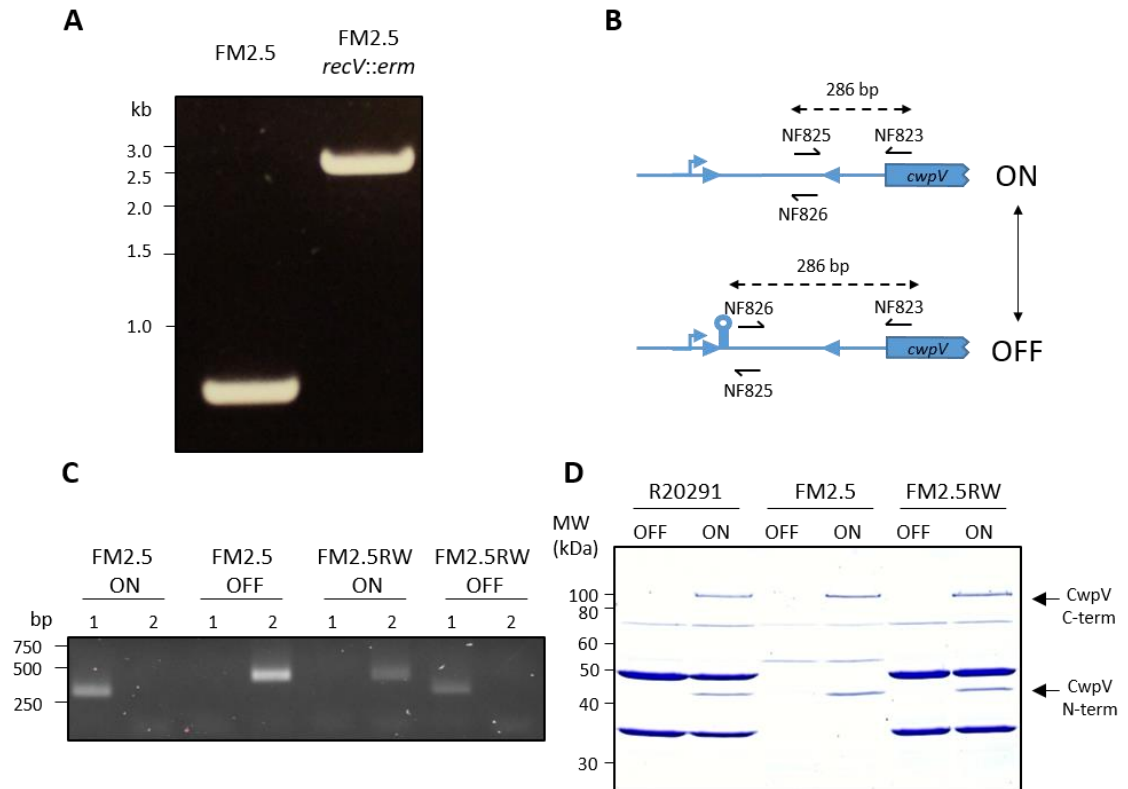
Fig.4.2: Creation of *cwpV* ON and OFF strains. (A) *recV* was inactivated in FM2.5 via ClosTron mutagenesis. pASF075 was conjugated into FM2.5 and FM2.5RW and transconjugants were restreaked onto BHI agar containing 5 μg/ml lincomycin to select for group II intron insertion. Potential mutants were then screened via PCR using primers NF1215 and RF438, which flank the intron insertion site. An increase of approximately 2.8 kb relative to FM2.5 confirms intron insertion. **(B)** The orientation of the *cwpV* switch was determined via PCR using primer pairs NF825/NF823 and NF826/NF823. FM2.5 *recV::erm cwpV* ON and FM2.5RW *recV::erm cwpV* OFF were isolated. To create mutants with the switch in the opposite orientations allele exchange mutagenesis was performed. Primer pair RF798/RF799 were used to PCR amplify a 2.4 kb fragment from FM2.5RW *cwpV* OFF and FM2.5 *cwpV* ON gDNA. These PCR fragments contained the *cwpV* switch, in the OFF and ON orientation respectively, and 1.2 kb upstream and downstream of the switch. These fragments were ligated into pMTLSC7215 via the *Pme*I restriction site, creating pJAK073 and pJAK074 respectively. pJAK073 was conjugated into FM2.5 *cwpV* ON and pJAK074 was conjugated into FM2.5RW *cwpV* OFF. Mutagenesis was performed as previously described (Cartman et al., 2012). **(C)** FM2.5 *recV::erm cwpV* OFF (FM2.5 *cwpV* OFF) and FM2.5RW *cwpV* ON (FM2.5RW ON) were selected via PCR screening using NF823/NF825 (Lanes '1') and NF823/NF826 (lanes '2'). **(D)** S-layer extracts were isolated from overnight cultures of *cwpV* ON and OFF strains and analysed by SDS-PAGE. The two subunits of CwpV are labelled.

4.2.1.3 cwpV Expression Has No Effect on Growth of FM2.5

As CwpV is a paralogue of SlpA, we hypothesised that CwpV overexpression in FM2.5 may complement the growth defect of this mutant (Kirk and Fagan, 2016). This would allow CwpV expressing cells within a population of FM2.5 to grow quicker than the surrounding cells, leading to disproportionate numbers of *cwpV* ON cells within a population, consistent with our previous observation.

4.2.1.3.1 Growth of FM2.5 *cwpV* ON and OFF Strains in Liquid Media

To test this hypothesis, growth assays were performed by monitoring the optical density of growing cultures of FM2.5 *cwpV* ON and OFF (Fig.4.3A). No significant difference in growth was observed between the ON and OFF strains. Although the growth rate of these two strains was slower than the wild type, surprisingly, both the *cwpV* ON and OFF strains reached a final cell density comparable to R20291. O/N cultures of FM2.5 ON and FM2.5 OFF were also analysed via TEM of thin sections, and no discernible morphological difference was observed between the two strains (Fig.4.3B).

4.2.1.3.2 Colony Morphology of *cwpV* ON and OFF Strains

Previous studies have shown colonies of *cwpV* ON strains have smaller, smoother edged morphologies than OFF strains (Reynolds et al., 2011). Similar colony morphologies were observed in *flgB* ON and OFF strains, although not linked to the orientation of the *flgB* switch (Anjuwon-Foster and Tamayo, 2017). After 48 h growth on BHI agar, the same difference was observed between R20291 *cwpV* ON and R20291 *cwpV* OFF. No difference was observed between FM2.5 *cwpV* ON and FM2.5 *cwpV* OFF or FM2.5RW *cwpV* ON and FM2.5RW *cwpV* OFF (Fig.4.3C). The FM2.5 and FM2.5RW ON and OFF strains all formed small, smooth edged colonies. FM2.5 produced smaller colonies after 48 h growth than any of the R20291 or FM2.5RW strains.
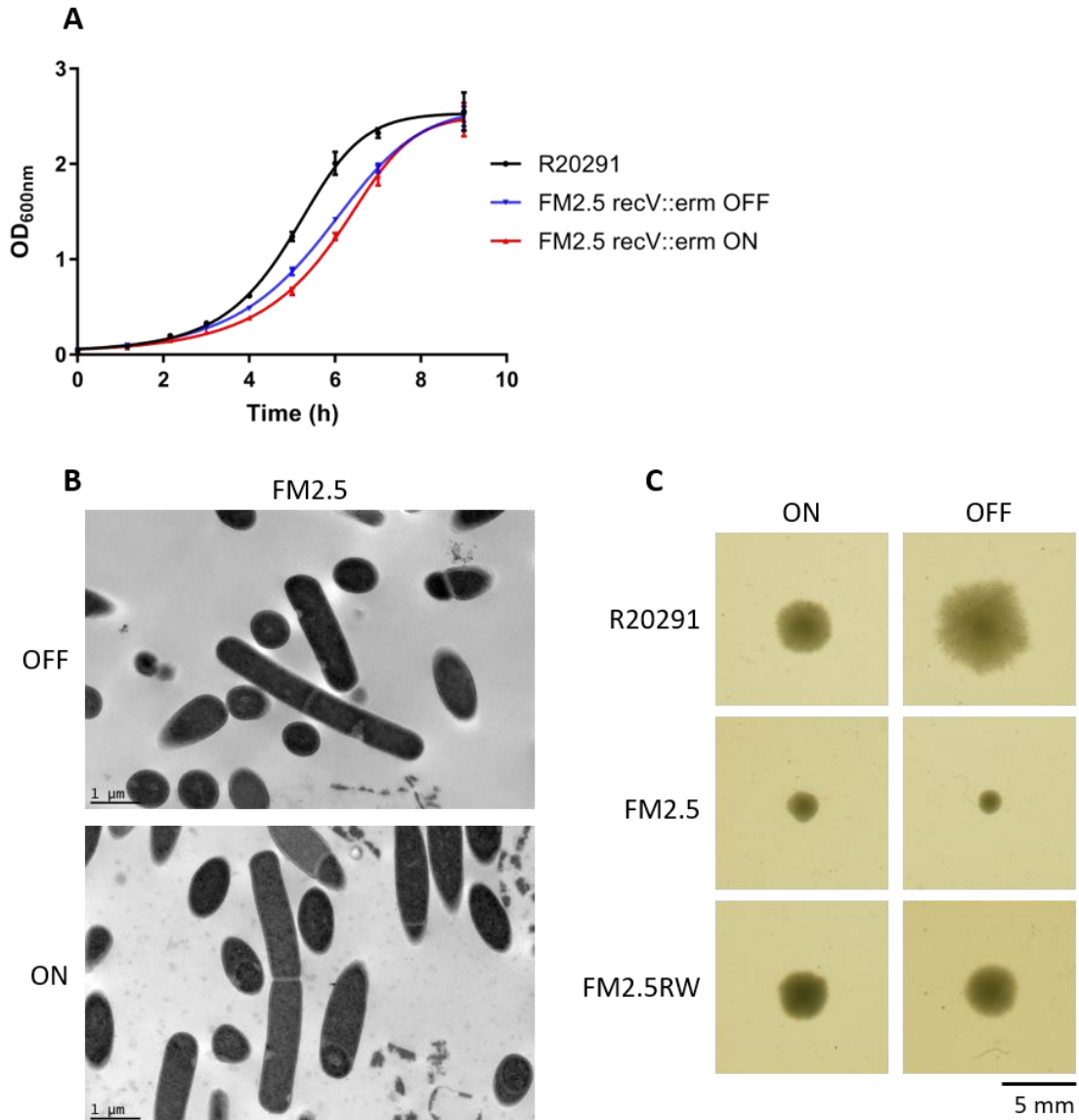
Fig.4.3: Growth of *cwpV* locked ON and OFF strains: (A) Growth assays were performed by monitoring OD$_{600nm}$ of growing cultures of FM2.5 *cwpV* ON and OFF. Both strains showed similar growth rates and reached the same final cell density which, interestingly, was comparable to the final cell density of growing cultures of R20291. These assays were performed in technical triplicates using biological duplicates. Shown are the means and standard deviations. **(B)** TEM analysis of O/N cultures of FM2.5 *cwpV* ON and FM2.5 *cwpV* OFF revealed no morphological difference between these two strains. **(C)** Colony morphologies of *cwpV* ON and OFF strains after 48 h growth on BHI agar. Smaller, smooth edged colonies of R20291 *cwpV* ON were observed, compared to R20291 *cwpV* OFF as previously reported (Reynolds et al., 2011). No morphological difference was observed between colonies of FM2.5RW *cwpV* ON and FM2.5RW *cwpV* OFF. This was also observed for colonies of FM2.5 *cwpV* ON and FM2.5 *cwpV* OFF, although these colonies were much smaller than colonies of R20291 and FM2.5 strains.

4.2.2 Genomic Analysis of FM2.5

4.2.2.1 SMRT Sequencing of the FM2.5 Genome

As *slpA* is essential for *C. difficile* growth, the genome of FM2.5 was sequenced to detect any suppressor mutations. This was initially done with Illumia sequencing. Analysis of the data only revealed a single SNP within the FM2.5 genome within *slpA,* as previously reported (Kirk and Fagan, 2016). However, Illumina sequencing results in short read lengths meaning that large insertions, deletions and, inversions in the genome are often overlooked. In addition, due to the relatively poor coverage of the sequencing data, other SNPs in the genome may have been missed. The genome of FM2.5 was therefore sequenced using a method resulting in longer read lengths; PacBio's single molecule real-time (SMRT) sequencing (Rhoads and Au, 2015). Sequence analysis and genome assembly was performed by Nadia Fernandes (the University of Sheffield, personal communication). Due to the large read lengths, a single contig, covering the entire genome, was assembled from the data.

The newly assembled genome was aligned to the published R20291 genome using the progressiveMauve alignment plugin for Geneious R10. Many SNPs were identified within the FM2.5 genome. Most of these were contained within stretches of adenines and thymine and were assumed to be an error with the published genome sequence, as sequencing errors occur commonly in AT rich regions using Illumina sequencing. Several other SNPs were identified in non-coding regions within the genome. However, in addition to the previously identified SNP within *slpA*, SNPs were identified within the ORF of three genes. These result in the substitution of tyrosine 146 to asparagine in CodY, isoleucine 58 to methionine in CDR20291_2781, and glutamine 138 to lysine in CDR20291_1323.

In addition, multiple insertions in the FM2.5 genome were identified. A 149 bp insertion was identified within *CD0667*, resulting in a 34 aa truncation of CDR20291_0667 and alteration of the C-terminal primary sequence of the truncated protein. Within the putative transcription

regulator, *CD1R20291_542,* an insertion of approximately 15 kb was identified, predicted to contain multiple ORFs including genes involved in iron uptake. Two insertions, of approximately 6 kb and 1.5 kb, were identified in rRNA coding regions. The significance of these mutations is yet to be explored.

It was unsurprising that the *cwpV* switch was identified in the ON orientation in the FM2.5 genome. However, two other short sequences were found in inverse orientations relative to the published genome. These regions were upstream of *CDR20291_1514* and *CDR20291_3128*. The inverse orientation of the *CDR20291_1514* region has been identified previously (Stabler et al., 2010), however the *CDR20291_3128* switch has not been previously described. BPROM analysis of the sequences upstream of *CDR20291_3128* and *CDR20291_1514*, predicted the promoters of these genes are contained within the invertable regions, which would result in gene expression being turned ON or OFF depending on switch orientation (Fig.4.4A). BPROM analysis also predicted a promoter immediately downstream of *CDR20291_1514*, suggesting the *CDR20291_1514* switch only controls the expression of CDR20291_1514, a putative c-phosphodiesterase (Bordeleau et al., 2011). *CDR20291_3128* is the first gene in an operon, along with *CDR20291_3127* and *CDR20291_3126*. CDR20291_3127 is a predicted two-component sensor histidine kinase, CDR20291_3128 and CDR20291_3126 are predicted two-component response regulators.

4.2.2.2 Characterisation of *CDR20291_1514* and *CDR20291_3128* Switch Inversion

As the *CDR20291_1514* and *CDR20291_3128* switches have never been characterised, it was first necessary to confirm these regions are indeed phase variable and to identify the recombinase(s) responsible for switching. Firstly, switching was confirmed via PCR using multiple primer pairs in which one primer annealed within and the other primer annealed outside of the invertable region (Fig.4.4A). It was observed that the *CDR20291_1514* and *CDR20291_3128* switches exist in both orientations in populations of R20291 (Fig.4.4B).

To determine if RecV is responsible for switching at the *CDR20291_1514* and *CDR20291_3128* loci, the orientation of these switches was examined in our *recV* mutants. Although the *CDR20291_1514* and *CDR20291_3128* switches were found in both orientations in populations of R20291, these switches were locked in the *recV* mutants (Fig.4.4B), confirming that RecV is responsible for inversion of the *CDR20291_1514* and *CDR20291_3128* switches.

Including the *CDR20291_1514* and *CDR20291_3128* switches, four phase variable regions have now been shown to be dependent on RecV for switching. All four phase variable switches are flanked by inverted repeats. The repeats flanking these switches differ in sequence. The *flgB* switch is flanked by the imperfect repeats 5'-AAGTT(A/G)CTAT(A/T)TTACAAAAAA-3' (Anjuwon-Foster and Tamayo, 2017), the *cwpV* switch is flanked by the imperfect repeats 5'-AAGTA(T/G)CCTTTAGAATTA(G/A)AA -3' (Emerson et al., 2009), the *CDR20291_3128* switch is flanked by the imperfect repeats 5'-TCC(C/A)TT(T/C)TCCAACAA(A/T)(T/C)GG(C/A)(T/G)ACTTTAA-3', and the *CDR20291_1514* switch is flanked by the perfect inverted repeat 5'-AAACTATCCATTTTACAAGAAATG-3'. These inverted repeats all contain a poorly conserved 21 bp sequence (Fig.4.4C)
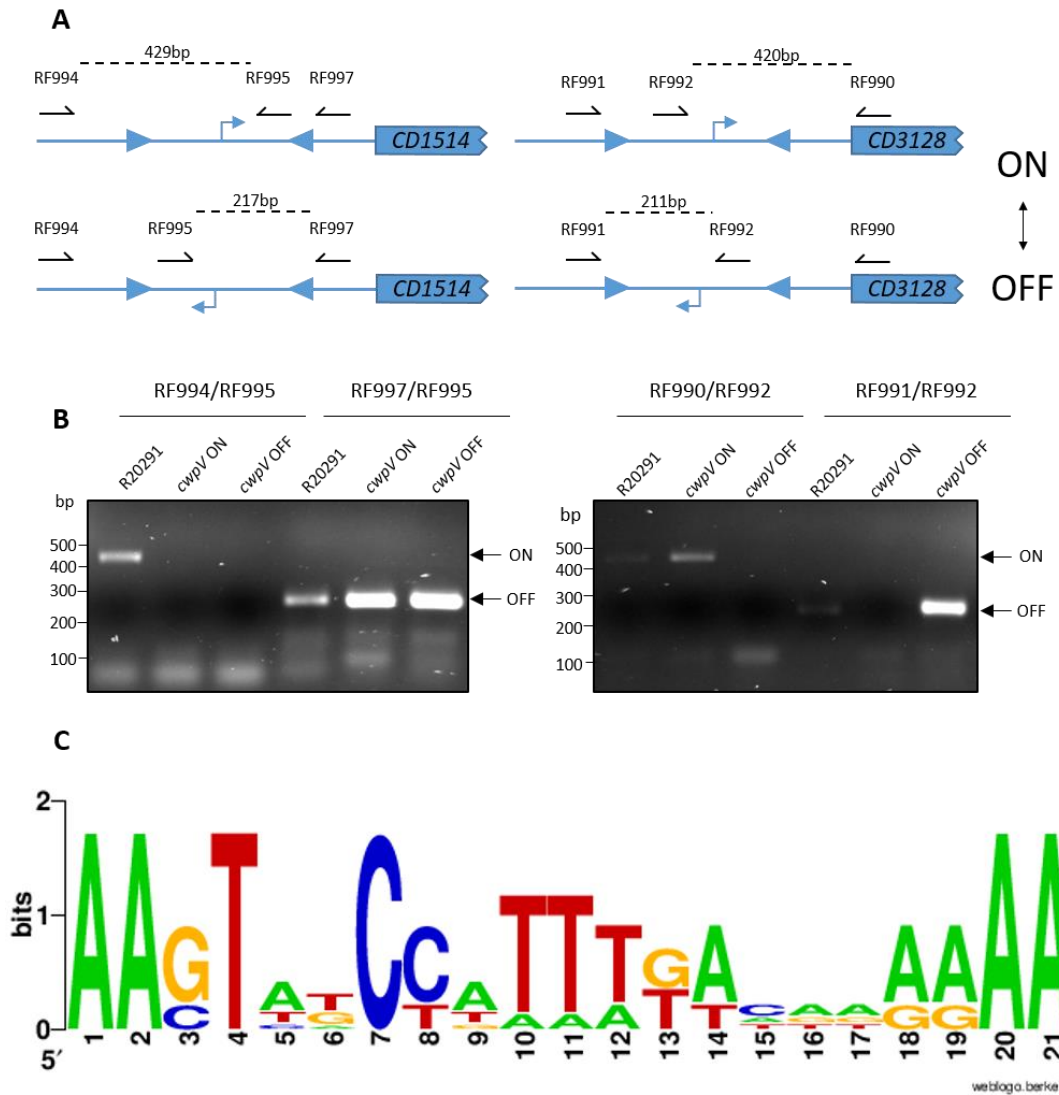
Fig.4.4: Identification of two novel phase-variable switches in the R20291 genome. (A) Schematic representation of the *CDR20291_1514* and *CDR20291_3128* promoter regions. BPROM analysis predicted that the promoter regions upstream of these genes are contained within the invertable region. The ON and OFF orientations of these switches were therefore modelled using the predicted promoter regions, as only one orientation would allow transcription. If these switches are in the ON orientation, PCR using RF994/RF995 *(CDR20291_1514)* or RF990/RF992 (*CDR20291_3128*) would result in products of 429 and 420 bp respectively. When in the OFF orientation, PCR using RF997/RF995 (*CDR20291_1514*) or RF991/RF992 (*CDR20291_3128*) would result in products of 217 and 211 bp respectively. **(B)** PCR analysis of R20291, R20291 *cwpV* ON, and R20291 cwpV OFF to determine *CDR20291_3128* and *CDR20291_1514* switch orientation. PCR analysis shows the *CDR20291_1514* and *CDR20291_3128* switches exist in both orientations in R20291 populations, confirming these regions as being phase variable. However, PCR products were only observed for one orientation in the *recV* inactivated strains suggesting that RecV is responsible for phase switching at these regions. **(C)** A logo-plot showing a 21 bp conserved sequence found within the inverted repeats that flank four phase variable switches.

4.2.3 Orientation of the Phase Variable Switches

4.2.3.1 Phase ON Frequencies of the *flgB, CDR20291_1514* and *CDR20291_3128* switches

As the *CDR20291_1514* and *CDR20291_3128* switches were identified in the inverse orientations in the FM2.5 genome, it was necessary to quantify the orientation frequency of all four phase variable switches. PCR analysis confirmed phase switching at the *flgB, CDR20291_1514,* and *CDR20291_3128* loci in FM2.5 and FM2.5RW (Fig.4.5A), as both orientations of the switches were observed in populations of these strains (data not shown). Orientation frequency was quantified using asymmetric digests of PCR fragments (Fig.4.5A and B). There was no significant difference in the orientation frequency of either the *CDR20291_1514* or *CDR20291_3128* between populations of R20291, FM2.5 and FM2.5RW. *CDR20291_1514* was shown to be ON in the minority of cells within a population (mean of 8.6% in R20291), while *CDR20291_3128* was shown to be ON in the majority of cells within a population (mean of 76.8% in R20291). There was, however, a significant difference in the observed frequencies of *flgB* ON cells within population of R02921 (median 0%) and FM2.5 (median 100% P<0.0005). However, no significant difference was observed in the frequency of *flgB* ON variants between populations of FM2.5 and FM2.5RW.

4.2.3.2 Orientation of the switch regions in the recV Inactivated Mutants

As R20291 *cwpV* ON and FM2.5RW *cwpV* ON display different colony morphologies (Fig.4.3C), the change in colony morphology previously reported to be due to CwpV expression, must in fact be CwpV-independent. The orientations of the other phase variable switches were determined in the *recV* inactivated mutants to confirm phenotypes associated with CwpV expression were not actually due to the other phase variable regions. As expected, the three other switch regions were in the same orientation in FM2.5 *cwpV* ON and FM2.5 *cwpV* OFF (all OFF), and FM2.5RW *cwpV* ON and FM2.5RW *cwpV* OFF, where only the *CDR20291_1514* switch

was in the OFF orientation. However, while the three other switches were in the OFF orientation in R20291 *cwpV* OFF, in R20291 *cwpV* ON the *CDR20291_3128* switch is locked in the ON orientation (Fig 4.5C). CDR20291_3128 expression may, therefore, be responsible for altered colony morphology.
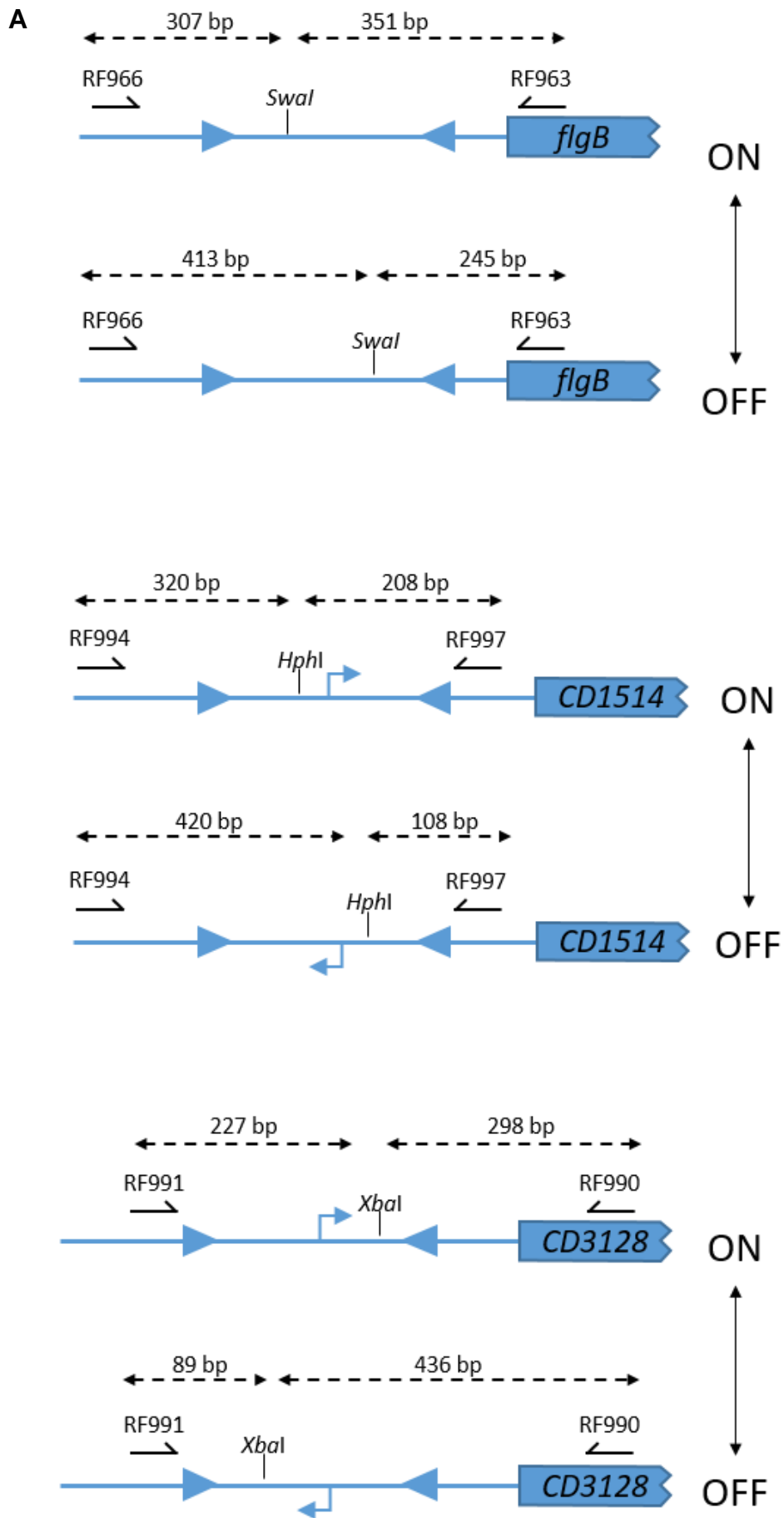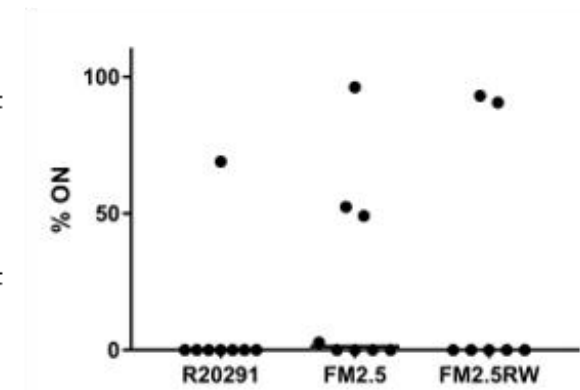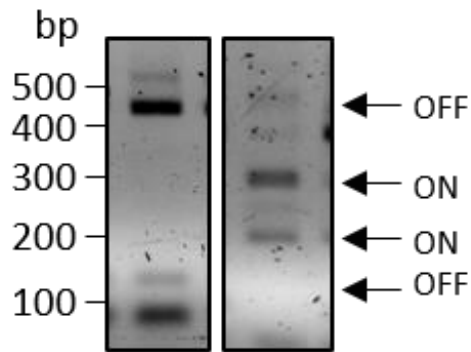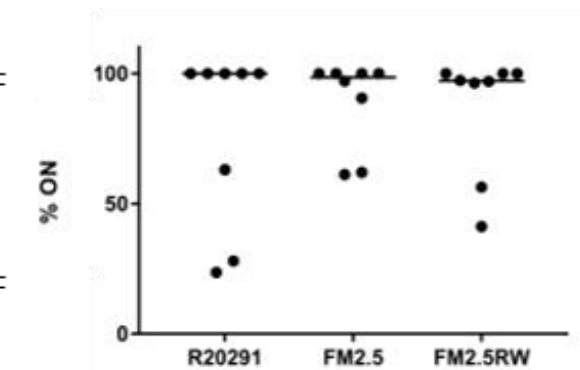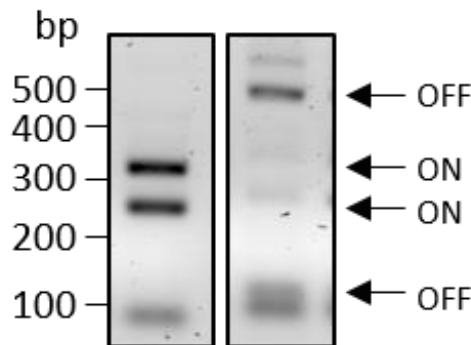
B

flgB
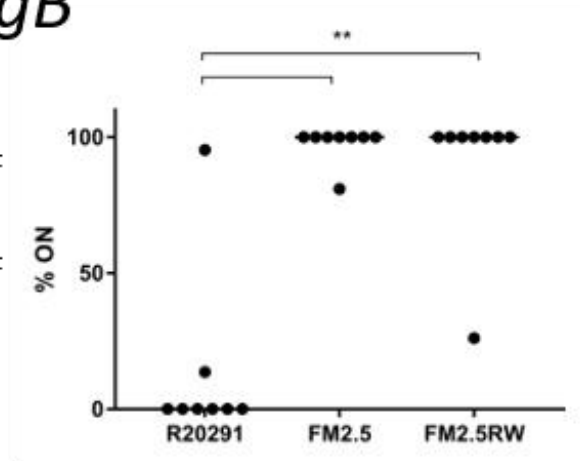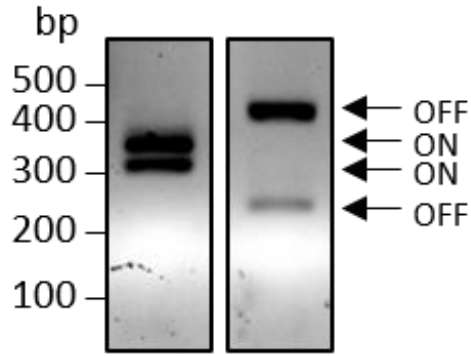
CD1514

CD3128

C

	cwpV	*flgB*	CD1514	CD3128
R20291 *recV::erm*	ON	OFF	OFF	ON
R20291 *recV::erm*	OFF	OFF	OFF	OFF
FM2.5 *recV::erm*	ON	OFF	OFF	OFF
FM2.5 *recV::erm*	OFF	OFF	OFF	OFF
FM2.5RW *recV::erm*	ON	ON	OFF	ON
FM2.5RW *recV::erm*	OFF	ON	OFF	ON

Fig.4.5: Orientation of the *flgB*, *CDR20291_3128* and *CDR20291_1514* genetic switches. (A) Schematic representation of the *flgB, CDR20291_1514,* and *CDR20291_3128* promoter regions. PCR amplification using primer pairs RF966/RF963, RF994/RF997, or RF991/RF990, respectively, results in a small DNA fragment containing an asymmetrically located restriction site. Digestion of these fragments using either *Swa*I (*flgB*), *Hph*I (*CDR20291_1514*), or *Xba*I (*CDR20291_3128*) results in differently sized fragments depending on the orientation of the switch. **(B)** Digested PCR fragments were run on 2% agarose gels and densitometric analysis performed, enabling the calculation of the proportion of each these switches in the ON orientation. This analysis was performed using eight biological repeats. **(C)** Analysis of the orientation of all four switches in the *recV* insertional mutants. Primer pairs RF964/RF963, RF995/RF994, RF992/RF990, NF825/NF823 (ON) and RF964/RF966, RF995/RF997, RF991/RF992, NF823/NF826 (OFF) were used to determine the orientation of the *flgB, CDR20291_1514, CDR20291_3128* and *cwpV* switches respectively.

4.2.4 Manipulation of the Native *slpA* Locus

The native *slpA* locus is notoriously difficult to manipulate. However, the isolation of S-layer deficient mutants provides a unique opportunity to genetically modify this locus. We have already replaced the native *slpA* with a watermarked variant (Kirk and Fagan, 2016). This paved the way for substitution of the native *slpA* with a mutant variant, encoding SlpA lacking the variable domain (domain 2) of the LMW subunit. This mutant (FM2.5RΔD2) has been used to provide further insight into S-layer structure (Banerji, 2017). As there are thirteen published SLCTs, it would be interesting to express the different *slpA* variants in FM2.5. This would allow investigation of the contributions of the different SLCTs to virulence using an isogenic expression system.

4.2.4.1 Substitution of Native SLCT-4 slpA with SLCT-7 slpA

Allele exchange mutagenesis using pMTLSC7215 was initially chosen to substitute the native SLCT-4 *slpA* with SLCT-7 *slpA*. This method would allow the exchange of the native *slpA* in not only FM2.5, but also in R20291. 1 kb fragments immediately upstream and downstream of *slpA* were cloned into pMTLSC7215, flanking *slpA* from 630. The resulting plasmid was denoted as pJAK072.

Conjugation of pJAK072 into both R20291 and FM2.5 was attempted using varying heat shock treatments (Kirk and Fagan, 2016). However, transconjugants were only obtained for R20291. As pJAK072 contains 1.2 kb upstream of *slpA*, it was assumed that this plasmid contains the complete *slpA* promoter. Expression of SLCT-7 SlpA was therefore investigated in R20291/ pJAK072. S-layer extraction was performed using O/N cultures of this strain and analysed via SDS-PAGE and western blotting using SLCT-4 and SLCT-7 specific antibodies. Only bands representing SLCT-4 HMW and LMW SLPs were observed in the SDS-PAGE and confirmed by western blot. These results suggest that plasmid borne *slpA* is not expressed in R20291/pJAK072.

Allele exchange mutagenesis was attempted as standard using this strain. The first recombination event, resulting in integration of the plasmid into the chromosome, was confirmed via PCR (Fig.4.6A and B). During the second recombination selection step, Avidocin-CD291.2 was added to the harvested cells. Avidocin-CD 291.2 kills SLCT-4, but not SLCT-7 expressing *C. difficile,* allowing selection of cells which have lost the SLCT-7 *slpA*. Eight colonies were isolated after selection for the second recombination event. S-layer extracts from these colonies were analysed by SDS-PAGE and western blotting. (Fig.4.6C and D). Of the eight colonies, one did not grow in TY broth and four still expressed SLCT-4 SlpA. One of the isolates seemed to contain an S-layer different to either SLCT-7 or SLCT-4 and showed extremely poor growth in TY broth, although it is possible that this was a contaminant. The two remaining isolates did not seem to express SlpA and were denoted JK1 and JK2. Primers RF110/RF111 were used to amplify the 5' region of *slpA*, shown to contain mutations in FM2.5 and FM2.6, from JK1 and JK2. Sequencing of these fragments revealed the same mutation, an insertion of a single adenine at position 283, as FM2.5. Interestingly, CwpV was observed in the S-layer extracts of both JK1 and JK2 (Fig.4.6C).

As pMTLSC7215 is notoriously difficult to conjugate into R20291, and as FM2.5RW and FM2.5RΔD2 were made using pMTL960 based vectors, the pJAK072 insert region was amplified using RF835/RF836 and ligated into pRPF185 via the *Kpn*I and *Sac*I sites. The resulting plasmid, pJAK076 was successfully conjugated into FM2.5. However, large colonies of FM2.5/pJAK076, indicative of *slpA* function complementation, were not observed. Due to the possibility that *slpA* under control of its native promoter may cause toxic effects when introduced on a multi-copy plasmid, the *slpA* promoter was removed from pJAK076. This was achieved through inverse PCR of pJAK076 using RF901/RF902, and ligation of the resulting fragment. The resulting plasmid, pJAK086, therefore contained a 115 bp upstream of *slpA*. However, when conjugated into FM2.5, again, large colonies were not observed.
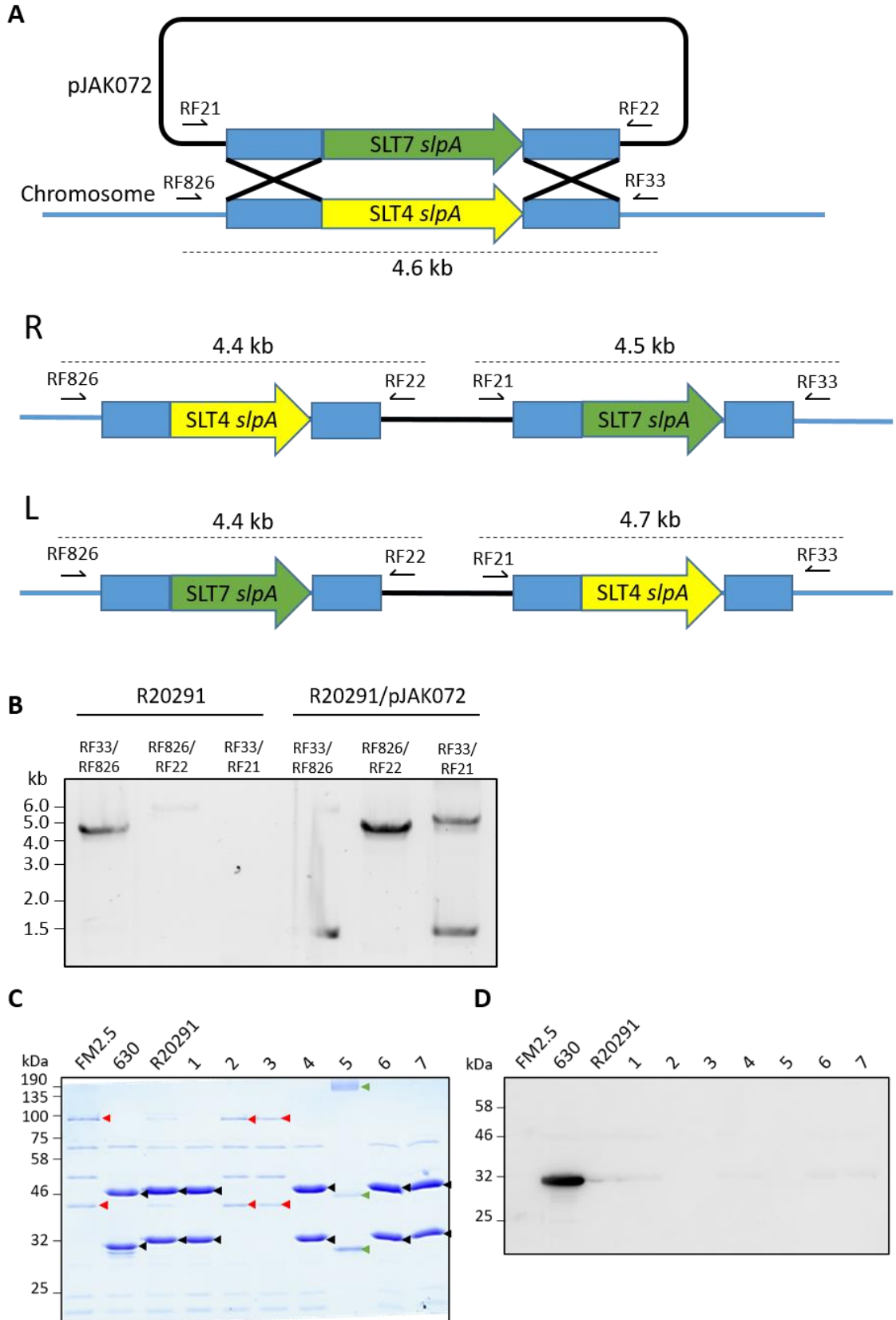
Fig.4.6: Manipulation of the native *slpA* locus. **(A)** 1 kb downstream of *slpA* was amplified from R20291 gDNA using RF659/RF660. *slpA* was amplified from 630 gDNA using RF750/RF658. These fragments were assembled into pMTLSC7215, linearised by PCR using primers RF311/312, by Gibson assembly. The resulting plasmid was linearised using RF779/RF780 and was used in a second Gibson assembly reaction with 1 kb upstream of *slpA*, which was amplified from R20291 gDNA using RF655/RF656. The resulting plasmid was denoted pJAK072 and was transformed into the conjugation donor CA434. Schematic representation of pJAK072 and the result of the first recombination event leading to insertion of the plasmid into the *C. difficile* genome. Recombination can happen at either the right homology arm or left homology arm (blue boxes), resulting in either sequence R or sequence L. **(B)** PCR analysis of potential single recombinants selected by an increase in colony size (Cartman et al., 2012). PCR was performed using combinations of primers flanking the plasmid insertion site, RF33 and RF826, and primers specific for the plasmid, RF21 and RF22. R20291 was included as a negative control. Bands at approximately 4.4 and 4.7 kb when primer pairs RF826/RF22 and RF33/RF21, respectively, were used and a lack of a band when primer pair RF33/RF826 were used indicates successful insertion of the plasmid into the chromosome. The single recombinants were restreaked to non-selective BHI and incubated for 72 h. Cells were then harvested and incubated with an excess of Avidocin-CD 291.2 for 1 h, prior to being plated onto CDMM containing 5-Fluorocytosine. After 96 h incubation eight colonies arose and were restreaked to BHI. Plasmid loss was confirmed via sensitivity to thiamphenicol. The eight restreaked colonies were used to inoculate O/N cultures and S-layer extracts isolated. S-layer extracts were analysed by SDS-PAGE **(C)** and western blotting using an anti-SLCT-7 LMW SLPA antibody **(D)**. No isolates were shown to express SLCT-7 SlpA, although two new *slpA* deficient mutants were isolated (lanes 2 and 3). One isolate also showed poor growth and an abnormal S-layer (colony 5). The LMW and HMW SLPs are labelled with a black arrow, CwpV is labelled with a red arrow, and unidentified proteins are labelled with a green arrow.

4.2.4.2 Characterisation of the slpA Promoter

Complementation of FM2.5 with native levels of SlpA, using mutant variants of *slpA* or variants from different S-layer cassettes, may be achievable by inserting *slpA*, under control of its native promoter, into the *pyrE* locus. This, however, would require characterisation of the *slpA* promoter. SlpA is one of the most abundantly expressed proteins in *C. difficile*, constituting approximately 15% of the total protein content of the cell suggesting that *slpA* expression is driven by an extremely strong promoter (Fagan and Fairweather, 2011). Therefore, characterisation of the *slpA* promoter may also have practical applications such as high-level expression of proteins for use in *in vitro* analysis.

4.2.4.2.1 In silico Analysis of the *slpA* Promoter

Firstly, DNA sequences upstream of all 13 recorded *slpA* cassettes (Dingle et al., 2013) were subjected to BPROM promoter prediction and sequence alignment. Distinct promoter families were identified via the alignments (Fig.4.7A). *slpA* cassettes 1-8 and H2/6, contain near identical sequences upstream of *slpA* (>90% identity). Cassettes 10 and 12 showed greater identity between them than when compared to others (approximately 89%) and Cassettes 9 and 11 showed relatively poor sequence identity to other cassettes types (<80%).

BPROM analysis of sequences from cassettes 1-8 identified conserved -10 and -35 promoter regions. This region, although differing in sequence relative to cassettes 1-8, was also quite well conserved within cassettes 9-12 (Fig.4.7B).

The sequence immediately upstream of the -35 region (-42 to -64) in cassettes 1-8 showed sequence similarity to the consensus sequence of upstream promoter elements (UPE) of *E. coli* (nnAAA(A/T)(A/T)T(A/T)TTTTnnAAAAnnn) (Fig.4.7B) (Estrem et al., 1998)*.* This sequence was not identified in cassettes 9-12, which may suggest that cassettes contain weaker promoters, or increase promoter activity using a different mechanism.

4.2.4.2.2 The *slpA* promoters contains a UPE

To confirm the presence of an upstream promoter element within the *slpA* promoter, a promoter activity assay was performed. The promoter was cloned to either include up to the - 40 or -80 regions and introduced into pRPF185 between the *Kpn*I and *BamH*I restriction site. This placed *gusA* under control of the *slpA* promoter without the potential UPE (pJAK106, referred to as *PslpA*-SHORT) or with the potential UPE (pJAK107, referred to as *PslpA*-LONG). β-glucuronidase assays were performed using R20291/*PslpA*-SHORT and R20291/*PslpA*-LONG to determine whether the sequence upstream of the -35 region increased promoter activity. pRPF144, a plasmid bearing *gusA* under the control of the constitutive *cwp2* promoter, was used as a positive control.

As expected, pRPF144 resulted in greater enzyme activity relative to the negative control, pSEW027, which lacks *gusA* ($P<0.0005$). *PslpA*-SHORT resulted in significantly higher enzyme activity relative to pRPF144 ($P<0.0005$). *PslpA*-LONG resulted in significantly higher enzyme activity (approximately 7-fold) compared to *PslpA*-SHORT ($P<0.0005$). This result strongly suggests that the *slpA* promoter contain an UPE. It is worth noting that before measuring the enzyme activity samples were centrifuged to remove cell debris and GusA activity seemed to be localised to the pellet. This could have led to localised substrate saturation in the *PslpA*-LONG samples, meaning the difference in enzyme activity between samples of *PslpA*-SHORT and *PslpA*-LONG may be greater than recorded.

A

|	h6/2	1	2	3	4	5	6	7	8	9	10	11	12
h6/2		97.642%	92.453%	98.585%	96.698%	98.585%	99.528%	92.925%	98.585%	69.820%	70.270%	70.136%	68.778%
1	97.642%		91.509%	99.057%	98.113%	98.585%	98.113%	91.981%	98.585%	68.919%	69.820%	70.588%	68.326%
2	92.453%	91.509%		92.453%	90.566%	91.981%	91.981%	96.698%	91.981%	67.117%	68.468%	68.326%	67.421%
3	98.585%	99.057%	92.453%		98.113%	99.057%	99.057%	92.925%	99.057%	69.369%	70.270%	70.588%	68.778%
4	96.698%	98.113%	90.566%	98.113%		97.642%	97.170%	91.038%	97.642%	67.660%	66.809%	68.220%	65.812%
5	98.585%	98.585%	91.981%	99.057%	97.642%		99.057%	92.453%	100%	68.468%	69.369%	70.455%	67.873%
6	99.528%	98.113%	91.981%	99.057%	97.170%	99.057%		92.453%	99.057%	69.369%	70.270%	70.588%	68.778%
7	92.925%	91.981%	96.698%	92.925%	91.038%	92.453%	92.453%		92.453%	67.568%	68.919%	68.326%	67.873%
8	98.585%	98.585%	91.981%	99.057%	97.642%	100%	99.057%	92.453%		68.468%	69.369%	70.455%	67.873%
9	69.820%	68.919%	67.117%	69.369%	67.660%	68.468%	69.369%	67.568%	68.468%		76.949%	77.143%	74.411%
10	70.270%	69.820%	68.468%	70.270%	66.809%	69.369%	70.270%	68.919%	69.369%	76.949%		76.829%	88.776%
11	70.136%	70.588%	68.326%	70.588%	68.220%	70.455%	70.588%	68.326%	70.455%	77.143%	76.829%		72.581%
12	68.778%	68.326%	67.421%	68.778%	65.812%	67.873%	68.778%	67.873%	67.873%	74.411%	88.776%	72.581%	

B

C
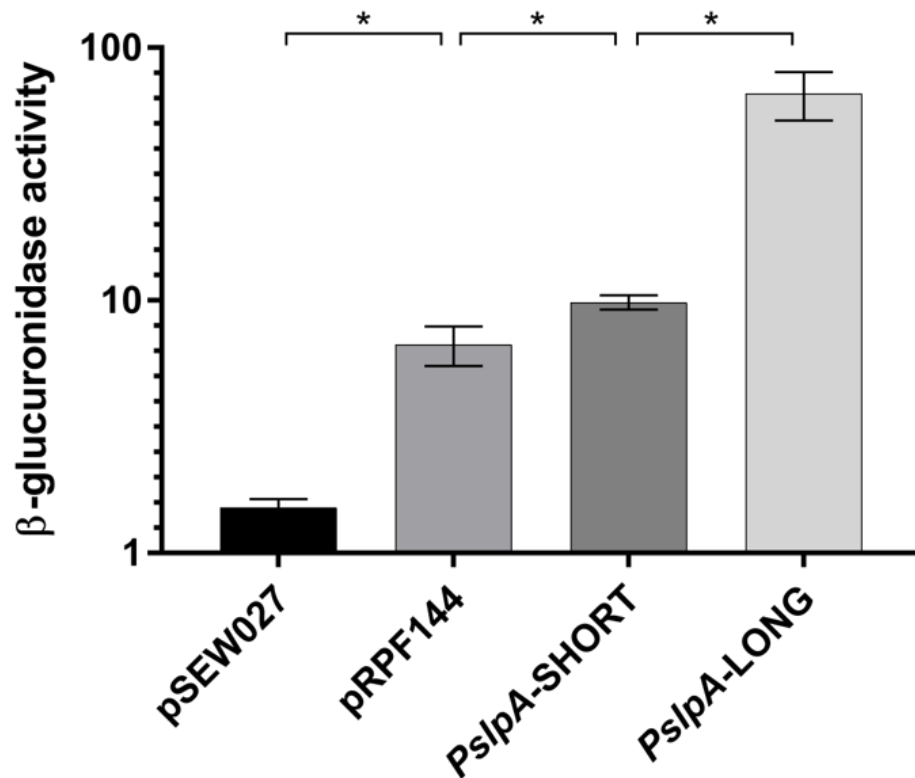
Fig.4.7: Characterisation of the *slpA* promoter. (A) Sequence identity between the sequences upstream of slpA from all 13 documented slpA cassettes. The entire non-coding region upstream of the *slpA* from all SLCTs were aligned using the Geneious ClustalW alignment tool. Numbers across the top and sown the left correspond to the SLCT and the % sequence identity is presented. **(B)** Sequence alignment of the cassette 1-8 sequences reveals high sequence conservation near the predicted promoter regions. The region immediately upstream of the -35 region resembles an UPE from *E. coli* promoters (Estrem et al., 1998). **(C)** The slpA promoter was cloned to include up to the -35 region using primer pair RF1052/RF1053, or to approximately -80 of the transcriptional start site using primer pair RF1051/RF1053. These fragments were ligated into pRPF185 via the KpnI and SacI restriction sites, creating PslpA-sHORT and PslpA-LONG respectively, placing *gusA* under the control of the slpA promoters. These plasmids were conjugated into R20291. pRPF144, in which *gusA* is under the control of the cwp2 promoter, was included as a positive control, and pSEW027, which lacks *gusA*, was used as a negative control. R20291 strains bearing these plasmids were grown O/N in biological duplicate and subcultured to OD600nm 0.05 and were harvested at approximately OD600nm 0.7. Bacterial pellets were then used in a β-glucuronidase assay as previously described (Dupuy and Sonenshein, 1998). PslpA-SHORT did result in significantly higher enzyme activity relative to pRPF144. PslpA-LONG, containing the putative UPE, resulted in an even greater GusA expression. The mean from experiments using biological duplicates performed in technical triplicate are plotted. Error bars correspond to the standard deviation. Statistical significance was calculated using two-tailed T-tests using Welch's correction (* P<0.0005).

4.2.5 S-layer Pore Size and Innate Immune Effector Resistance

The SlpA deficient mutant, FM2.5, shows increased sensitivity to the innate immune effectors LL37 and lysozyme (Kirk, 2017). To investigate the mechanism(s) behind this resistance, LL37 and lysozyme sensitivity assays were performed using FM2.5RΔD2 (Banerji, 2017). This strain lacks the variable domain in the LMW subunit of SlpA (Fig.4.8A), causing an apparent increase in S-layer pore size (Banerji, 2017).

Growing cultures of R20291, FM2.5 and FM2.5RΔD2 were treated with lysozyme (500 µg/ml) or LL37 (5 µg/ml), and the optical density of the culture was recorded hourly. As previously observed, treatment of FM2.5 with lysozyme results in significantly slowed growth relative to R20291 and treatment with LL-37 completely killed FM2.5 (Fig.4.8B). Treatment of FM2.5RΔD2 with lysozyme results in slowed growth, comparable to the effect of seen with FM2.5 (Fig.4.8B). In contrast, treatment of growing cultures of FM2.5RΔD2 with LL37 did not result in complete killing as with FM2.5. However, a larger drop in cell density was observed when FM2.5RΔD2 was treated with LL37 relative to R20291.

A
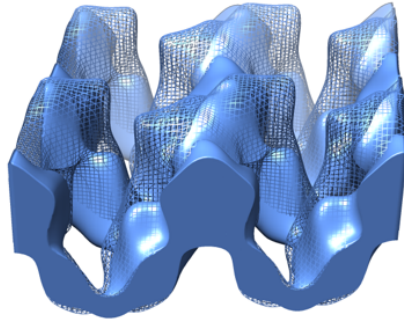
B Lysozyme
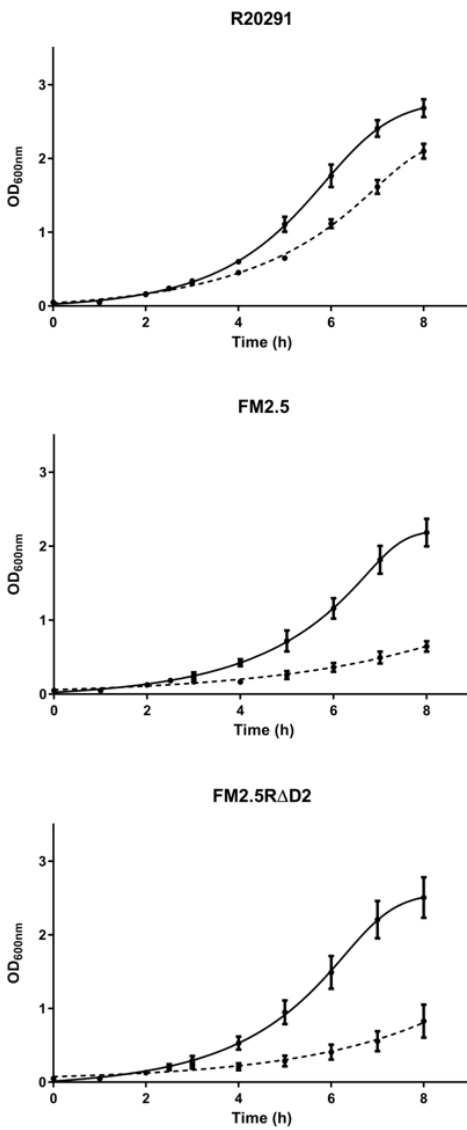
C LL-37

Fig.4.8: Sensitivity of FM2.5RΔD2 to the innate immune effectors, lysozyme and LL37. (A) 3D model of the native R20291 S-layer at 2 Å generated using data from electron diffraction experiments (Banerji, 2017) showing the structure of the S-layer of R20291 (blue mesh) compared to the S-layer from FM2.5RΔD2 (blue). Bro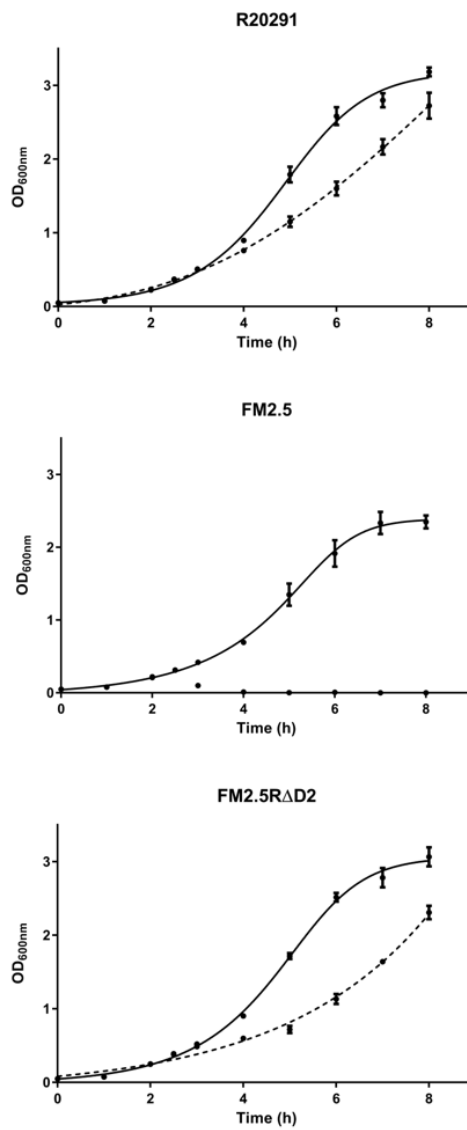th based killing assays were performed where the OD_{600nm} of growing cultures of R20291, FM2.5, and FM2.5RΔD2 was monitored every hour after addition of lysozyme (500 µg/ml.) or LL37 (5 µg/ml). LL37 and lysozyme were added to cultures 2.5 h post-inoculation (black arrow). **(B)** Addition of lysozyme to growing cultures of FM2.5RΔD2 results in a reduced growth rate comparable to that of FM2.5. **(C)** Addition of LL37 to growing cultures of FM2.5RΔD2 results in impaired growth relative to R20291. However, addition of 5 µg/ml of LL37 to growing cultures of FM2.5 results in complete cell death, suggesting increased LL37 resistance in FM2.5RΔD2 relative to FM2.5. LL37 and lysozyme challenge experiments were performed in triplicate using biological duplicates. Error bars show the standard deviation of the data.

4.3 Discussion

Unlike other bacterial species, the S-layer of *C. difficile* is essential for growth *in vitro* (Dembek et al., 2015), although the reasons for this remain unknown. Due to the lack of *slpA* mutants, the functions of the S-layer are poorly understood. Recently, spontaneous *slpA* mutants of *C. difficile,* have been isolated (Kirk et al, 2017). These strains present a unique opportunity to study the functions of the S-layer and attempt to understand why the S-layer appears to be essential in WT *C. difficile*. Here we have made attempts to understand the effects of *slpA* loss to gain understanding on S-layer function.

The most obvious difference observed with the SlpA deficient mutants was increased CwpV expression. Interestingly this was observed in all four independent *slpA* mutants (FM2.5, FM2.6, JK1, and JK2). As *cwpV* is under the control of a phase variable promoter, where expression of CwpV is dependent on the orientation of a 195 bp switch region upstream of the *cwpV* ORF, we explored the possibility that the *cwpV* switch is biased towards the ON orientation in populations of FM2.5. It was confirmed that in FM2.5 the switch region is found more frequently in the ON orientation. We originally hypothesised that this observation was due to increased growth rate of FM2.5 *cwpV* ON variants, which would result in higher proportions of *cwpV* ON cells within a population. However, no difference in growth rate or cell morphology was observed in *recV* insertional mutants with *cwpV* expression locked either ON or OFF. This suggests that the overexpression of CwpV in populations of FM2.5 is due to changes in regulation of RecV-dependent site-specific recombination. Interestingly, the *cwpV* region is found in the ON orientation in FM2.5RW at similar frequencies to R20291, suggesting that SlpA expression may influence switching of the *cwpV* switch. Surprisingly, the colony morphology of FM2.5 and FM2.5RW *cwpV* ON and OFF strains did not match the previously reported colony morphologies (Reynolds et al., 2011). This may be due to the orientation of the *CDR20291_3128*

switch as the its orientation differs between FM2.5RW *cwpV* ON and R20291 *cwpV* ON, which theoretically have the same surface architecture.

The genome of FM2.5 has previously been sequenced using illumina (Kirk et al, 2017). This method was chosen due to its routine use and relatively low cost. As no obvious suppressor mutations were observed in the illumina sequence, SMRT sequencing was employed. This method provides much larger read-lengths, averaging 10 kb, enabling easier identification of large insertions, deletions and inversions within the genome (Rhoads and Au, 2015). This is particularly useful as the *C. difficile* genome contains many repeat sequences. Using this method additional mutations were found within the FM2.5 genome. Although most of these mutations occur in genes of unknown function, a SNP within the gene encoding the regulatory protein, CodY, was identified. CodY is a negative regulator of both toxin production and sporulation (Dineen et al., 2010; Nawrocki et al., 2016), and it would be interesting to see whether this mutation contributes to the inability of FM2.5 to efficiently sporulate and produce toxins.

As FM2.5 shows increased CwpV expression, it was unsurprising that the orientation of the *cwpV* switch was observed in the ON orientation in the FM2.5 genome. Interestingly however, other phase variable regions were found to be in their inverse orientations relative to the published R20291 genome. This includes the putative phase variable switch upstream of *CDR20291_1514* and a previously undescribed phase variable region upstream of *CDR20291_3128*. The presence of these switches in both orientation within populations of *C. difficile* was confirmed via PCR analysis and the switch inversion was shown to be dependent on RecV. RecV has been shown to control the orientation of the *cwpV* switch (Emerson et al., 2009) and the *flgB* switch (Anjuwon-Foster and Tamayo, 2017), meaning that this recombinase is responsible for the phase switching of all switch regions identified to date. This is particularly interesting as the inverted repeat regions flanking these genetic switches share little sequence similarity. Asymmetric digest analysis also shows that the orientations of the switch regions are

regulated differently. *cwpV* switch regulation is, as previously mentioned, abnormal in FM2.5 and corrected in FM2.5RW. *flgB* orientation regulation also seems to be abnormal in FM2.5, but is not corrected in FM2.5RW. Finally, both the *CDR20291_3128* and *CDR20291_1514* orientation switching are not affected by loss of *slpA.* It is worth mentioning, that while asymmetric digests are useful for comparing different samples, they may not be sensitive enough to detect rare inversion events. Therefore, the reported proportion of *CDR20291_3128* ON and *CDR20291_1514* ON cells within a population may not accurately reflect the actual proportions. It may therefore be necessary to use more sensitive techniques, such as reverse-transcription PCR, to determine the frequency of *CDR20291_3128* and *CDR20291_1514* being in the ON orientations. As the orientation of the *flgB* switch has been shown to change over time within a population (Anjuwon-Foster and Tamayo, 2017), it may also be necessary to calculate phase switching at different stages of the bacterial growth. The difference in inverted repeat sequence may provide some insight into the mechanism(s) in which these regions are regulated. As the orientation switching of the different regions seem to differ, as not all switches are affected by SlpA loss, it is likely that they are regulated by different mechanisms. One possible explanation is the presence of co-factors which may be required for RecV function. Co-factors required for orientation switching may differ for the different switches. As c-di-GMP is known to be involved in regulation of expression of the *flgB* operon and CDR20291_1514 encodes a c-di-GMP PDE, c-di-GMP may act as a co-factor for RecV-dependent orientation switching. This would allow coupling of expression of the regulatory *CDR20291_3128* operon, surface antigen expression, motility, and c-di-GMP degradation to levels of intracellular c-di-GMP and, therefore, virulence and biofilm production (Bordeleau and Burrus, 2015).

With the confirmation that two other phase variable regions, other than *cwpV* and *flgB*, were under control of RecV, the orientations of the four genetic switches in the *recV* inactivated mutants were determined. As FM2.5 *cwpV* OFF and FM2.5RW *cwpV* ON were created using allele exchange, the orientations of the other switches were the same between the ON and OFF

strains. However, R20291 *cwpV* ON was created through expression of *recV* from a plasmid. This means that the orientations of the other switches may not necessarily match R20291 *cwpV* OFF. In fact, in R20291 *cwpV* ON, *CDR20291_3128* was also in the ON orientation, while in R20291 *cwpV* OFF, *CDR20291_3128* was in the OFF orientation. Interestingly *CDR20291_3128* was also found to be in the ON orientation in the FM2.5RW *recV*::*erm* strains. This observation may explain why both FM2.5RW *cwpV* ON and OFF strains share a colony similar to that of R20291 *cwpV* ON. Whilst this trait was assumed to be due to expression of CwpV, it now seems more likely to be due to expression of the *CDR20291_3128* operon. To confirm this, it would be necessary to show that the orientation of the *CDR20291_3128* switch assumed to be ON does result in transcription of the downstream genes. In addition, it would be necessary to show that expression of the *CDR20291_3128* operon alone is sufficient to result in altered colony morphology. This could easily be performed using R20291 *cwpV* OFF, as all genetic switches in this strain are in the OFF orientation. The orientation of the *CDR20291_3128* switch is in the OFF orientation in the FM2.5 *recV*::*erm* strains, which do show a small, smooth edged colony morphology. However, this may be explained by the huge alteration of surface architecture of these strains relative to wild-type.

Manipulation of the *slpA* locus is notoriously difficult. Previous attempts to inactivate *slpA* have proven unsuccessful, and it is now known that *slpA* is essential for growth (Dembek et al., 2015). The isolation of FM2.5 raised hopes that the *slpA* locus may now be more easily manipulated. This has proven to be true to some extent. Insertion of *slpA,* encoding a functional protein, into the native locus has been successful, e.g. insertion of intact *slpA* containing a watermark allowed construction of strain FM2.5RW. Additionally, through insertion of *slpA* lacking the coding region for domain 2 of the LMW subunit, strain FM2.5RΔD2 was created. This demonstrates that we are now able to introduce mutant variants of SLCT-4 *slpA* into the FM2.5 genome, potentially only if the gene expresses a functional SlpA. This strain has allowed us to further probe the functions of the S-layer. Treatment of growing cultures of FM2.5RΔD2 with

lysozyme resulted in slowed growth, similar to the effect seem in FM2.5 However, addition of LL37 to growing cultures of FM2.5RΔD2 did not result in complete cell death, as observed in FM2.5. As FM2.5RΔD2 has been reported to contain larger pores in the S-layer (Banerji, 2017) it may therefore be hypothesised that S-layer-dependent resistance to lysozyme may be caused by the S-layer acting as a physical barrier to lysozyme, preventing the enzyme from reaching the cell wall. The S-layer therefore acts as a molecular sieve, as previously hypothesised (Fouet et al., 1999). However, as increasing the pore size of the S-layer had minor effects on resistance to LL37, the S-layer may confer resistance to LL37 by altering the charge of the cell surface, preventing the CAMPs from reaching the cell membrane. This is a similar mechanism to lipopolysaccharide dependent CAMP resistance seen in Gram negative bacteria (Band and Weiss, 2015).

This exemplifies how manipulation of *slpA* may help further understand the functions of the S-layer and the underlying mechanisms. However, even in FM2.5, manipulation of the native *slpA* locus has proven difficult. The *slpA* promoter has now been identified in R20291 which seems to include a UPE. Due to the sequence similarity between the R20291 promoter region and those from strains bearing S-layer cassettes 1-8, it is very likely this promoter region is conserved between these strains. However, it is still not understood how the promoters of cassettes 9-12 are so apparently strong, without an apparent UPE. High levels of *slpA* expression in these strains may be attributed to the presence of an additional promoter, although experimental evidence is still required. Identification and characterization of the *slpA* promoter may allow deletion of native *slpA* in FM2.5, and insertion of *slpA* under control of its native promoter into the *pyrE* locus. This would enable the study of all thirteen different SLCTs using an isogenic expression system.

Chapter V

SecA2 and Biogenesis of the S-layer

5.1. Introduction

In addition to poor understanding of its function, little is known about the biogenesis of the S-layer. SlpA is secreted via the accessory Sec secretion pathway. *C. difficile* expresses two, independently functional SecA proteins. SecA1 is thought to be responsible for the secretion of the majority of proteins while SecA2 secretes a small subset of proteins. SlpA is dependent on SecA2 for secretion (Fagan and Fairweather, 2011). However, the reason for *C. difficile* expressing two, functionally independent SecA proteins is unknown. As SlpA is so abundant, it is possible that an additional SecA is required to prevent saturation of the Sec secretion pathway. However, due to the lack of an additional SecY, it is likely that SecA2 uses the canonical channel for protein translocation. There is also the possibility that SecA2 provides temporal regulation on SlpA secretion. However, this is unlikely as SlpA is expressed constitutively and is essential for growth *in vitro* (Dembek et al., 2015). Another possibility is that SecA2 enables localised secretion. SecA2 mediated secretion is a poorly understood process and *M. tuberculosis* is the only organism to date for which the structure of both SecA2 and SecA1 is known (Sharma et al., 2003; Swanson et al., 2015).

In rod-shaped bacteria, synthesis of the cell wall occurs along the long axis of the cell and at the septum (Vollmer and Bertsche, 2008). There are two multi-enzyme complexes responsible for peptidoglycan synthesis, the elongasome and the divisome. These complexes are organised by the cytoskeletal proteins, the actin homologue MreB and the tubulin homologue FtsZ, respectively. MreB is thought to localise around the long axis of the cell, ensuring peptidoglycan synthesis occurs at these sites, maintaining cell shape and integrity (as

reviewed in (Errington, 2015)). However, the exact pattern of localisation is disputed. There are two *C. difficile* MreB homologues, MreB1 and MreB2, both of which are essential for growth *in vitro* (Dembek et al., 2015). FtsZ undergoes GTP-dependent polymerisation forming filamentous arches which form the Z-ring at the mid cell, the molecular scaffold essential for the localisation of at least 10 proteins involved in cell division (as reviewed in (Ortiz et al., 2016)).

To further understand the mechanisms behind SecA2 substrate specificity and function, we have performed X-ray crystallisation of SecA2. SecA proteins are composed of several functional domains. NBD1 and NBD2 form the DEAD motor box domain which binds and hydrolyses ATP providing energy for protein transport, the PPXD domain is involved in substrate binding, the HSD links the specificity domain to the motor domain, the HWD links the IRA1 domain thought to insert the substrate into the channel, and the CTD which interacts with SecB in Gram-negative bacteria and phospholipids (Feltcher and Braunstein, 2012). We have identified differences in SecA2 structure when compared to the structure of *Mycobacterium tuberculosis* SecA1 (*Mtb*SecA1). Some of these differences are conserved between *C. difficile* SecA2 and *M. tuberculosis* SecA2 (*Mtb*SecA2). In addition, we provide evidence for MreB and FtsZ involvement in SecA2 dependent secretion and suggest that S-layer biogenesis is localised to areas of cell wall synthesis.

5.2 Results

5.2.1 3D Modelling of SecA Proteins

5.2.1.1 Purification of SecA1 and SecA2

Before *in vitro* analysis was performed, SecA1 and SecA2 were purified from *C. difficile* lysates. Plasmids expressing N-terminally strep-tagged variants of *C. difficile* strain 630 SecA2 and SecA1*,* pRPF186 and pRPF193 respectively (Fagan and Fairweather, 2011), were conjugated into 630. The resulting strains were used to over-express the two proteins that were then

purified using Strep-tag affinity chromatography. However, the resulting yields of SecA2 and SecA1 were very low (approximately 2 mg and 0.3 mg, from 10 l of *C. difficile* culture, respectively). SecA1 and SecA2 preparations were very pure, although SecA1 samples were contaminated with a protein with a MW of approximately 45 kDa (Fig.5.1A).

As SecA1 is mainly localised at the membrane (Fagan and Fairweather, 2011), the effect of treatment of cell lysates with detergent on the solubility of SecA1 was investigated. Supernatants of cell lysates treated with varying concentrations of N-lauroylsarcosine were analysed via western blot. None of the treatments substantially increased the solubility of SecA1 (Fig.5.1B).
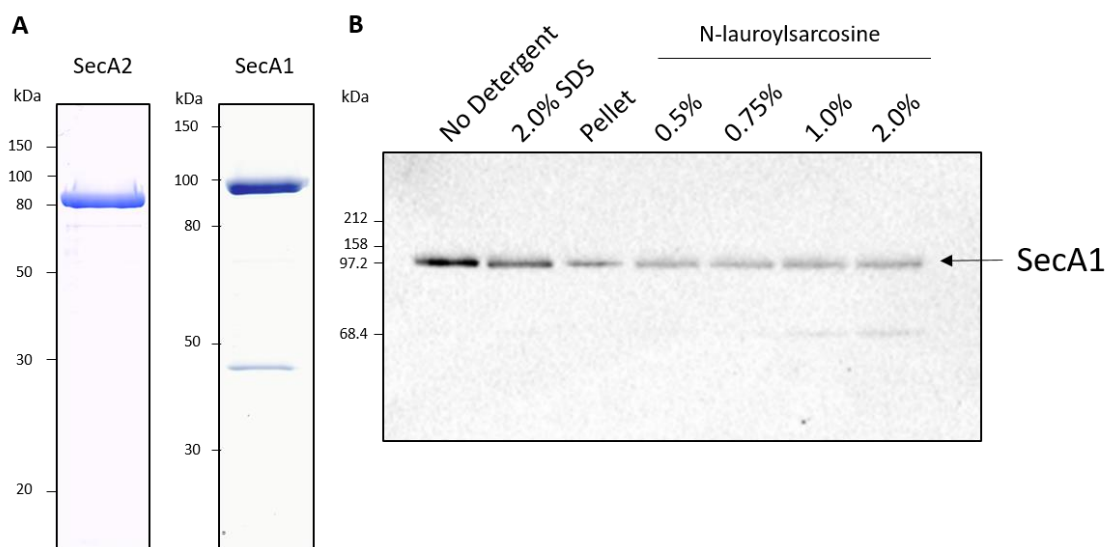
Fig.5.1: Purification of SecA1 and SecA2. (A) 10 L of TYG was inoculated with 630/pRPF186 or 630/pRPF193 and grown to an approximate OD_{600} of 0.5. Protein expression was then induced through addition of 700 ng/ml of Atc. The cultures were then incubated anaerobically for a further 5-6 h, at which point cells were harvested via centrifugation (4,700 x g, 10 min). Cells were resuspended in 2x StrepTactin purification buffer and lysed using the CD27L bacteriophage endolysin (Mayer et al., 2008). The soluble fraction of the resulting lysates were diluted using 1x binding buffer and loaded onto two 1 ml StrepTactin columns. Protein was eluted from the column through the addition of an increasing gradient of elution buffer containing desthiobiotin. SDS-PAGE showing purified SecA2 and SecA1. Little contamination is seen in preparations of SecA2, although SecA1 contains a contaminating protein with a MW of approximately 45 kDa. **(B)** Cell lysates of 630/pRPF193 were incubated without detergent or with 2% SDS or 0.5, 0.75, 1, and 2% N-lauroylsarcosine. Samples were then incubated at 4°C for 30 min and centrifuged to remove cell debris (20,000 x g, 2 min). Supernatants were resolved by SDS-PAGE and analysed via western blotting using an anti-Strep-tag antibody. A pellet of untreated cell debris was also treated with detergent to determine protein content of the non-soluble fraction (lane Pellet). Although some SecA1 is lost in the non-soluble fraction, more SecA1 is present in the soluble fraction when not treated with detergent.

5.2.1.2 SecA1 and SecA2 X-Ray Crystallography

Purified samples of SecA1 and SecA2 were concentrated using Amicon centrifugation filters. However, neither protein could reach concentrations above approximately 1 mg/ml before precipitation was observed. Protein samples, at approximately 1 mg/ml, were used in four 96-well crystal screen trials. No crystal growth was observed for SecA1. However, SecA2 crystals were observed in five different conditions in the PACT crystal screen. The largest, non-needle like crystals were observed in conditions B10 (0.2 M magnesium chloride, 0.1 M MES pH 6.0, 20 % w/v PEG 6000) and C10 (0.2 M magnesium chloride 0.1 M HEPES pH 7.0, 20 % w/v PEG 6000) (Fig.5.2A). X-ray diffraction data was obtained to approximately 3.3 Å using these crystals. Optimisation of crystal growth conditions, data analysis, and 3D model assembly were performed by Dr Paula Salgado (University of Newcastle). Optimisation of crystal growth resulted in several conditions that yielded crystals. The structure of SecA proteins vary dependent on the presence of bound ATP or ADP. Crystallisation trials containing the unhydrolysable ATP analogue, AMP-PNP, or ADP were also performed. Different conditions were used to obtain crystals of SecA2 in its unbound, ADP bound, and AMP-PNP bound forms. Diffraction data was collected for all these SecA2 isoforms. Diffraction data was collected to 2.65 Å for SecA2-ADP. Phasing was inferred by molecular replacement using the SecA structure of *Thermatoga maritima*. Although we now have a complete model of SecA2, further refinement is still necessary as some areas do not fit within the electron density plot. This is especially true for residues 230-335 which correspond to the PPXD domain (Fig.5.2B and C).

The structure of SecA2 contains all domains expected for a SecA protein including the HSD, HWD, IRA1, PPXD, and a DEAD box motor domain consisting of NBD1 and NBD2, although as expected, no CTD was identified within the model as this domain is believed to be very flexible. As the structure of SecA1 could not be determined, *Mtb*SecA1 is shown for comparison (Fig.5.2D). When compared to *Mtb*SecA1, there is a striking difference in the conformation of

the PPXD domain, which is rotated towards NBD2 in SecA2. In addition, NBD2 lacks the two alpha helices that form the VAR domain.

A

B10

C10

100 µm

100 µm

B

C

D
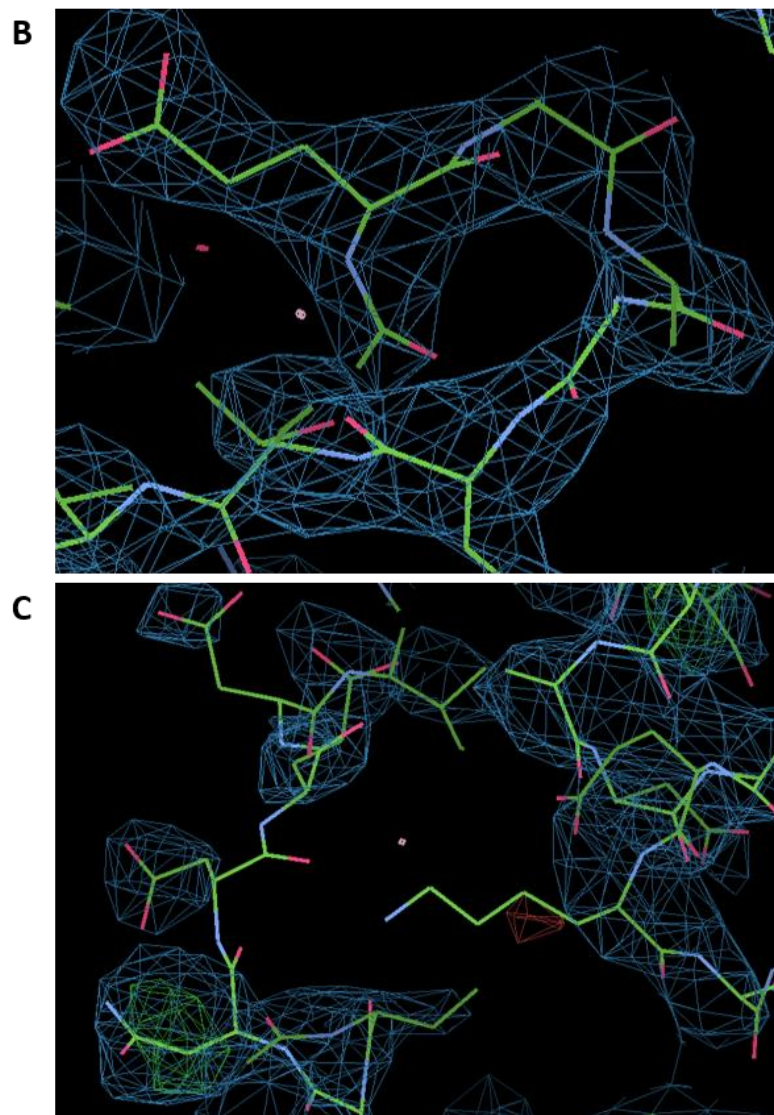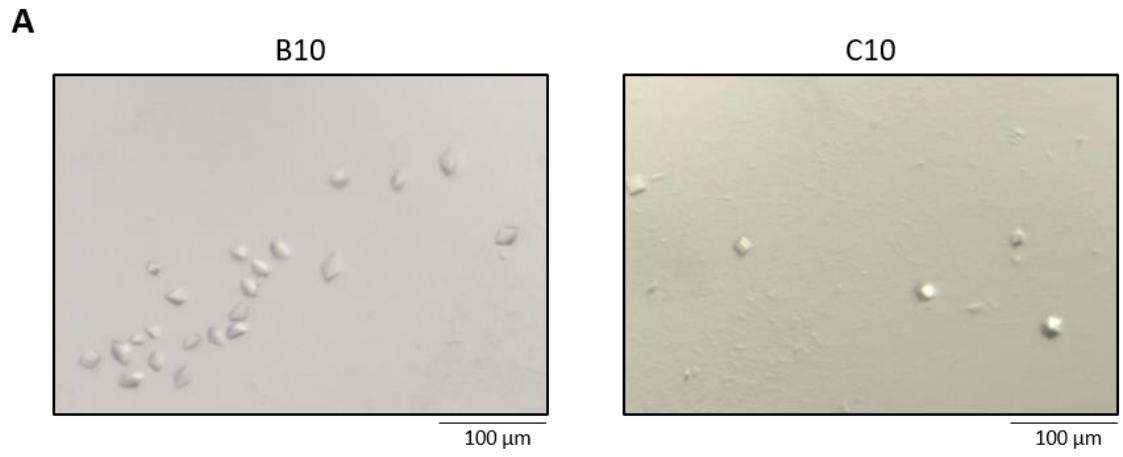
Clostridium difficile
SecA2

Mycobacterium tuberculosis
SecA1

SecA1

NBD1			PPXD		NBD2	HSD	HWD	IRA1	CTD

SecA2

NBD1			PPXD		NBD2	HSD	HWD	IRA1
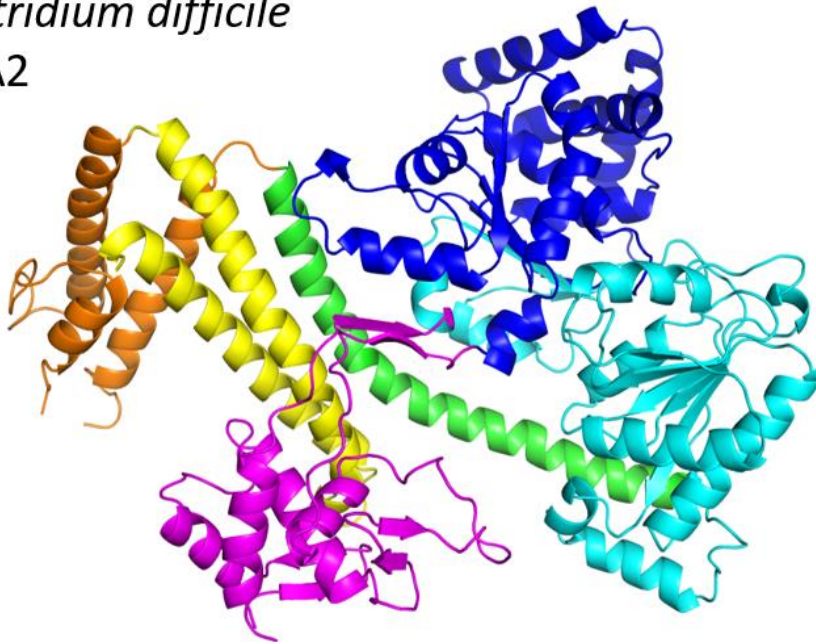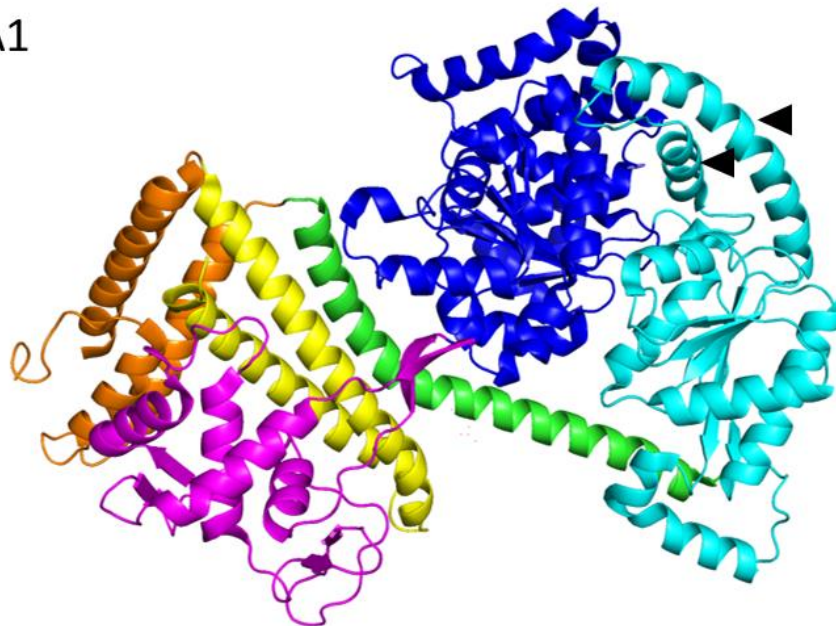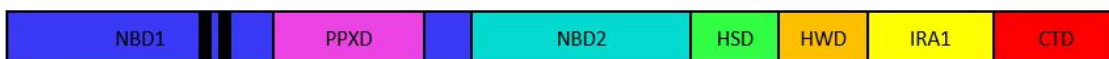
Fig.5.2: X-ray crystallography of SecA2. (A) Purified and concentrated samples of SecA1 and SecA2 were used in four different 96-well crystallisation screens. Although no crystal growth was recorded for SecA1, five different conditions resulted in SecA2 crystal growth. Of these conditions the best were PACT B10 (0.2 M magnesium chloride, 0.1 M MES pH 6.0, 20 % w/v PEG 6000) and C10 (0.2 M magnesium chloride 0.1 M HEPES pH 7.0, 20 % w/v PEG 6000). Using these crystals, diffraction data was collected to approximately 3.3 Å. Analysis of diffraction data was performed by Dr Paula Salgardo (The University of Newcastle). Optimisation of crystal growth was performed, including the addition of either AMP-PNP or ADP. Diffraction data was collected for three SecA2 isoforms. The most complete structure with the highest resolution resulting from these data was SecA2-ADP, however further refinement is required. **(B)** Although the model fits in the electron density map well in most areas, **(C)** other areas are poorly modelled. This may be due to bias introduced by using the *Thermotoga maritima* SecA structure for molecular replacement. **(D)** *M. tuberculosis* is the only organism to date for which the structures of both SecA1 and SecA2 have been resolved (Sharma et al., 2003; Swanson et al., 2015).The *C. difficile* SecA2 structure clearly contains the domains associated with SecA proteins; the walker box containing NBD1 (blue), NBD1 (cyan), PPXD (magenta), HSD (green), HWD (yellow), and IRA1 (orange). The two helices that form the VAR domain, absent in SecA2, are labelled (black arrows). Images were created using PyMOL molecular graphics system. SecA1 structural data was obtained from the protein data bank. PBD ID1NKT.

5.2.2 SecA2 Interacts with Members of the CWP family and Cell Wall Synthesis Machinery

SecA2K106, a mutant variant of the protein unable to hydrolyse ATP (Fagan and Fairweather, 2011), was purified as before. However, SDS-PAGE analysis of purified Strep-tagged SecA2K106R showed many protein contaminants in the sample. (Fig.5.3A). Protein samples were analysed via western blots, using anti-strep tag and anti-LMW SLP antibodies (Fig.5.3B). Strep-tagged SecA2K106R and the major SecA2 substrate, SlpA, were identified in the sample. As SecA2K106R is unable to hydrolyse ATP, it is likely trapped in the secretion complex, bound to its substrate and other proteins involved in secretion. The contaminating protein bands were therefore hypothesised to be SecA2 interacting proteins.

Peptide mass fingerprinting was used to identify the potential SecA2 binding partners. Mass spectrometry was performed by Dr Richard Beniston (The University of Sheffield BioMicS facility). Proteins of interest that were identified in the protein sample included eight members of the Cell Wall Protein family including the known SecA2 substrates SlpA and CwpV, SecA1, and the cell wall synthesis machinery; MreB2, MreB1, and FtsZ. Although this experiment proved useful for identifying potential SecA2 binding partners, it is worth noting that there seemed to be cross contamination between samples, likely due to insufficient washing between each analysis.
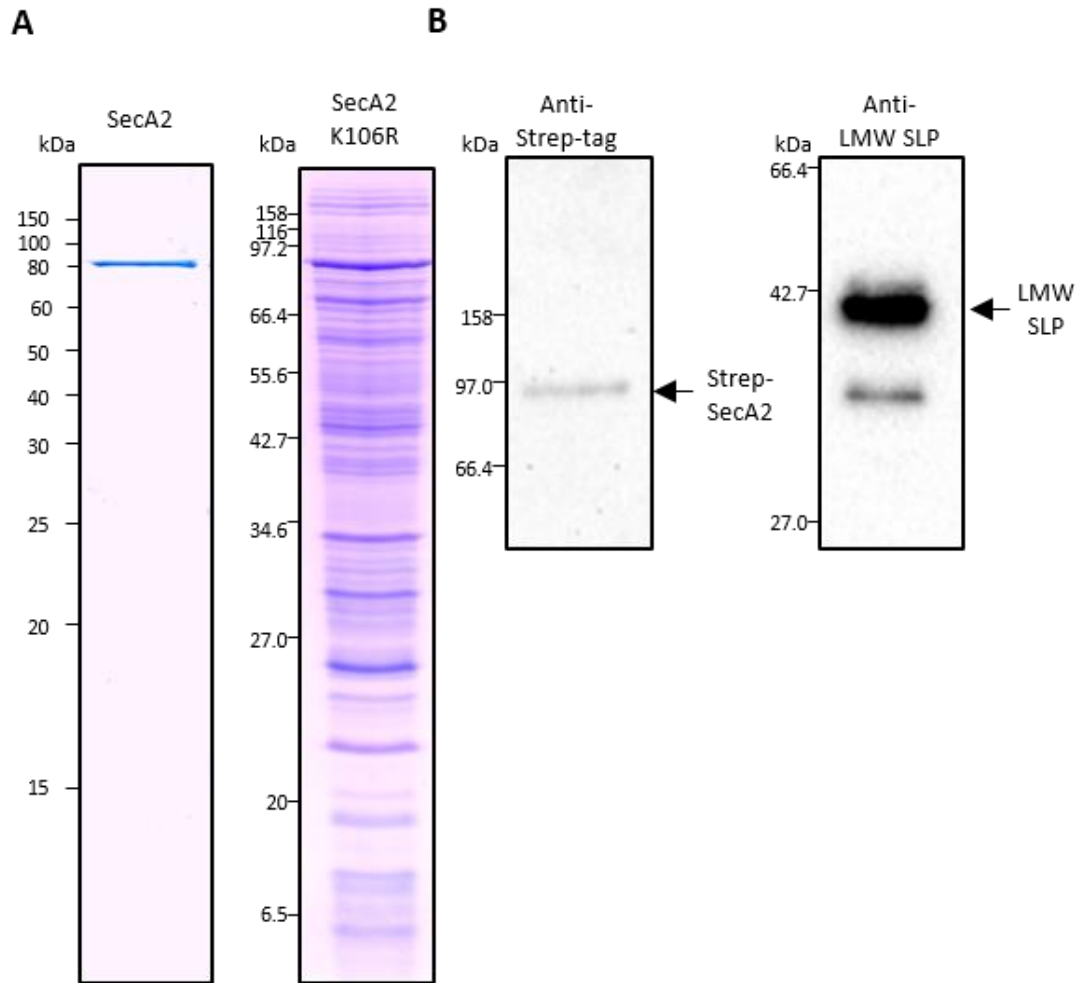
Fig.5.3: Pull down assay identifying SecA2 interacting proteins. (A) SecA2K106 was purified using Strep-tag affinity chromatography. Protein containing fractions were pooled and analysed using SDS-PAGE. When SecA2 is purified in this manner, the resulting protein solution is very pure. SecA2K106R preparations however, contain many contaminating proteins. **(B)** Western blot analysis of the SecA2K106 preparations confirms the presence of Strep-tagged SecA2K106R and SlpA (LMW SLP), the major substrate of SecA2.

5.2.3 MreB2 and SecA2 secretion

5.2.3.1 SecA2 Interaction Experiments

To investigate the potential interaction between SecA2 and the MreB proteins, bacterial two-hybrid experiments were performed. Interaction between T18-tagged bait proteins and T25-tagged prey protein results in functional complementation of adenylate cyclase (Karimova et al., 1998). Co-expression of these proteins in an *E. coli* strain that lacks an endogenous adenylate cyclase and contains a cAMP responsive *lacZ*, results in blue colonies when grown in the presence of X-gal. Proteins were N and C terminally tagged with either the T18 or T25 fragments of CyaA. *mreB1* and *mreB2* were amplified using primer pairs RF200/RF201 and RF196/RF197 respectively, and cloned into pUT18 between *Kpn*I and *Xba*I. *secA1* and *secA2* were amplified using primer pairs RF194/RF195 and RF192/RF193 respectively, and cloned into pKNT25 and pKT25 between *Kpn*I and *Xba*I. No interaction was observed between SecA1 or SecA2 with MreB1 or MreB2 using this method (data not shown).

A similar experiment was performed using split-SNAP. This method involves tagging proteins with the N (nSNAP) or C (cSNAP) terminal half of SNAP, a variant of human O^6-alkylguanine-DNA-alkyltransferase that binds to a small fluorescent substrate. Interaction between cSNAP and nSNAP-tagged proteins results in complementation of SNAP, enabling binding of the fluorescent marker (Mie et al., 2016). This method was chosen as it can be performed in *C. difficile.* This is beneficial as SecA2 may only bind to MreB when bound to its substrate, SlpA, not present in *E. coli*. pRPF150 encodes an N-terminally Strep-tagged SecA2. This plasmid was linearised using primer pair NF1957/NF1958 and ligated, to introduce an *Xho*I site between *secA2* and the strep-tag, resulting in pJAK012. SNAP was amplified from pFT46 using primer pair RF208/RF209 and was introduced into pJAK012 between *Sac*I and *Xho*I, resulting in pJAK013. Primer pair RF430/RF431 was used to amplify a 1.2 kb fragment, upstream of the *secA2* start codon, from 630 gDNA. Primer pair RF432/RF433 was used to amplify a 1.7 kb fragment from pJAK013

containing the SNAP tag coding sequence and 1.2kb downstream of the *secA2* start codon. These fragments were ligated into pMTLSC7315 using Gibson assembly. The resulting plasmid, pJAK047, was used to N-terminally SNAP tag chromosomal *secA2* (described below). To generate n-SNAP-secA2, *cSNAP* was deleted from pJAK047 by inverse PCR using RF464/RF466. The resulting plasmid, pJAK052, was used to insert *nSNAP* into the 5' end of chromosomal *secA2* via allele exchange mutagenesis. Expression of cSNAP fused prey proteins was induced in this strain (630 *nSNAP-secA2)* from plasmids. SNAP functional complementation was investigated using epifluorescence microscopy and in-gel fluorescence analysis (performed by Alice Lanne, personal communication). Again, no interaction was observed between SecA1 or SecA2 and MreB1 or MreB2 (data not shown). However, this method also failed to demonstrate an interaction between nSNAP-SecA2 and cSNAP-SecY suggesting that this method may not be suitable for investigating SecA2 protein-protein interaction (data not shown).

5.2.3.2 Creation of a Conditional MreB2 *C. difficile*

As direct interaction between SecA2 and MreB2 was not observed using the above methods, the effect of MreB perturbation on SlpA secretion was investigated. Creation of an *mreB2* conditional lethal *C. difficile* strain was attempted. *mreB2,* under control of a tetracycline inducible promoter, was inserted into the *pyrE* locus using allele exchange mutagenesis (Fig.5.4A and B). Attempts to knock-out the native *mreB2* were then performed in the resulting strain, 630 *pyrE::Ptet-mreB2.* However, despite numerous attempts a successful knock-out mutant was not isolated.

5.2.3.3 A22 Treatment affects SlpA secretion

A more indirect route to study the role of MreB in SlpA secretion was performed using A22. A22 is a small MreB perturbing compound that interacts with the MreB nucleotide-binding pocket, preventing ATP binding and MreB polymerisation (Bean et al., 2009). A22 was added to exponentially growing cultures of 630 and its effect on SlpA secretion was investigated

(performed by Alice Lanne, personal communication). Western blot analysis revealed an increase in the concentration of intracellular, unprocessed SlpA in the presence of A22 (Fig.5.4C), suggesting that MreB perturbation inhibits SlpA secretion. Interestingly, an increase in intracellular processed SlpA was also observed which suggests that Cwp84, the protease responsible for SlpA cleavage, is also secreted by SecA2.
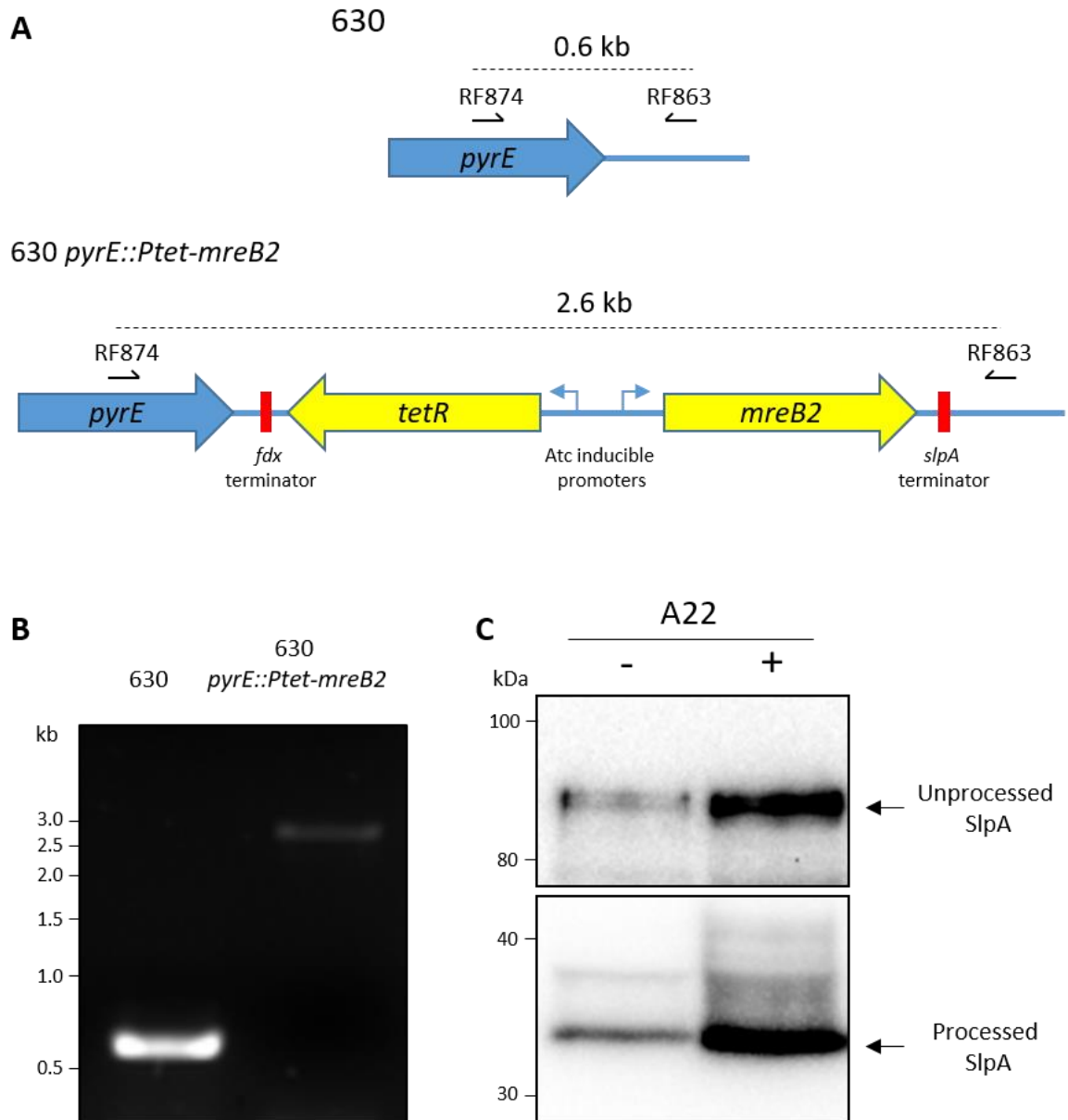
Fig.5.4: Investigating the role of MreB2 in SecA2 dependent secretion. (A) To investigate MreB2's role in SecA2 secretion, creation of a conditional lethal 630 strain was attempted. 1.2 kb upstream and downstream of the non-coding region downstream of *pyrE* were amplified from 630 gDNA using primer pairs RF837/RF838 and RF841/RF842. *mreB2* under control of a tetracycline inducible promotor was amplified from pAMBL019 using RF839/RF840. These three fragments were inserted into PCR linearised pMTLSC7315 by Gibson assembly. The resulting plasmid, pJAK080, was conjugated into 630 and allele exchange mutagenesis performed. **(B)** Mutants were screened via PCR using RF874/RF863. An increase in size of approximately 2 kb confirmed insertion of the conditional *mreB2* in the *pyrE* locus. Deletion of the native *mreB2* in the resulting strain was attempted, to create a conditional lethal strain. However, we failed to isolated a successful deletion mutant. Without a conditional-lethal mutant, more indirect methods for investigating the role of MreB2 in SecA2 dependent secretion were used. **(C)** 630 was grown

to approximately OD_{600nm} 0.6 and A22 (400 µg/ml) was added to the culture. After 1 h incubation, cells were treated with low pH glycine to remove the S-layer, then washed, and lysed. Cell lysates were then used in western blot analysis using anti-LMW SLP antibodies (performed by Alice Lanne, data unpublished). Bands corresponding to the Cwp84-processed and unprocessed SlpA are labelled.
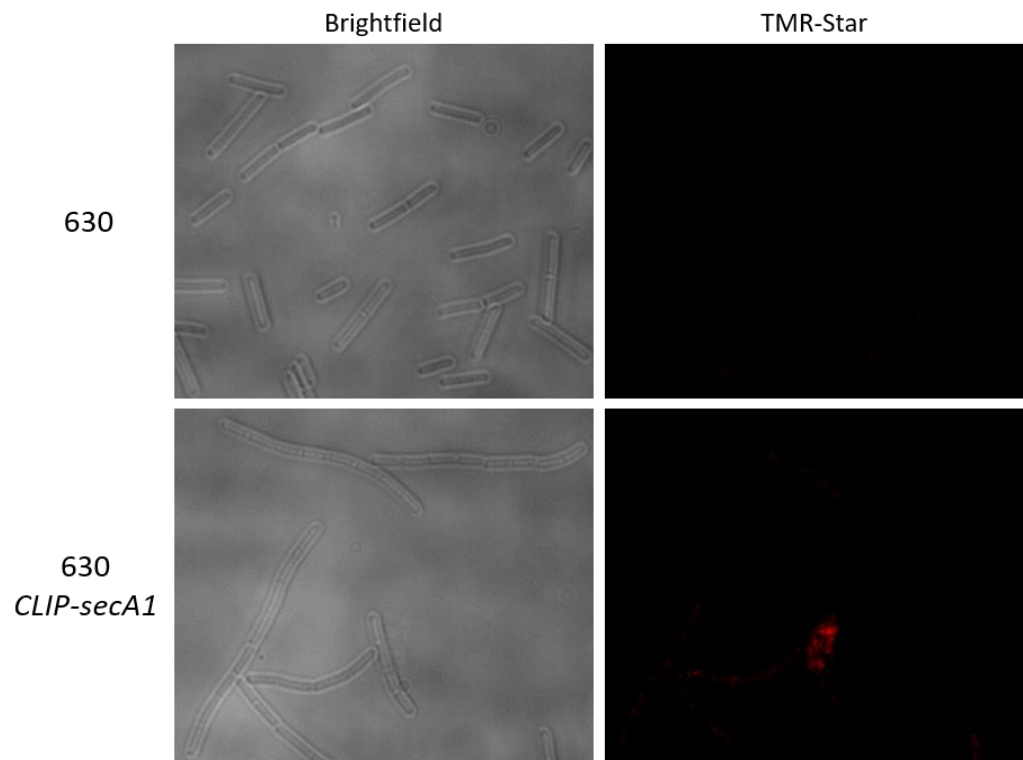
5.2.4 Localisation of SecA2 Secretion

5.2.4.1 SNAP and CLIP-Tagged SecA2 and SecA1

We hypothesised that SecA2-dependent secretion is localised to particular areas of the cell. To test this, SecA1 and SecA2 were tagged with fluorescent reporters and localisation of these proteins was investigated microscopically.

To determine the effect of tagging SecA2 on its function, N and C terminally tagged SecA2 was constitutively expressed in a conditional lethal 630 strain (630 *Ptet-secA2*). Although C-terminally tagged SecA2 did not fully restore growth in the mutant strain when grown without Atc, N-terminally tagged SecA2 did (data not shown). Due to the lack of a conditional-lethal *secA1* strain in which to test tagged protein function, N-terminal tagging was chosen for both proteins. Allele exchange mutagenesis was performed to add *SNAP* and *CLIP* to the 5' end of *secA2* and *secA1* on the chromosome respectively (Fig.5.5A and B), although mutants were difficult to isolate. These mutants grew normally in TY broth, 630 *SNAP-secA2* did show abnormal growth on solid media, growing as a mixture of large and small colonies.

Brightfield and epifluorescence microscopy was performed using these mutants. The morphology of 630 *SNAP-secA2* appeared normal and, although there was a strong signal from SNAP-SecA2, the protein did not seem to show any organised localisation (Fig.5.6). Cells of 630 *CLIP-secA1* however, formed long chains and little signal was observed from CLIP-SecA1. As 630 *SNAP-secA2* showed abnormal growth on agar and 630 *CLIP-secA1* showed altered cell morphology, tagged SecA1 and SecA2 may be only partially functional.

Fig.5.5: Creation of *630 SNAP-secA2* and *630 CLIP-secA1*. (A) *secA1* was amplified using primer pair RF234/RF235 and inserted into pJAK013 via the *Xho*I and *Bam*HI restriction sites, resulting in pJAK015. *C. difficile* codon optimised CLIP tag (Integrated DNA Technologies) was amplified using primer pair RF226/RF227 and inserted into pJAK015 via the *Sac*I and *Xho*I restriction sites, resulting in pJAK034. 1.7 kb, containing the CLIP tag and 1.2 kb of the *secA1* ORF, was amplified using RF436/RF437. 1.2 kb upstream of *secA1* was amplified using RF434/RF435. These fragments were assembled into RF311/RF312 linearised pMTLSC7315 using Gibson assembly. The resulting plasmid, pJAK049, and pJAK047 were used to add *CLIP* and *SNAP* with triple alanine linkers, to the chromosomal *secA1* and *secA2*, respectively, by allele exchange mutagenesis. **(B)** An increase of approximately 600 bp when analysed by PCR, using primers flanking the insert site, confirms successful insertion of the tags into the chromosome.

CLIP TMR-Star treated

	Brightfield	TMR-Star

630

630
CLIP-secA1

SNAP TMR-Star treated
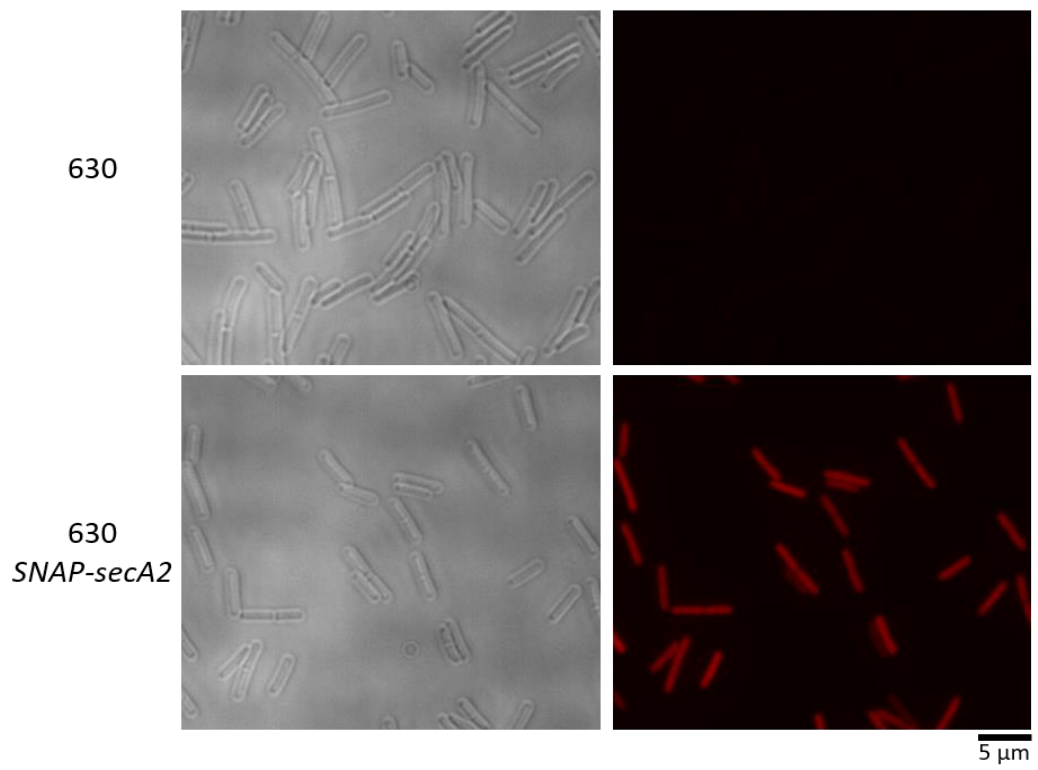
630

630
SNAP-secA2

5 µm

Fig.5.6: Microscopic analysis of SecA1 and SecA2 localisation. Brightfield and fluorescence microscopic analysis was performed using 630 *SNAP-secA2* and 630 *CLIP-secA1,* which revealed that although the cellular morphology of 630 *SNAP-secA2* appeared normal, 630 *CLIP-secA1* formed long chains. In addition, the signal from CLIP-SecA1 was very weak. SNAP-SecA2 however gave very strong signal, although no organised protein localisation was observed.

5.2.4.3 S-layer Biogenesis Co-localises to Areas of Cell Wall Synthesis

Although localisation studies focusing on the SecA proteins have resulted in non-conclusive results, the localisation of newly synthesised SlpA proved to be an easier target for study.

To determine if SlpA biogenesis is localised to areas of cell wall synthesis as hypothesised, a fluorescently labelled D-amino acid (HADA) was employed (Kuru et al., 2015). Growing *C. difficile* in the presence of HADA results in saturation of the cell wall with this fluorescent marker. When cells are washed of excess HADA and allowed to grow for 30 min, areas of newly synthesised cell wall are non-fluorescent (Fig.5.7A). pSEW027, a plasmid bearing SLCT-7 *slpA* under control of a tetracycline inducible promoter, was introduced into R20291. As the LMW SLP is variable amongst strains, antibodies against these subunits show high specificity. When induced for 5 minutes with Atc, newly secreted SLCT-7 SlpA was visualised using immunofluorescence microscopy, which resulted in punctate staining of the cell (Fig.5.7B).

Combination of these staining methods revealed co-localisation of cell wall synthesis with S-layer biogenesis. However, immunostaining of the cells was not efficient and further optimisation of this experiment is required to validate the results obtained.
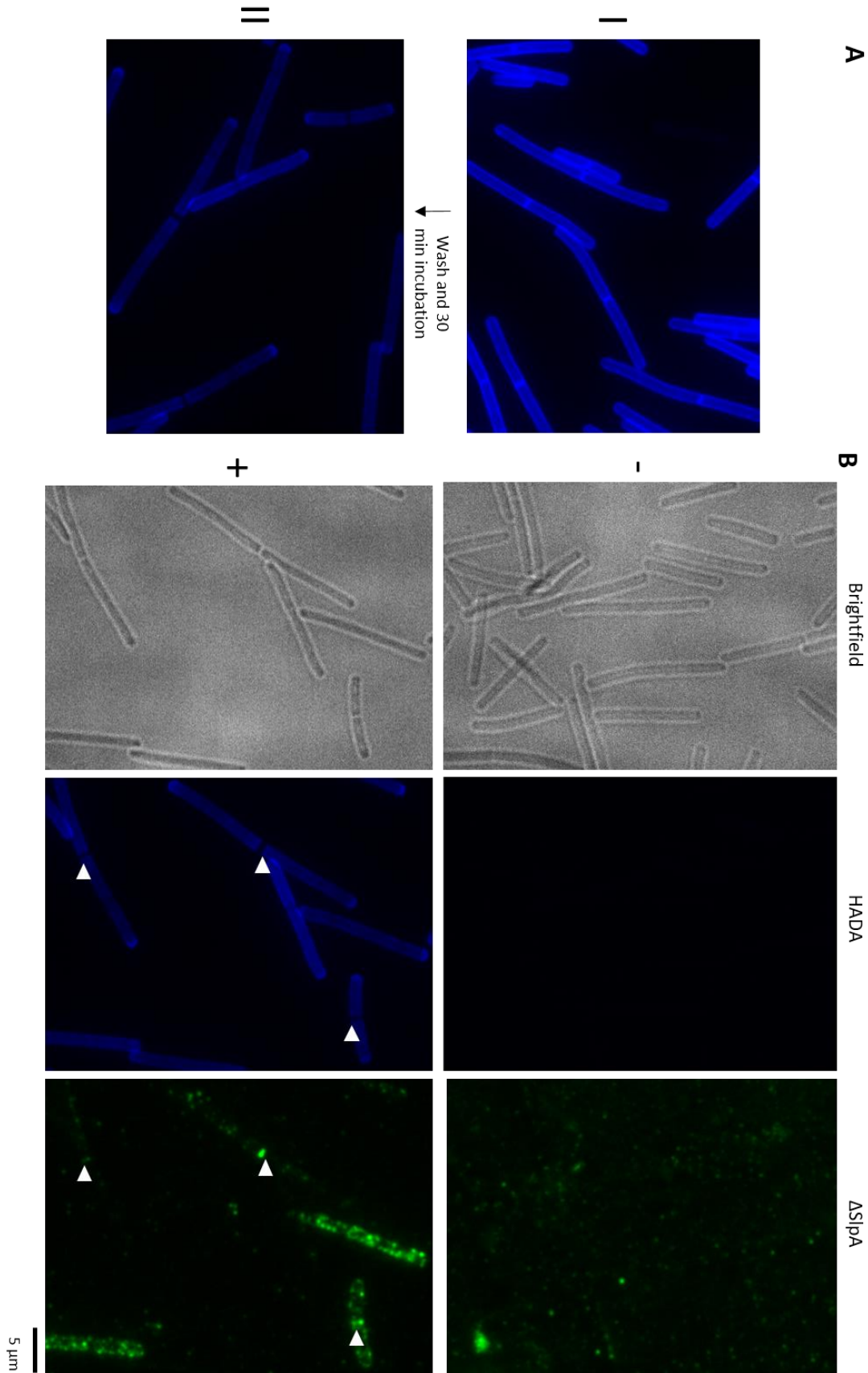
Fig.5.7: Co-localisation of cell wall synthesis and S-layer biogenesis. (A) Incubation of growing cultures of *C. difficile* with the fluorescently labelled D-amino acid HADA resulted in saturation of the cell wall with this fluorescent marker. These cells were then washed and allowed to grow for a further 30 minutes. Areas of new cell wall growth appeared dark, this was especially apparent at new septa. **(B)** R20291/pSEW027 was treated with HADA as described. 5 min prior to harvesting, SLCT-7 SlpA expression was induced using 100 ng/ml Atc. Harvested cells were fixed and immunostained using an anti SLCT-7 antibody. Newly created septa (white arrows) were obvious in the HADA images which also seemed to contain large amounts of newly secreted SLCT-7 SlpA, suggesting that cell wall synthesis and S-layer biogenesis are co-localised.

5.3 Discussion

As with many elements of *C. difficile* physiology, SecA2-dependent secretion and S-layer biogenesis are poorly understood. Using a combination of structural biology, microscopy, and biochemical analysis we have taken the first steps towards understanding the function of SecA2 and its role in S-layer biogenesis.

Over-expression of SecA2 and SecA1 in *C. difficile* allowed isolation of very pure samples of these proteins using a simple one-step affinity chromatography protocol. This also allowed isolation of these proteins from a native expression system. Although we were unable to crystallise SecA1, we have now solved the crystal structure of SecA2 in its holo and apo forms, the highest resolution structure being the ADP-bound form. When compared to SecA from other organisms, including *M. tuberculosis* for which both the SecA1 and SecA2 structures are known, SecA2 shares all structural features expected for a SecA protein, including the HWD, HSD, IRA1, HSD, DEAD box motor domain composed of NBD1 and NBD2, and PPXD domains although it lacks a CTD. In addition, NBD2 is truncated, lacking the pair of alpha helices that compose the VAR subdomain of other SecA proteins. The VAR domain forms a cap on the nucleotide binding site, suggesting that the nucleotide binding motif of SecA2 is more solvent-exposed. The VAR domain modulates the function of the DEAD motor box, loss of which results in rapid ADP dissociation and increased ATPase activity (Das et al., 2012). The PPXD of SecA2 is also rotated towards NBD2 relative to the positioning of the *Mtb*SecA1 PPXD. A similar lack of the VAR and C-terminal domains, and a shift in the positioning of the PPXD have also been observed for SecA2 from *Mycobacterium tuberculosis* (Swanson et al., 2015). These features may therefore be common characteristics of SecA2 proteins and may provide some insight into SecA2 function and specificity. However, as our model still requires refinement, the structure reported here is likely to change upon further scrutiny.

The purpose of having two independently functioning SecA proteins is not understood. As SecA2 likely uses the canonical SecY (Fagan and Fairweather, 2011), it is unlikely that the additional SecA is required to prevent saturation of the Sec secretion pathway with the highly abundant substrate SlpA. It is therefore logical to hypothesis that SecA2 is required for localised secretion of a small subset of proteins. The pull-down experiment using SecA2K106R supports this hypothesis. When this protein, unable to hydrolyse ATP, was isolated, several proteins were co-purified. This included the cell wall synthesis machinery MreB1, MreB2 and FtsZ. These experiments have since been repeated using both SecA2K106R and SecA1K106R (Peter Oatley, personal communication). FtsZ and MreB2 were only co-purified with SecA2K106, MreB1 was not identified. Tagged MreB has been reported to show abnormal behaviour in many studies, the most famous of which demonstrated that YFP-MreB mis-localises in *E. coli* (Margolin, 2012). It also appears that the N-terminal tags applied to *C. difficile* SecA proteins in this study result in partial loss of function. This may explain the failure of the split-SNAP and bacterial two-hybrid experiments to demonstrate an interaction between either MreB1 or MreB2 with SecA2. *In vitro* binding experiments such as thermophoresis or isothermal titration calorimetry may therefore be more suitable for studying the interactions between these proteins. Although split-SNAP and bacterial two-hybrid experiments failed to demonstrate an interaction between SecA2 and MreB2, A22 perturbation of the MreB cytoskeleton did inhibit SlpA secretion, suggesting MreB is involved in SecA2-dependent secretion. However, A22 is known to cause MreB-independent toxicity (Karczmarek et al., 2007) and the inhibition of SlpA synthesis cannot, therefore, conclusively be linked to MreB perturbation. A conditional-lethal MreB strain would be ideal to study the effect of MreB perturbation on SlpA secretion. Although attempts to create this mutant were unsuccessful, insertion of the entire operon containing MreB, under control of an inducible promoter, into the *pyrE* locus may be necessary to facilitate deletion of the native *mreB2*. Although localisation of SecA2 and SecA1 could not be directly observed, through a combination of immunofluorescence and HADA labelling of the cell wall, we have shown co-

localisation of cell wall synthesis and S-layer biogenesis. However, immunostaining of SlpA was not optimal and further study is necessary to validate these experiments.

In addition to the cell wall synthesis machinery, several members of the Cell Wall Protein family were also identified in the pull-down samples, including the only confirmed SecA2 substrates, SlpA and CwpV (Fagan and Fairweather, 2011). SecA2 dependent secretion may therefore be a characteristic of members of this protein family, or a small subset. Experiments performed by Peter Oatley also identified Cwp2 and Cwp6 in later SecA2K106R samples (personal communication). The presence of intracellular, processed SlpA in A22 treated samples may also suggests Cwp84 secretion is SecA2 dependent.

Work presented here represents a huge step in understanding the role of SecA2 and the mechanisms behind its specificity. The structure of SecA2, although requiring further refinement, is one of only two SecA2 structures reported to date and may provide insights into SecA2 function. This is also the first instance of the cell wall machinery being implicated in S-layer biogenesis, and further study will hopefully confirm this hypothesis.

Chapter VI

Discussion

The S-layer of *C. difficile* completely encases the cell and its protein constituents represent approximately 15% of the total protein content of the cell. It is therefore likely that S-layer biogenesis is associated with a high metabolic cost and it likely provides vital functions to the cell, especially as this structure is essential for *C. difficile* growth *in vitro (Dembek et al., 2015).* Exposing cultures of *C. difficile* to excess Avidocin-CD allowed the isolation of spontaneous mutants lacking a functional S-layer. The S-layer mutants and subsequent revertant strains have provided a unique opportunity to study the function of the *C. difficile* S-layer, showing the complex to be multi-functional (Fig.6.1). The properties of these mutants allowing their growth, whilst lacking a gene essential for growth in wild-type strains, remain elusive. It is likely, however, that these mutants contain suppressor mutations in their genome to allow their growth. The genome sequence of FM2.5 revealed several differences compared to the wild-type. Although many of these differences occur in genes of unknown function, a substitution within *codY* results in a change in the primary sequence of the resulting protein. CodY is a regulatory protein that regulates expression of toxins, sporulation and metabolism (Dineen et al., 2010; Nawrocki et al., 2016). This is a particularly interesting observation as we have demonstrated that FM2.5 has diminished ability to sporulate and produce toxins. However, as FM2.5RW does not show diminished sporulation or toxin production and likely contains the same mutation in *codY,* the phenotypes observed in FM2.5 are unlikely to be due to *codY*.

Avidocin-CDs use the S-layer as a binding receptor. As these structures are created using bacteriophage RBPs mined from the genomes of *C. difficile,* we have also shown that the S-layer acts as a receptor for bacteriophage binding. These findings suggest that antigenic variation of

SLCTs is largely due to selective pressure from bacteriophage predation rather than evasion of the host immune system and may help determine the host-range of *C. difficile* bacteriophage (Nale et al., 2016).

FM2.5 is avirulent in the hamster model of infection, likely due to decreased toxin production. Surprisingly FM2.5 was still able to persist in the gut, right up to the experimental end-point 14 days after initial infection. This observation suggests that SlpA does not function as a colonisation factor as previously suggested (Merrigan et al., 2013). Alternatively, the presence of several adhesins on the cell surface may represent functional redundancy, meaning loss of one adhesin form the surface may not result in loss of adherence to the host. FM2.5 is also able to persist in the gut despite showing increased sensitivity to killing by innate immune effectors such as LL-37 and lysozyme. LL-37 is present in the gut, being secreted by immune cells and gut epithelial cells, at concentrations of up to 5 µg/ml. This may be further increased in response to infection (Bowdish et al., 2005). Infection with FM2.5 also did not result in any symptoms of *C. difficile* disease. Taken together, these observations suggest that FM2.5 is unable to stimulate an immune response. As toxins can stimulate the immune system, causing release of inflammatory cytokines (Shen, 2012), the toxin synthesis defect displayed by FM2.5 will likely prevent host immune activation. SlpA is also able to stimulate the immune response via TLR4 (Ryan et al., 2011). Perhaps, therefore, loss of toxin production and/or SlpA loss is sufficient to prevent the excessive activation of the host immune system that characterises severe *C. difficile* infection.

Although it is difficult to speculate on the exact causes, FM2.5 shows significantly reduced sporulation rates. As the spore acts as the transmissive agent, it is likely that FM2.5 will show lower transmission rates, although this needs to be confirmed experimentally. As FM2.5 is avirulent, and has lower sporulation rates, it represents little clinical threat. This highlights the promise Avidocin-CDs show as treatment for CDI, as *C. difficile* must trade virulence for

resistance to these agents. As Avidocin-CDs have been shown to not affect the composition of the microbiota, the use of these agents as either prophylactics or treatment would allow the recovery of the microbiota from previous damage (Gebhart et al., 2015). It is unlikely that resistant mutants such as FM2.5 would be able to persist in the gut in the presence of a healthy microbiota, although this would be interesting to investigate experimentally. In addition, we have demonstrated that killing of *C. difficile* with Avidocin-CDs does not results in the release of intracellular toxin *in vitro.* This is an important observation as treatments that result in the release of toxin may exacerbate disease symptoms. However, as Avidocin-CDs kill *C. difficile* in a strain-specific manner, depending on the SLCT of the strain, treatment using these compounds poses a serious problem. To be effective treatments, either the SLCT of the infecting strain must be known, or a cocktail of different Avidocin-CDs covering all SLCTs must be administered. This highlights the need to develop Avidocin-CDs that target SLCT-3 and SLCT-5, which we currently lack. Ideally a broad range RBP would be used that interacts with all SLCTs, although due to the variation between SLCTs, discovery of such a RBP seems unlikely.

The isolation of FM2.5 also presents a unique opportunity to study the contributions of the different SLCTs to virulence. Although we have demonstrated that manipulation of the native *slpA* locus is difficult, further understanding of the *slpA* promotor region and SlpA expression may enable insertion of *slpA* under control of its native promotor into the *pyrE* locus of FM2.5, allowing deletion of the native *slpA.* This would allow insertion of any of the SLCT *slpA* into the genome of FM2.5, resulting in a library of isogenic strains expressing different SlpAs at native levels. This method would also allow the insertion of mutant variants of *slpA* which would allow us to study the contributions of the different components of the SlpA complex to virulence and its other functions.

Indeed, one mutant, whose creation would likely have not been possible without the isolation of FM2.5, allowed further insight into the mechanisms behind S-layer-dependent

resistance to lysozyme and LL-37. FM2.5RΔD2, presumed to contain an S-layer with larger pores than usual, shows increased sensitivity to lysozyme relative to wild-type but is resistant to LL-37. As lysozyme is so large, it is possible that the S-layer excludes this enzyme by acting as a physical barrier; the S-layer may not contain pores large enough for passage to the cell wall. Increasing the pore size of the S-layer may therefore allow lysozyme to pass through the S-layer. LL-37 on the other hand is a very small peptide. It is unlikely that the S-layer can exclude this molecule simply through size exclusion. It is more likely that, as with the LPS of Gram-negative bacteria, the S-layer is able to sequester this cationic peptide at the cell surface via ionic interactions. Increasing the pore size of the S-layer would therefore have little to no effect on the sensitivity of the cell to LL-37. It would be interesting to determine if overexpression of one of the CWPs, such as CwpV, could functionally complement the loss of SlpA in FM2.5 for resistance to innate immune effectors.

We have also demonstrated that SlpA contributes to the regulation of CwpV expression as FM2.5, but not FM2.5RW, shows increased CwpV expression due to larger proportion of the population containing the *cwpV* switch in the ON orientation. CwpV expression does not result in a growth benefit to FM2.5, which would otherwise result in higher proportions of CwpV ON cells within a population. These results suggest that CwpV expression is regulated by SlpA. It seems unlikely that extracellular SlpA could regulate phase switching and so I propose that intracellular SlpA down-regulates expression either through interaction with RecV or indirectly.

In addition to allowing investigation of S-layer function, isolation of FM2.5 and sequencing of its genome led to the identification of two additional RecV-dependent phase variable regions. One of these regions was upstream of *CDR20291_1514*, and has been previously suggested to be phase variable. *CDR20291_1514* encodes a c-di-GMP PDE which controls the turnover of the nucleotide secondary messenger, c-di-GMP, and its expression has previously been shown to increase motility in strain 630 (*CD630_1616*) (Bordeleau et al., 2011).

The phase variable region upstream of *CDR20291_3128* has not been previously described. *CDR20291_3128* is the first gene in an operon that encodes a histidine sensor kinase and two response regulators likely to form a three-component regulatory system. Although the function of this three-component system is unknown, its expression seems to affect colony morphology of *C. difficile.* Experiments to determine the effect of expression of this operon on colony morphology are required to confirm this hypothesis. However, colony morphology is often associated with changes in motility. CwpV expression causes aggregation (Reynolds et al., 2011), expression of the *flgB* operon is essential for motility (Anjuwon-Foster and Tamayo, 2017), and c-di-GMP is a secondary messenger that regulates motility and aggregation (Purcell et al., 2012). Therefore, we hypothesis that RecV-dependent phase variation is involved in the regulation of genes responsible for switching from sessile to motile lifestyles in *C. difficile.* During infection *C. difficile* switches from a motile to a surface adherent state, and it would be interesting to determine whether RecV-dependent phase variation is involved in this change and if RecV activity is affected by c-di-GMP.

Another interesting question posed by the identification of these additional phase variable regions is how RecV recognises their different inverted repeats and why their regulation is different. One possibility is the presence of specific co-factors required for RecV activity. Pull down experiments using RecV may be used to identify RecV binding partners.

As the S-layer is bound to the cell wall, growth of the cell wall would result in the formation of gaps in the S-layer which could potentially be fatal to the cell as this structure is essential for growth (Dembek et al., 2015). We propose a new model for S-layer biosynthesis in which interaction of SecA2 with MreB or FtsZ localises SecA2-dependent secretion, and therefore S-layer biogenesis, to areas of peptidoglycan growth (Fig.6.1). This localisation of S-layer biogenesis was observed microscopically. Further optimisation of these experiments and utilisation of super-resolution methodologies will allow more accurate observations of S-layer

growth *in situ*. Investigation of the localisation of SecA1 secretion was not performed and secretion of SecA1 substrates may also be localised to areas of cell wall growth, as these areas have transient gaps in the S-layer. This may be required for large secreted proteins to pass through the S-layer.

We were unable to show direct interaction between SecA2 and MreB or FtsZ of the cell wall synthesis machinery. This may be due to the methods employed to investigate these interactions which are dependent on tagging the proteins of interest which can result in loss of function or cause steric hindrance at the interaction interfaces. *In vitro* investigation into the interactions between SecA2 and its binding partners may therefore be better suited.

Using X-ray crystallography, we have also created a 3D model of SecA2, the second SecA2 structure to date (Swanson et al., 2015). Interestingly, this structure demonstrated structural difference to *Mtb*SecA1, similar to *Mtb*SecA2, including loss of the VAR domain and a spatial rearrangement of the PPXD domain. However, unlike *Mtb*SecA2, the SecA2 of *C. difficile* does contain a HWD. Although the significance of these changes are not understood, it is likely that they contribute to the specificity of these proteins for their substrates.

In conclusion, we have now identified several functions of the S-layer including sporulation, toxin production, virulence, and growth. In addition, we also propose a new model for SecA2 dependent secretion, which may be a characteristic of members of the CWP family, in which its secretion is localised to areas of new peptidoglycan synthesis.
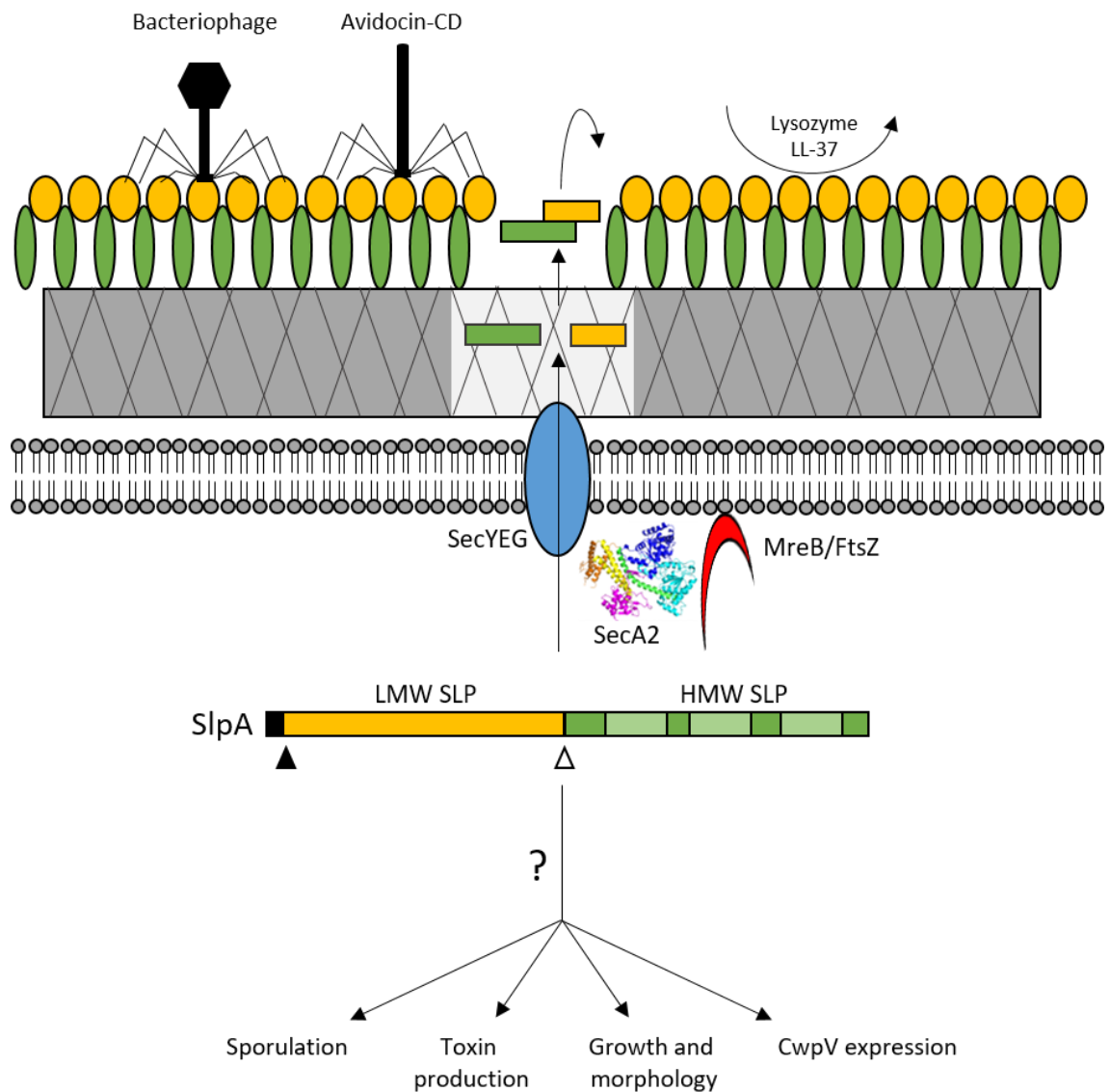
Fig.6.1: Biogenesis and function of the *C. difficile* S-layer. The S-layer is the receptor for binding of Avidocin-CDs and some bacteriophages, and its loss results in resistance to Avidocin-CDs. The S-layer also provides resistance to the innate immune effectors LL-37 and lysozyme. The mechanisms behind these resistances are unclear but are likely due to size exclusion and sequestering the antimicrobials at the cell surface. Loss of SlpA results in diminished sporulation efficiency, toxin production, altered growth and increased CwpV expression. The reasons for these observed phenotypes are unknown. SlpA is secreted by the accessory secretion pathway. This secretion is localised to areas of cell wall growth (light grey area) via interaction of SecA2, for which the structure is now known, with the cell wall synthesis machinery MreB and FtsZ. The S-layer is attached to the cell wall and is assembled in the gaps formed by cell wall growth.

References

Ahn, J.S., Chandramohan, L., Liou, L.E., and Bayles, K.W. (2006). Characterization of CidR-mediated regulation in *Bacillus anthracis* reveals a previously undetected role of S-layer proteins as murein hydrolases. Mol Microbiol *62*, 1158-1169.

Aktories, K., and Barth, H. (2004). *Clostridium botulinum* C2 toxin--new insights into the cellular up-take of the actin-ADP-ribosylating toxin. Int J Med Microbiol *293*, 557-564.

Alang, N., and Kelly, C.R. (2015). Weight gain after fecal microbiota transplantation. Open Forum Infect Dis *2*, ofv004.

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

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

Awram, P., and Smit, J. (1998). The *Caulobacter crescentus* paracrystalline S-layer protein is secreted by an ABC transporter (type I) secretion apparatus. J Bacteriol *180*, 3062-3069.

Awram, P., and Smit, J. (2001). Identification of lipopolysaccharide O antigen synthesis genes required for attachment of the S-layer of *Caulobacter crescentus*. Microbiology *147*, 1451-1460.

Bacci, S., Molbak, K., Kjeldsen, M.K., and Olsen, K.E. (2011). Binary toxin and death after *Clostridium difficile* infection. Emerg Infect Dis *17*, 976-982.

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

Band, V.I., and Weiss, D.S. (2015). Mechanisms of Antimicrobial Peptide Resistance in Gram-Negative Bacteria. Antibiotics (Basel) *4*, 18-41.

Banerji, O. (2017). Structural studies of the *Clostridium difficile* surface layer. In Molecular biology and Biotechnology (The University of Sheffield).

Baranova, E., Fronzes, R., Garcia-Pino, A., Van Gerven, N., Papapostolou, D., Pehau-Arnaudet, G., Pardon, E., Steyaert, J., Howorka, S., and Remaut, H. (2012). SbsB structure and lattice reconstruction unveil Ca2+ triggered S-layer assembly. Nature *487*, 119-122.

Bartlett, J.G. (1994). *Clostridium difficile*: history of its role as an enteric pathogen and the current state of knowledge about the organism. Clin Infect Dis *18 Suppl 4*, S265-272.

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

Bartsch, S.M., Umscheid, C.A., Fishman, N., and Lee, B.Y. (2013). Is fidaxomicin worth the cost? An economic analysis. Clin Infect Dis *57*, 555-561.

Bean, G.J., Flickinger, S.T., Westler, W.M., McCully, M.E., Sept, D., Weibel, D.B., and Amann, K.J. (2009). A22 disrupts the bacterial actin cytoskeleton by directly binding and inducing a low-affinity state in MreB. Biochemistry *48*, 4852-4857.

Bensing, B.A., Gibson, B.W., and Sullam, P.M. (2004). The *Streptococcus gordonii* platelet binding protein GspB undergoes glycosylation independently of export. J Bacteriol *186*, 638-645.

Bharat, T.A.M., Kureisaite-Ciziene, D., Hardy, G.G., Yu, E.W., Devant, J.M., Hagen, W.J.H., Brun, Y.V., Briggs, J.A.G., and Lowe, J. (2017). Structure of the hexagonal surface layer on *Caulobacter crescentus* cells. Nat Microbiol *2*, 17059.

Blaser, M.J. (1993). Role of the S-layer proteins of *Campylobacter fetus* in serum-resistance and antigenic variation: a model of bacterial pathogenesis. Am J Med Sci *306*, 325-329.

Bordeleau, E., and Burrus, V. (2015). Cyclic-di-GMP signaling in the Gram-positive pathogen *Clostridium difficile*. Curr Genet *61*, 497-502.

Bordeleau, E., Fortier, L.C., Malouin, F., and Burrus, V. (2011). c-di-GMP turn-over in *Clostridium difficile* is controlled by a plethora of diguanylate cyclases and phosphodiesterases. PLoS Genet *7*, e1002039.

Bouwknegt, M., van Dorp, S., and Kuijper, E. (2015). Burden of *Clostridium difficile* infection in the United States. N Engl J Med *372*, 2368.

Bowdish, D.M., Davidson, D.J., Lau, Y.E., Lee, K., Scott, M.G., and Hancock, R.E. (2005). Impact of LL-37 on anti-infective immunity. J Leukoc Biol *77*, 451-459.

Bradshaw, W.J., Kirby, J.M., Roberts, A.K., Shone, C.C., and Acharya, K.R. (2017). Cwp2 from *Clostridium difficile* exhibits an extended three domain fold and cell adhesion *in vitro*. FEBS J.

Brahamsha, B. (1996). An abundant cell-surface polypeptide is required for swimming by the nonflagellated marine *Cyanobacterium Synechococcus*. Proc Natl Acad Sci U S A *93*, 6504-6509.

Brandt, L.J., Aroniadis, O.C., Mellow, M., Kanatzar, A., Kelly, C., Park, T., Stollman, N., Rohlke, F., and Surawicz, C. (2012). Long-Term Follow-Up of Colonoscopic Fecal Microbiota Transplant for Recurrent *Clostridium difficile* Infection. Am J Gastroenterol *107*, 1079-1087.

Braunstein, M., Espinosa, B.J., Chan, J., Belisle, J.T., and Jacobs, W.R., Jr. (2003). SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. Mol Microbiol *48*, 453-464.

Breukink, E., Nouwen, N., van Raalte, A., Mizushima, S., Tommassen, J., and de Kruijff, B. (1995). The C terminus of SecA is involved in both lipid binding and SecB binding. J Biol Chem *270*, 7902-7907.

Burke, K.E., and Lamont, J.T. (2014). *Clostridium difficile* infection: a worldwide disease. Gut Liver *8*, 1-6.

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

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

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

CDC (2013). Antibiotic resistance threaths in the United States, C.f.D.C.a. Prevention, ed.

Chumbler, N.M., Rutherford, S.A., Zhang, Z., Farrow, M.A., Lisher, J.P., Farquhar, E., Giedroc, D.P., Spiller, B.W., Melnyk, R.A., and Lacy, D.B. (2016). Crystal structure of *Clostridium difficile* toxin A. Nat Microbiol *1*, 15002.

Chung, S., Shin, S.H., Bertozzi, C.R., and De Yoreo, J.J. (2010). Self-catalyzed growth of S-layers via an amorphous-to-crystalline transition limited by folding kinetics. Proc Natl Acad Sci U S A *107*, 16536-16541.

Clemente, J.C., Ursell, L.K., Parfrey, L.W., and Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. Cell *148*, 1258-1270.

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.*, et al.* (2016). The binary toxin CDT enhances Clostridium difficile virulence by suppressing protective colonic eosinophilia. Nat Microbiol *1*, 16108.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Curry, S.R., Marsh, J.W., Muto, C.A., O'Leary, M.M., Pasculle, A.W., and Harrison, L.H. (2007). *tcdC* genotypes associated with severe TcdC truncation in an epidemic clone and other strains of *Clostridium difficile*. J Clin Microbiol *45*, 215-221.

Dalbey, R.E., and Wickner, W. (1985). Leader peptidase catalyzes the release of exported proteins from the outer surface of the *Escherichia coli* plasma membrane. J Biol Chem *260*, 15925-15931.

Das, S., Grady, L.M., Michtavy, J., Zhou, Y., Cohan, F.M., Hingorani, M.M., and Oliver, D.B. (2012). The variable subdomain of *Escherichia coli* SecA functions to regulate SecA ATPase activity and ADP release. J Bacteriol *194*, 2205-2213.

de la Fuente-Nunez, C., Mertens, J., Smit, J., and Hancock, R.E. (2012). The bacterial surface layer provides protection against antimicrobial peptides. Appl Environ Microbiol *78*, 5452-5456.

de la Riva, L., Willing, S.E., Tate, E.W., and Fairweather, N.F. (2011). Roles of cysteine proteases Cwp84 and Cwp13 in biogenesis of the cell wall of *Clostridium difficile*. J Bacteriol *193*, 3276-3285.

Deakin, L.J., Clare, S., Fagan, R.P., Dawson, L.F., Pickard, D.J., West, M.R., Wren, B.W., Fairweather, N.F., Dougan, G., and Lawley, T.D. (2012). The *Clostridium difficile* spo0A gene is a persistence and transmission factor. Infect Immun *80*, 2704-2711.

Debast, S.B., Bauer, M.P., Kuijper, E.J., European Society of Clinical, M., and Infectious, D. (2014). European Society of Clinical Microbiology and Infectious Diseases: update of the treatment guidance document for *Clostridium difficile* infection. Clin Microbiol Infect *20 Suppl 2*, 1-26.

Dembek, M., Barquist, L., Boinett, C.J., Cain, A.K., Mayho, M., Lawley, T.D., Fairweather, N.F., and Fagan, R.P. (2015). High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. MBio *6*, e02383.

Dembek, M., Reynolds, C.B., and Fairweather, N.F. (2012). *Clostridium difficile* cell wall protein CwpV undergoes enzyme-independent intramolecular autoproteolysis. J Biol Chem *287*, 1538-1544.

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

Dingle, K.E., Didelot, X., Ansari, M.A., Eyre, D.W., Vaughan, A., Griffiths, D., Ip, C.L., Batty, E.M., Golubchik, T., Bowden, R.*, et al.* (2013). Recombinational switching of the *Clostridium difficile* S-layer and a novel glycosylation gene cluster revealed by large-scale whole-genome sequencing. J Infect Dis *207*, 675-686.

Drudy, D., Fanning, S., and Kyne, L. (2007). Toxin A-negative, toxin B-positive *Clostridium difficile*. Int J Infect Dis *11*, 5-10.

Dupuy, B., Govind, R., Antunes, A., and Matamouros, S. (2008). *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. J Med Microbiol *57*, 685-689.

Dupuy, B., and Sonenshein, A.L. (1998). Regulated transcription of *Clostridium difficile* toxin genes. Mol Microbiol *27*, 107-120.

Dworkin, J., Tummuru, M.K., and Blaser, M.J. (1995). A lipopolysaccharide-binding domain of the *Campylobacter fetus* S-layer protein resides within the conserved N terminus of a family of silent and divergent homologs. J Bacteriol *177*, 1734-1741.

ECDC (2013). Point prevelence survey of healthcare-associated infections and antimicrobial use in European acute hospitals 2011-2012 (European Centre for Disease Prevention and Control).

Economou, A., and Wickner, W. (1994). SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. Cell *78*, 835-843.

Egea, P.F., and Stroud, R.M. (2010). Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. Proc Natl Acad Sci U S A *107*, 17182-17187.

Eiseman, B., Silen, W., Bascom, G.S., and Kauvar, A.J. (1958). Fecal enema as an adjunct in the treatment of pseudomembranous enterocolitis. Surgery *44*, 854-859.

El Meouche, I., Peltier, J., Monot, M., Soutourina, O., Pestel-Caron, M., Dupuy, B., and 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*, e83748.

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

Engelhardt, H. (2007). Are S-layers exoskeletons? The basic function of protein surface layers revisited. J Struct Biol *160*, 115-124.

Erlandson, K.J., Miller, S.B., Nam, Y., Osborne, A.R., Zimmer, J., and Rapoport, T.A. (2008). A role for the two-helix finger of the SecA ATPase in protein translocation. Nature *455*, 984-987.

Errington, J. (2015). Bacterial morphogenesis and the enigmatic MreB helix. Nat Rev Microbiol *13*, 241-248.

Estrem, S.T., Gaal, T., Ross, W., and Gourse, R.L. (1998). Identification of an UP element consensus sequence for bacterial promoters. Proc Natl Acad Sci U S A *95*, 9761-9766.

Etienne-Toumelin, I., Sirard, J.C., Duflot, E., Mock, M., and Fouet, A. (1995). Characterization of the *Bacillus anthracis* S-layer: cloning and sequencing of the structural gene. J Bacteriol *177*, 614-620.

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

Fagan, R.P., and Fairweather, N.F. (2011). *Clostridium difficile* has two parallel and essential Sec secretion systems. J Biol Chem *286*, 27483-27493.

Fagan, R.P., and Fairweather, N.F. (2014). Biogenesis and functions of bacterial S-layers. Nat Rev Microbiol *12*, 211-222.

Fagan, R.P., Janoir, C., Collignon, A., Mastrantonio, P., Poxton, I.R., and Fairweather, N.F. (2011). A proposed nomenclature for cell wall proteins of *Clostridium difficile*. J Med Microbiol *60*, 1225-1228.

Feltcher, M.E., and Braunstein, M. (2012). Emerging themes in SecA2-mediated protein export. Nat Rev Microbiol *10*, 779-789.

Figueroa, I., Johnson, S., Sambol, S.P., Goldstein, E.J., Citron, D.M., and Gerding, D.N. (2012). Relapse versus reinfection: recurrent *Clostridium difficile* infection following treatment with fidaxomicin or vancomycin. Clin Infect Dis *55 Suppl 2*, S104-109.

Fimlaid, K.A., Bond, J.P., Schutz, K.C., Putnam, E.E., Leung, J.M., Lawley, T.D., and Shen, A. (2013). Global analysis of the sporulation pathway of *Clostridium difficile*. PLoS Genet *9*, e1003660.

Ford, M.J., Nomellini, J.F., and Smit, J. (2007). S-layer anchoring and localization of an S-layer-associated protease in *Caulobacter crescentus*. J Bacteriol *189*, 2226-2237.

Fouet, A., Mesnage, S., Tosi-Couture, E., Gounon, P., and Mock, M. (1999). *Bacillus anthracis* surface: capsule and S-layer. J Appl Microbiol *87*, 251-255.

Francis, M.B., Allen, C.A., Shrestha, R., and Sorg, J.A. (2013). Bile acid recognition by the *Clostridium difficile* germinant receptor, CspC, is important for establishing infection. PLoS Pathog *9*, e1003356.

Freeman, J., Bauer, M.P., Baines, S.D., Corver, J., Fawley, W.N., Goorhuis, B., Kuijper, E.J., and Wilcox, M.H. (2010). The changing epidemiology of *Clostridium difficile* infections. Clin Microbiol Rev *23*, 529-549.

Gebhart, D., Lok, S., Clare, S., Tomas, M., Stares, M., Scholl, D., Donskey, C.J., Lawley, T.D., and Govoni, G.R. (2015). A modified R-type bacteriocin specifically targeting *Clostridium difficile* prevents colonization of mice without affecting gut microbiota diversity. MBio *6*.

Gebhart, D., Williams, S.R., Bishop-Lilly, K.A., Govoni, G.R., Willner, K.M., Butani, A., Sozhamannan, S., Martin, D., Fortier, L.C., and Scholl, D. (2012). Novel high-molecular-weight, R-type bacteriocins of *Clostridium difficile*. J Bacteriol *194*, 6240-6247.

George, W.L., Sutter, V.L., Goldstein, E.J., Ludwig, S.L., and Finegold, S.M. (1978). Aetiology of antimicrobial-agent-associated colitis. Lancet *1*, 802-803.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.

Goorhuis, A., Bakker, D., Corver, J., Debast, S.B., Harmanus, C., Notermans, D.W., Bergwerff, A.A., Dekker, F.W., and Kuijper, E.J. (2008). Emergence of *Clostridium difficile* infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clin Infect Dis *47*, 1162-1170.

Green, E.R., and Mecsas, J. (2016). Bacterial Secretion Systems: An Overview. Microbiol Spectr *4*.

Gupta, A., and Khanna, S. (2014). Community-acquired *Clostridium difficile* infection: an increasing public health threat. Infect Drug Resist *7*, 63-72.

Gupta, S., Allen-Vercoe, E., and Petrof, E.O. (2016). Fecal microbiota transplantation: in perspective. Therap Adv Gastroenterol *9*, 229-239.

Gweon, T.G., Lee, K.J., Kang, D.H., Park, S.S., Kim, K.H., Seong, H.J., Ban, T.H., Moon, S.J., Kim, J.S., and Kim, S.W. (2015). A case of toxic megacolon caused by *Clostridium difficile* infection and treated with fecal microbiota transplantation. Gut Liver *9*, 247-250.

Hall, A. (2012). Rho family GTPases. Biochem Soc Trans *40*, 1378-1382.

Hall, I., and O'Toole, E. (1935). Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, *Bacillus difficilis*. Am J Dis Child *49*, 390.

Hammond, G.A., and Johnson, J.L. (1995). The toxigenic element of *Clostridium difficile* strain VPI 10463. Microb Pathog *19*, 203-213.

He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D.J., Martin, M.J., Connor, T.R., Harris, S.R., Fairley, D., Bamford, K.B.*, et al.* (2013). Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. Nat Genet *45*, 109-113.

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

Higgins, D., and Dworkin, J. (2012). Recent progress in *Bacillus subtilis* sporulation. FEMS Microbiol Rev *36*, 131-148.

Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., and Pease, L.R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene *77*, 61-68.

Houwink, A.L. (1953). A macromolecular mono-layer in the cell wall of *Spirillum* spec. Biochim Biophys Acta *10*, 360-366.

Hunt, J.F., Weinkauf, S., Henry, L., Fak, J.J., McNicholas, P., Oliver, D.B., and Deisenhofer, J. (2002). Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. Science *297*, 2018-2026.

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

Johnson, S. (2009). Recurrent *Clostridium difficile* infection: a review of risk factors, treatments, and outcomes. J Infect *58*, 403-410.

Johnson, S., Louie, T.J., Gerding, D.N., Cornely, O.A., Chasan-Taber, S., Fitts, D., Gelone, S.P., Broom, C., Davidson, D.M., and Polymer Alternative for, C.D.I.T.i. (2014). Vancomycin, metronidazole, or tolevamer for *Clostridium difficile* infection: results from two multinational, randomized, controlled trials. Clin Infect Dis *59*, 345-354.

Johnson, S., Samore, M.H., Farrow, K.A., Killgore, G.E., Tenover, F.C., Lyras, D., Rood, J.I., DeGirolami, P., Baltch, A.L., Rafferty, M.E., *et al.* (1999). Epidemics of diarrhea caused by a clindamycin-resistant strain of *Clostridium difficile* in four hospitals. N Engl J Med *341*, 1645-1651.

Karczmarek, A., Martinez-Arteaga, R., Alexeeva, S., Hansen, F.G., Vicente, M., Nanninga, N., and den Blaauwen, T. (2007). DNA and origin region segregation are not affected by the transition from rod to sphere after inhibition of *Escherichia coli* MreB by A22. Mol Microbiol *65*, 51-63.

Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc Natl Acad Sci U S A *95*, 5752-5756.

Karlsson, S., Burman, L.G., and Akerlund, T. (1999). Suppression of toxin production in *Clostridium difficile* VPI 10463 by amino acids. Microbiology *145 (Pt 7)*, 1683-1693.

Kern, V.J., Kern, J.W., Theriot, J.A., Schneewind, O., and Missiakas, D. (2012). Surface-layer (S-layer) proteins sap and EA1 govern the binding of the S-layer-associated protein BslO at the cell septa of *Bacillus anthracis*. J Bacteriol *194*, 3833-3840.

Khanna, S., Pardi, D.S., Aronson, S.L., Kammer, P.P., Orenstein, R., St Sauver, J.L., Harmsen, W.S., and Zinsmeister, A.R. (2012). The epidemiology of community-acquired *Clostridium difficile* infection: a population-based study. Am J Gastroenterol *107*, 89-95.

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

Kirk, J.A., Banerji, O., and Fagan, R.P. (2017). Characteristics of the *Clostridium difficile* cell envelope and its importance in therapeutics. Microbial biotechnology *10*, 1-220.

Kirk, J.A., and 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. Science Translational Medicine.

Koval, S.F., and Hynes, S.H. (1991). Effect of paracrystalline protein surface layers on predation by *Bdellovibrio bacteriovorus*. J Bacteriol *173*, 2244-2249.

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

Kuehne, S.A., Collery, M.M., Kelly, M.L., Cartman, S.T., Cockayne, A., and Minton, N.P. (2014). Importance of toxin A, toxin B, and CDT in virulence of an epidemic Clostridium difficile strain. J Infect Dis *209*, 83-86.

Kuru, E., Tekkam, S., Hall, E., Brun, Y.V., and Van Nieuwenhze, M.S. (2015). Synthesis of fluorescent D-amino acids and their use for probing peptidoglycan synthesis and bacterial growth *in situ*. Nat Protoc *10*, 33-52.

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

Lawley, T.D., Clare, S., Walker, A.W., Stares, M.D., Connor, T.R., Raisen, C., Goulding, D., Rad, R., Schreiber, F., Brandt, C*., et al.* (2012). Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing *Clostridium difficile* disease in mice. PLoS Pathog *8*, e1002995.

Lessa, F.C., Mu, Y., Bamberg, W.M., Beldavs, Z.G., Dumyati, G.K., Dunn, J.R., Farley, M.M., Holzbauer, S.M., Meek, J.I., Phipps, E.C*., et al.* (2015). Burden of *Clostridium difficile* infection in the United States. N Engl J Med *372*, 825-834.

Louie, T.J., Cannon, K., Byrne, B., Emery, J., Ward, L., Eyben, M., and Krulicki, W. (2012). Fidaxomicin preserves the intestinal microbiome during and after treatment of *Clostridium difficile* infection (CDI) and reduces both toxin reexpression and recurrence of CDI. Clin Infect Dis *55 Suppl 2*, S132-142.

Lyerly, D.M., Saum, K.E., MacDonald, D.K., and Wilkins, T.D. (1985). Effects of *Clostridium difficile* toxins given intragastrically to animals. Infect Immun *47*, 349-352.

Lynch, M., Walsh, T.A., Marszalowska, I., Webb, A.E., MacAogain, M., Rogers, T.R., Windle, H., Kelleher, D., O'Connell, M.J., and Loscher, C.E. (2017). Surface layer proteins from virulent *Clostridium difficile* ribotypes exhibit signatures of positive selection with consequences for innate immune response. BMC Evol Biol *17*, 90.

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*., et al.* (2009). Toxin B is essential for virulence of *Clostridium difficile*. Nature *458*, 1176-1179.

Margolin, W. (2012). The price of tags in protein localization studies. J Bacteriol *194*, 6369-6371.

Martin, J.S., Monaghan, T.M., and Wilcox, M.H. (2016). *Clostridium difficile* infection: epidemiology, diagnosis and understanding transmission. Nat Rev Gastroenterol Hepatol *13*, 206-216.

Mayer, M.J., Narbad, A., and Gasson, M.J. (2008). Molecular characterization of a *Clostridium difficile* bacteriophage and its cloned biologically active endolysin. J Bacteriol *190*, 6734-6740.

McCarren, J., and Brahamsha, B. (2007). SwmB, a 1.12-megadalton protein that is required for nonflagellar swimming motility in *Synechococcus*. J Bacteriol *189*, 1158-1162.

McCarren, J., and Brahamsha, B. (2009). Swimming motility mutants of marine *Synechococcus* affected in production and localization of the S-layer protein SwmA. J Bacteriol *191*, 1111-1114.

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

Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M., and Fouet, A. (2000). Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO J *19*, 4473-4484.

Mesnage, S., Tosi-Couture, E., Mock, M., Gounon, P., and Fouet, A. (1997). Molecular characterization of the *Bacillus anthracis* main S-layer component: evidence that it is the major cell-associated antigen. Mol Microbiol *23*, 1147-1155.

Mie, M., Naoki, T., and Kobatake, E. (2016). Development of a Split SNAP-CLIP Double Labeling System for Tracking Proteins Following Dissociation from Protein-Protein Complexes in Living Cells. Anal Chem *88*, 8166-8171.

Mullane, K.M., Miller, M.A., Weiss, K., Lentnek, A., Golan, Y., Sears, P.S., Shue, Y.K., Louie, T.J., and Gorbach, S.L. (2011). Efficacy of fidaxomicin versus vancomycin as therapy for *Clostridium difficile* infection in individuals taking concomitant antibiotics for other concurrent infections. Clin Infect Dis *53*, 440-447.

Nakayama, K., Takashima, K., Ishihara, H., Shinomiya, T., Kageyama, M., Kanaya, S., Ohnishi, M., Murata, T., Mori, H., and Hayashi, T. (2000). The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. Mol Microbiol *38*, 213-231.

Nale, J.Y., Spencer, J., Hargreaves, K.R., Buckley, A.M., Trzepinski, P., Douce, G.R., and Clokie, M.R. (2016). Bacteriophage Combinations Significantly Reduce *Clostridium difficile* Growth *In Vitro* and Proliferation *In Vivo*. Antimicrob Agents Chemother *60*, 968-981.

Nawrocki, K.L., Edwards, A.N., Daou, N., Bouillaut, L., and McBride, S.M. (2016). CodY-Dependent Regulation of Sporulation in *Clostridium difficile*. J Bacteriol *198*, 2113-2130.

Nguyen-Mau, S.M., Oh, S.Y., Kern, V.J., Missiakas, D.M., and Schneewind, O. (2012). Secretion genes as determinants of *Bacillus anthracis* chain length. J Bacteriol *194*, 3841-3850.

Noonan, B., and Trust, T.J. (1995). Molecular analysis of an A-protein secretion mutant of *Aeromonas salmonicida* reveals a surface layer-specific protein secretion pathway. J Mol Biol *248*, 316-327.

Oprita, R., Bratu, M., Oprita, B., and Diaconescu, B. (2016). Fecal transplantation - the new, inexpensive, safe, and rapidly effective approach in the treatment of gastrointestinal tract diseases. J Med Life *9*, 160-162.

Ortiz, C., Natale, P., Cueto, L., and Vicente, M. (2016). The keepers of the ring: regulators of FtsZ assembly. FEMS Microbiol Rev *40*, 57-67.

Papanikou, E., Karamanou, S., Baud, C., Frank, M., Sianidis, G., Keramisanou, D., Kalodimos, C.G., Kuhn, A., and Economou, A. (2005). Identification of the preprotein binding domain of SecA. J Biol Chem *280*, 43209-43217.

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

Park, E., and Rapoport, T.A. (2011). Preserving the membrane barrier for small molecules during bacterial protein translocation. Nature *473*, 239-242.

Parlitz, R., Eitan, A., Stjepanovic, G., Bahari, L., Bange, G., Bibi, E., and Sinning, I. (2007). *Escherichia coli* signal recognition particle receptor FtsY contains an essential and autonomous membrane-binding amphipathic helix. J Biol Chem *282*, 32176-32184.

Peltier, J., Courtin, P., El Meouche, I., Lemee, L., Chapot-Chartier, M.P., and Pons, J.L. (2011). *Clostridium difficile* has an original peptidoglycan structure with a high level of N-acetylglucosamine deacetylation and mainly 3-3 cross-links. J Biol Chem *286*, 29053-29062.

Pepin, J., Saheb, N., Coulombe, M.A., Alary, M.E., Corriveau, M.P., Authier, S., Leblanc, M., Rivard, G., Bettez, M., Primeau, V.*, et al.* (2005). Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*-associated diarrhea: a cohort study during an epidemic in Quebec. Clin Infect Dis *41*, 1254-1260.

Pereira, F.C., Saujet, L., Tome, A.R., Serrano, M., Monot, M., Couture-Tosi, E., Martin-Verstraete, I., Dupuy, B., and Henriques, A.O. (2013). The spore differentiation pathway in the enteric pathogen *Clostridium difficile*. PLoS Genet *9*, e1003782.

Petrof, E.O., Gloor, G.B., Vanner, S.J., Weese, S.J., Carter, D., Daigneault, M.C., Brown, E.M., Schroeter, K., and Allen-Vercoe, E. (2013). Stool substitute transplant therapy for the eradication of *Clostridium difficile* infection: 'RePOOPulating' the gut. Microbiome *1*, 3.

Pruitt, R.N., Chambers, M.G., Ng, K.K., Ohi, M.D., and Lacy, D.B. (2010). Structural organization of the functional domains of *Clostridium difficile* toxins A and B. Proc Natl Acad Sci U S A *107*, 13467-13472.

Purcell, E.B., McKee, R.W., McBride, S.M., Waters, C.M., and Tamayo, R. (2012). Cyclic diguanylate inversely regulates motility and aggregation in *Clostridium difficile*. J Bacteriol *194*, 3307-3316.

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

Qazi, O., Hitchen, P., Tissot, B., Panico, M., Morris, H.R., Dell, A., and Fairweather, N. (2009). Mass spectrometric analysis of the S-layer proteins from *Clostridium difficile* demonstrates the absence of glycosylation. J Mass Spectrom *44*, 368-374.

Ratner, M. (2016). Seres's pioneering microbiome drug fails mid-stage trial. Nat Biotech *34*, 1004-1005.

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

Rhoads, A., and Au, K.F. (2015). PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics *13*, 278-289.

Ridlon, J.M., Kang, D.J., and Hylemon, P.B. (2006). Bile salt biotransformations by human intestinal bacteria. J Lipid Res *47*, 241-259.

Rigel, N.W., and Braunstein, M. (2008). A new twist on an old pathway- accessory Secretion systems. Mol Microbiol *69*, 291-302.

Ristl, R., Steiner, K., Zarschler, K., Zayni, S., Messner, P., and Schaffer, C. (2011). The S-layer glycome-adding to the sugar coat of bacteria. Int J Microbiol *2011*.

Robson, A., Gold, V.A., Hodson, S., Clarke, A.R., and Collinson, I. (2009). Energy transduction in protein transport and the ATP hydrolytic cycle of SecA. Proc Natl Acad Sci U S A *106*, 5111-5116.

Ryan, A., Lynch, M., Smith, S.M., Amu, S., Nel, H.J., McCoy, C.E., Dowling, J.K., Draper, E., O'Reilly, V., McCarthy, C.*, et al.* (2011). A role for TLR4 in *Clostridium difficile* infection and the recognition of surface layer proteins. PLoS Pathog *7*, e1002076.

Sara, M., and Sleytr, U.B. (1987). Molecular sieving through S layers of *Bacillus stearothermophilus* strains. J Bacteriol *169*, 4092-4098.

Savidge, T.C., Pan, W.H., Newman, P., O'Brien, M., Anton, P.M., and Pothoulakis, C. (2003). *Clostridium difficile* toxin B is an inflammatory enterotoxin in human intestine. Gastroenterology *125*, 413-420.

Schaffer, C., and Messner, P. (2017). Emerging facets of prokaryotic glycosylation. FEMS Microbiol Rev *41*, 49-91.

Schaffer, C., Wugeditsch, T., Neuninger, C., and Messner, P. (1996). Are S-layer glycoproteins and lipopolysaccharides related? Microb Drug Resist *2*, 17-23.

Schirmer, J., and Aktories, K. (2004). Large clostridial cytotoxins: cellular biology of Rho/Ras-glucosylating toxins. Biochim Biophys Acta *1673*, 66-74.

Schmidt, T.G., and Skerra, A. (2007). The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. Nat Protoc *2*, 1528-1535.

Scholl, D., Cooley, M., Williams, S.R., Gebhart, D., Martin, D., Bates, A., and Mandrell, R. (2009). An engineered R-type pyocin is a highly specific and sensitive bactericidal agent for the food-borne pathogen *Escherichia coli* O157:H7. Antimicrob Agents Chemother *53*, 3074-3080.

Schwan, C., Kruppke, A.S., Nolke, T., Schumacher, L., Koch-Nolte, F., Kudryashev, M., Stahlberg, H., and Aktories, K. (2014). *Clostridium difficile* toxin CDT hijacks microtubule organization and reroutes vesicle traffic to increase pathogen adherence. Proc Natl Acad Sci U S A *111*, 2313-2318.

Sekulovic, O., Ospina Bedoya, M., Fivian-Hughes, A.S., Fairweather, N.F., and Fortier, L.C. (2015). The *Clostridium difficile* cell wall protein CwpV confers phase-variable phage resistance. Mol Microbiol *98*, 329-342.

Sharma, V., Arockiasamy, A., Ronning, D.R., Savva, C.G., Holzenburg, A., Braunstein, M., Jacobs, W.R., Jr., and Sacchettini, J.C. (2003). Crystal structure of *Mycobacterium tuberculosis* SecA, a preprotein translocating ATPase. Proc Natl Acad Sci U S A *100*, 2243-2248.

Shen, A. (2012). *Clostridium difficile* toxins: mediators of inflammation. J Innate Immun *4*, 149-158.

Siboo, I.R., Chambers, H.F., and Sullam, P.M. (2005). Role of SraP, a Serine-Rich Surface Protein of *Staphylococcus aureus*, in binding to human platelets. Infect Immun *73*, 2273-2280.

Sillanpaa, J., Martinez, B., Antikainen, J., Toba, T., Kalkkinen, N., Tankka, S., Lounatmaa, K., Keranen, J., Hook, M., Westerlund-Wikstrom, B.*, et al.* (2000). Characterization of the collagen-binding S-layer protein CbsA of *Lactobacillus crispatus*. J Bacteriol *182*, 6440-6450.

Smits, W.K., Lyras, D., Lacy, D.B., Wilcox, M.H., and Kuijper, E.J. (2016). *Clostridium difficile* infection. Nat Rev Dis Primers *2*, 16020.

Sockett, R.E. (2009). Predatory lifestyle of *Bdellovibrio bacteriovorus*. Annu Rev Microbiol *63*, 523-539.

Sorg, J.A., and Sonenshein, A.L. (2008). Bile salts and glycine as cogerminants for *Clostridium difficile* spores. J Bacteriol *190*, 2505-2512.

Stabler, R.A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T.D., Sebaihia, M., Quail, M.A., Rose, G.*, et al.* (2009). Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol *10*, R102.

Stabler, R.A., Valiente, E., Dawson, L.F., He, M., Parkhill, J., and Wren, B.W. (2010). In-depth genetic analysis of *Clostridium difficile* PCR-ribotype 027 strains reveals high genome fluidity including point mutations and inversions. Gut Microbes *1*, 269-276.

Swanson, S., Ioerger, T.R., Rigel, N.W., Miller, B.K., Braunstein, M., and Sacchettini, J.C. (2015). Structural Similarities and Differences between Two Functionally Distinct SecA Proteins, *Mycobacterium tuberculosis* SecA1 and SecA2. J Bacteriol *198*, 720-730.

Takumi, K., Endo, Y., Koga, T., Oka, T., and Natori, Y. (1992). In vitro self-assembly of the S layer subunits from *Clostridium difficile* GAI 0714 into tetragonal arrays. Tokushima J Exp Med *39*, 95-100.

Tan, K.S., Wee, B.Y., and Song, K.P. (2001). Evidence for holin function of *tcdE* gene in the pathogenicity of *Clostridium difficile*. J Med Microbiol *50*, 613-619.

Tanaka, Y., Sugano, Y., Takemoto, M., Mori, T., Furukawa, A., Kusakizako, T., Kumazaki, K., Kashima, A., Ishitani, R., Sugita, Y.*, et al.* (2015). Crystal Structures of SecYEG in Lipidic Cubic Phase Elucidate a Precise Resting and a Peptide-Bound State. Cell Rep *13*, 1561-1568.

Terveer, E.M., van Beurden, Y.H., Goorhuis, A., Seegers, J.F.M.L., Bauer, M.P., van Nood, E., Dijkgraaf, M.G.W., Mulder, C.J.J., Vandenbroucke-Grauls, C.M.J.E., Verspaget, H.W.*, et al.* (2017). How to: Establish and run a stool bank. Clinical Microbiology and Infection.

Thompson, S.A. (2002). *Campylobacter* surface-layers (S-layers) and immune evasion. Ann Periodontol *7*, 43-53.

Thompson, S.A., Shedd, O.L., Ray, K.C., Beins, M.H., Jorgensen, J.P., and Blaser, M.J. (1998). *Campylobacter fetus* surface layer proteins are transported by a type I secretion system. J Bacteriol *180*, 6450-6458.

Tsirigotaki, A., De Geyter, J., Sostaric, N., Economou, A., and Karamanou, S. (2017). Protein export through the bacterial Sec pathway. Nat Rev Micro *15*, 21-36.

Usenik, A., Renko, M., Mihelic, M., Lindic, N., Borisek, J., Perdih, A., Pretnar, G., Muller, U., and Turk, D. (2017). The CWB2 Cell Wall-Anchoring Module Is Revealed by the Crystal Structures of the *Clostridium difficile* Cell Wall Proteins Cwp8 and Cwp6. Structure *25*, 514-521.

Valent, Q.A., de Gier, J.W., von Heijne, G., Kendall, D.A., ten Hagen-Jongman, C.M., Oudega, B., and Luirink, J. (1997). Nascent membrane and presecretory proteins synthesized in *Escherichia coli* associate with signal recognition particle and trigger factor. Mol Microbiol *25*, 53-64.

Van den Berg, B., Clemons, W.M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. Nature *427*, 36-44.

van der Sluis, E.O., Nouwen, N., Koch, J., de Keyzer, J., van der Does, C., Tampe, R., and Driessen, A.J. (2006). Identification of two interaction sites in SecY that are important for the functional interaction with SecA. J Mol Biol *361*, 839-849.

van der Woude, M.W., and Baumler, A.J. (2004). Phase and antigenic variation in bacteria. Clin Microbiol Rev *17*, 581-611, table of contents.

van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E.G., de Vos, W.M., Visser, C.E., Kuijper, E.J., Bartelsman, J.F., Tijssen, J.G.*, et al.* (2013). Duodenal infusion of donor feces for recurrent *Clostridium difficile*. N Engl J Med *368*, 407-415.

Vandamme, D., Landuyt, B., Luyten, W., and Schoofs, L. (2012). A comprehensive summary of LL-37, the factotum human cathelicidin peptide. Cell Immunol *280*, 22-35.

Venugopal, A.A., and Johnson, S. (2012). Fidaxomicin: a novel macrocyclic antibiotic approved for treatment of *Clostridium difficile* infection. Clin Infect Dis *54*, 568-574.

Viswanathan, V.K., Mallozzi, M.J., and Vedantam, G. (2010). *Clostridium difficile* infection: An overview of the disease and its pathogenesis, epidemiology and interventions. Gut Microbes *1*, 234-242.

Vollmer, W., and Bertsche, U. (2008). Murein (peptidoglycan) structure, architecture and biosynthesis in *Escherichia coli*. Biochim Biophys Acta *1778*, 1714-1734.

von Heijne, G. (1990). The signal peptide. J Membr Biol *115*, 195-201.

Waligora, A.J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A., and Karjalainen, T. (2001). Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. Infect Immun *69*, 2144-2153.

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.*, et al.* (2012). Changing epidemiology of *Clostridium difficile* infection following the introduction of a national ribotyping-based surveillance scheme in England. Clin Infect Dis *55*, 1056-1063.

Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., and Hochstrasser, D.F. (1999). Protein identification and analysis tools in the ExPASy server. Methods Mol Biol *112*, 531-552.

Willing, S.E., Candela, T., Shaw, H.A., Seager, Z., Mesnage, S., Fagan, R.P., and Fairweather, N.F. (2015). *Clostridium difficile* surface proteins are anchored to the cell wall using CWB2 motifs that recognise the anionic polymer PSII. Mol Microbiol *96*, 596-608.

Wren, B.W., and Tabaqchali, S. (1987). Restriction endonuclease DNA analysis of *Clostridium difficile*. J Clin Microbiol *25*, 2402-2404.

Wüst, J., Sullivan, N.M., Hardegger, U., and Wilkins, T.D. (1982). Investigation of an outbreak of antibiotic-associated colitis by various typing methods. Journal of Clinical Microbiology *16*, 1096-1101.

Xie, K., and Dalbey, R.E. (2008). Inserting proteins into the bacterial cytoplasmic membrane using the Sec and YidC translocases. Nat Rev Microbiol *6*, 234-244.

Zhang, F., Luo, W., Shi, Y., Fan, Z., and Ji, G. (2012). Should We Standardize the 1,700-Year-Old Fecal Microbiota Transplantation. Am J Gastroenterol *107*, 1755-1755.

Zimmer, J., Nam, Y., and Rapoport, T.A. (2008). Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature *455*, 936-943.

Appendix I

Table A1: Bacterial strains used in this study

C. difficile		
Strain	**Description**	**Source**
R20291	Ribotype 027 strain responsible for the 2004 Stoke-Mandeville outbreak.	(Stabler et al., 2009)
R20291 *recV*::*erm cwpV* ON	R20291 containing a ClosTroin inactivated *recV*, with the *cwpV* switch locked in the ON orientation.	(Sekulovic et al., 2015)
R20291 *recV*::*erm cwpV* OFF	R20291 containing a ClosTroin inactivated *recV*, with the *cwpV* switch locked in the OFF orientation.	(Sekulovic et al., 2015)
FM2.5	Spontaneous mutant resistant to killing by Avidocin-CDs. This strain contains a single base insertion early in the *slpA* gene resulting in a frame shift and loss of *slpA* function.	AvidBiotics
FM2.5RW	FM2.5 in which a functional copy of *slpA*, containing a silent mutation, has been returned to the chromosome.	This study
FM2.5 *recV*::*erm cwpV* ON	FM2.5 containing a ClosTroin inactivated *recV*, with the *cwpV* switch locked in the ON orientation.	This study
FM2.5 *recV*::*erm cwpV* OFF	FM2.5 containing a ClosTroin inactivated *recV*, with the *cwpV* switch locked in the OFF orientation.	This study
FM2.5RW *recV*::*erm cwpV* ON	FM2.5RW containing a ClosTroin inactivated *recV*, with the *cwpV* switch locked in the ON orientation.	This study
FM2.5RW *recV*::*erm cwpV* OFF	FM2.5RW containing a ClosTroin inactivated *recV*, with the *cwpV* switch locked in the OFF orientation.	This study
FM2.5RΔD2	FM2.5 in which a functional copy of *slpA*, in which the coding region for the domain 2 of the LMW subunit has been deleted, has been returned to the chromosome.	(Oishik Banerji, 2017)
FM2.6	Spontaneous mutant resistant to killing by Avidocin-CDs. This strain contains a single base substitution early in the *slpA* gene resulting in a stop codon and loss of *slpA* function.	AvidBiotics
FM2.6RW	FM2.6 in which a functional copy of *slpA*, containing a silent mutation, has been returned to the chromosome.	This study
630	Isolated from the stool of a healthy infant during an outbreak in Zurich.	(Wüst et al., 1982)

630 *pyrE:: Ptet-mreB2*	630 in which *mreB2* under control of a tetracycline inducible promotor has been inserted into the *pyrE* locus on the chromosome.	This study
***E. coli* strains**		
NEB-5α	Chemically competent cells used for routine cloning and plasmid propagation.	New England Biolabs
CA434	Strain HB101 carrying the IncPβ conjugative plasmid, R702. Used as a conjugation donor.	(Purdy et al., 2002)
Top10	Chemically competent cells used for routine cloning and plasmid propagation.	Lab stock

Appendix II

Table A2: Primers used in this study

Name	Sequence 5'-3'	Restriction site
NF823	GCATTTTTCCATCCATCTTG	
NF825	TTTAAGGTAAGTTTGATTTTTATGTTAATGAAT	
NF826	ATTCATTAACATAAAAATCAAACTTACCTTAAA	
NF863	GTAAAAAGAGTAGCACTTTTAATTCTAAAGG	
NF1215	CTTAATCACATATATCACACTTGTATAAGTTTACTGG	
NF1957	GAGTCAGTTATAGATTCGATACTTGAC	
NF1958	GAGTTTTTCAAATTGTGGATGACTCCAC	
RF21	GGATTTCACATTTGCCGTTTTGTAAAC	
RF22	GATCTTTTCTACGGGGTCTGAC	
RF33	GGCTTCTATAGCTTGGTGAA	
RF110	GACATAACTGCAGCACTACTTG	
RF111	CAGGATTAACAGTATTAGCTTCTGC	
RF192	GATCTCTAGAGATGTCAGTTATAGATTCGATACTTGAC	*Xba*I
RF193	GATCGGTACCGCGTTAAATTTATATAAGTATTGCACTGTTGC	*Kpn*I
RF194	GATCTCTAGAGATGAGCTTTATGGATAATTTATTCAACATG	*Xba*I
RF195	GATC GGTACCGCTCTTCCACAACAGTTTTTATATTTTTTACC	*Kpn*I
RF196	GATCTCTAGAGATGGCTAAGGAAAAAAGAAAGAAAAAAAG	*Xba*I
RF197	GATCGGTACCGCATTCTTAGTATTCATCATTAATACTTTTTC	*Kpn*I
RF200	GATCTCTAGAGATGGCTGGAGCTGATATAGG	*Xba*I
RF201	GATCGGTACCGCTTTTCTTTTGAAAGACCCACCAG	*Kpn*I
RF208	GATCGAGCTCGGAGGAACTACTATGG	*Sac*I
RF209	GATC CTCGAGAGCAGCTGCCCCAAGTCCTGGTTTCCCCAAAC	*Xho*I
RF226	GATCGAGCTCGGAGGAACTACTATGGATAAAGATTGTGAAATGAA AAG	*Sac*I
RF227	GATCCTCGAGAGCAGCTGC TCCTAATCCTGGTTTTCCTAATC	*Xho*I
RF234	GATCCTCGAGAGCTTTATGGATAATTTATTCAACATGG	*Xho*I

RF235	GATCGGATCCGCCATCTGCTCCATACTATGTTTATAC	*Bam*HI
RF311	TAGGGTAACAAAAAACACCG	
RF312	CCTTTTTGATAATCTCATGACC	
RF430	CGTAGAAATACGGTGTTTTTTGTTACCCTAGTAGTAAATATTATAATTCTGATGATAAAAATG	
RF431	ACAATCTTTATCCATATAAGTACCTCCTTTTCTATTGTC	
RF432	AAAGGAGGTACTTATATGGATAAAGATTGTGAAATG	
RF433	GGGATTTTGGTCATGAGATTATCAAAAAGGAATCAGCTCTAATAACTGGTCTATTAG	
RF434	CGTAGAAATACGGTGTTTTTTGTTACCCTAAGTAAGTTAATTATAATTTGGGTATATG	
RF435	ACAATCTTTATCCATATTTTACATCTCTCCTCGTTC	
RF436	GGAGAGATGTAAAATATGGATAAAGATTGTGAAATG	
RF437	GGGATTTTGGTCATGAGATTATCAAAAAGGTAAGTCTTCCCTTATCATAAGC	
RF438	CATTAATAGCTTGCTGAGTTACTCTTATATATCGC	
RF464	TTGTTGAAATACAGGATGGTGAAG	
RF466	GCAGCTGCTCTCGAGTCAG	
RF655	CGTAGAAATACGGTGTTTTTTGTTACCCTATAGTGATAAGTCTGTACTTAGAAGTGC	
RF656	ATTTTTCTTATTCATTTCTTAAATTCCTCCCAAC	
RF658	ATTATATTTATATTACATATCTAATAAATCTTTCATTTTGTTTATAAC	
RF659	GATTTATTAGATATGTAATATAAATATAATAATATAAAAAGGCTTC	
RF660	GGGATTTTGGTCATGAGATTATCAAAAAGGTTTCTATCCATCTCAAAGTCTTC	
RF750	CGTAGAAATACGGTGTTTTTTGTTACCCTAATGAATAAGAAAAATATAGCAATAGC	
RF779	CGTAGAAATACGGTGTTTTTTGTTACCCTATTATAGAGATGTGAGAAATATTAGGAATATATG	
RF780	AACTACAGTATAACCTTGTGTTCCAGTAGTTGCTGCAAACACAGGTGC	
RF798	GATCGTTTAAACCAAACAGAGTTATAAAACCTATGGAAAGG	*Pme*I
RF799	GATCGTTTAAACCTTCTTTAACCTCACTAGTTGGCTTAAAG	*Pme*I
RF826	GGTTATCAATTAGTAGATGCAGTTGC	
RF835	GATCGGTACC TAGTGATAAGTCTGTACTTAGAAGTG	*Kpn*I
RF836	GATCGGATCC TTTCTATCCATCTCAAAGTCTTCTTC	*Bam*HI
RF837	CGTAGAAATACGGTGTTTTTTGTTACCCTAAATATATCTTGCCCAAATGTC	
RF838	TTGATGCAGAATTCGTTATTACATCCCTAATTCCTTG	
RF839	TTAGGGATGTAATAACGAATTCTGCATCAAGCTAG	
RF840	TAAATAATTAAGTTTAAATAAAAAAGACTTCTCATGAGAG	
RF841	AAGTCTTTTTTATTTAAACTTAATTATTTATAGTGTTACTTAAAAAATG	
RF842	GGGATTTTGGTCATGAGATTATCAAAAAGGATAGTATATAACATTAATAAAATTTAAAATCAATAATTATC	
RF863	CAGCTAACCTTCATTTATAAGGCTAAC	
RF874	GTACAGTTGTTGAACAAATAAAAGACTTAG	
RF901	GGTACCATAAAAATAAGAAGCCTGC	
RF902	GAATAAAAATGGATTATTATAGAGATGTG	
RF963	TGGCATAGCATCATTTAATGTTTCTTC	

RF964	AGGCAACTTTATAAAGAAATATTTAAATTTATATTAAAATATTTTTAT ATTTTTATTAGG	
RF966	GTATACTTAAGTTAAAACTAAATAGGCAAATC	
RF990	CAACATCATCATCTACAACCAG	
RF991	GATAGGTGAAATTTTGGCTTTTAAAGTAG	
RF992	CATTGAGATGCAAAATATAAACCTTTTG	
RF994	GTTTCATTTATGTAATGAAAATGTTATATAAATG	
RF995	GTAGCATATTCATTGTCTTAACTATTG	
RF997	GCAAATAAGTACATCTCAAAAGTTTCC	
RF1051	GATC GGTACC CTCATATGCTTGGGAAAAATAATAATTTC	*Kpn*I
RF1052	GATCGGTACCGACTTGCAAAAAGAATAAAAATGGATTATTATAG	*Kpn*I
RF1053	GATCGAGCTCCATTATATTAAAATTACATCTCTTCATTATGTAC	*Sac*I

Appendix III

Table A3: Plasmids used in this study

Plasmid Name	Description	Source
pASF75	ClosTron vector carrying a group II intron targeted to *recV.*	(Emerson et al., 2009)
pJAK012	pMTL960 based vector containing strep-tagged *secA2* under control of *Pcwp2*. There is a *Xho*I restriction site between the strep-tag and *secA2.*	This study
pJAK013	pJAK012 in which the strep tag has been replaced with a SNAP tag.	This study
pJAK015	pJAK012 in which *secA2* has been replaced with a *secA1.*	This study
pJAK034	pJAK015 in which the SNAP tag has been replaced with a CLIP tag	This study
pJAK047	pMTLSC7315 based vector to insert a SNAP tag to the 5' of chromosomal *secA2*	This study
pJAK049	pMTLSC7315 based vector to insert a CLIP tag to the 5' end of chromosomal *secA1*	This study
pJAK052	pMTLSC7315 based vector to insert nSNAP tag to the 5' of chromosomal *secA2*	This study
pJAK072	pMTLSC7215 based vector to replace the native R20291 *slpA* with a SLCT-7 *slpA*	This study
pJAK073	pMTLSC7215 based vector to lock the *cwpV* switch in the OFF orientation	This study
pJAK074	pMTLSC7215 based vector to lock the *cwpV* switch in the ON orientation	This study
pJAK076	pMTL960 based vector to replace the native R20291 *slpA* with a SLCT-7 *slpA*	This study
pJAK080	pMTLSC7315 based vector to insert *mreB2* under control of a tetracycline inducible promoter into the *pyrE* locus on the chromosome	This study
pJAK084	pMTLSc7315 based vector to knock out the native *mreB2*	This study
pJAK086	pJAK076 with a truncated left homology arm	This study

pJAK106	pMTL960 based vector containing *gusA* under control of the *slpA* promoter (to the -35 region)	This study
pJAK107	pMTL960 based vector containing *gusA* under control of the *slpA* promoter (to include the potential UPE)	This study
pMTLSC7215	*codA* allele exchange vector for use in strain R20291	(Cartman et al., 2012)
pMTLSC7315	*codA* allele exchange vector for use in strain 630	(Cartman et al., 2012)
pRPF144	pMTL960 based vector containing strep-tagged *secA2* under control of *Pcwp2*	(Fagan and Fairweather, 2011)
pRPF185	pMTL960 based vector containing *gusA* under control of the tetracycline inducible promoter.	(Fagan and Fairweather, 2011)
pRPF186	pMTL960 based vector containing strep-tagged *secA2* under control of the tetracycline inducible promoter.	(Fagan and Fairweather, 2011)
pRPF187	pMTL960 based vector containing strep-tagged *secA2K106R* under control of the tetracycline inducible promoter.	(Fagan and Fairweather, 2011)
pRPF193	pMTL960 based vector containing strep-tagged *secA1* under control of the tetracycline inducible promoter.	(Fagan and Fairweather, 2011)

Appendix IV

As part of my PhD thesis I am including my published paper "Heat shock increases conjugation efficiency in *Clostridium difficile*", which was originally published in Anaerobe, https://doi.org/10.1016/j.anaerobe.2016.06.009

All work presented in this paper was performed by myself. I also wrote all sections of the paper, which was then edited by my supervisor and the corresponding author, Dr Robert Fagan.

This work is under a creative commons licence (https://creativecommons.org/licenses/by/4.0/) and I therefore have permission to use this article in my thesis.

Yours sincerely,

Joseph Kirk (PhD candidate)

Dr Robert Fagan

Contents lists available at ScienceDirect

Anaerobe
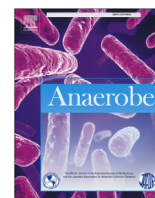
Clostridium difficile (including epidemiology)

Heat shock increases conjugation efficiency in *Clostridium difficile*☆

Joseph A. Kirk, Robert P. Fagan*

Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, S10 2TN, UK

ABSTRACT

Clostridium difficile infection has increased in incidence and severity over the past decade, and poses a unique threat to human health. However, genetic manipulation of *C. difficile* remains in its infancy and the bacterium remains relatively poorly characterised. Low-efficiency conjugation is currently the only available method for transfer of plasmid DNA into *C. difficile*. This is practically limiting and has slowed progress in understanding this important pathogen. Conjugation efficiency varies widely between strains, with important clinically relevant strains such as R20291 being particularly refractory to plasmid transfer. Here we present an optimised conjugation method in which the recipient *C. difficile* is heat treated prior to conjugation. This significantly improves conjugation efficiency in all *C. difficile* strains tested including R20291. Conjugation efficiency was also affected by the choice of media on which conjugations were performed, with standard BHI media giving most transconjugant recovery. Using our optimised method greatly increased the ease with which the chromosome of R20291 could be precisely manipulated by homologous recombination. Our method improves on current conjugation protocols and will help speed genetic manipulation of strains otherwise difficult to work with.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Clostridium difficile poses a significant threat to human health and is the most common cause of antibiotic associated diarrhoea [1]. Both incidence and mortality has increased over the last decade. This is due, at least in part, to the emergence of epidemic lineages such as ribotype 027 and 078 [1]. Despite the increase incidence and severity of *C. difficile* infections, relatively little is known about the molecular basis of disease. However, an increased recognition of the importance of *C. difficile* in human health has led to an increased interest in this bacterium. Despite this, genetic manipulation of *C. difficile* remains in its infancy. One limitation in *C. difficile* research is the reliance on low-efficiency conjugation, as this is the only useable method for DNA transfer into *C. difficile* [2]. This is less of a problem when using strains with relatively high conjugation efficiencies such as the laboratory strain CD630. However, clinically relevant epidemic strains, such as the ribotype 027 strain R20291, display very low conjugation efficiencies. This makes genetic manipulation of these strains particularly difficult.

Conjugation efficiencies of other bacteria, such as *Campylobacter*

jejuni and *Corynebacterium glutamicum*, have been shown to increase in response to heat treatment [3,4]. Here we demonstrate that conjugation efficiency in a panel of *C. difficile* strains is increased in response to heat treatment and we provide an optimised conjugation protocol that will increase the ease with which the species may be genetically manipulated.

2. Materials and methods

2.1. Growth and handling of organisms

C. difficile strains, outlined in Table 1, were propagated either in TY broth without thioglycolate [5] or on BHI agar. Cultures were incubated at 37 °C in an anaerobic environment composed of 80% nitrogen, 10% hydrogen and 10% carbon dioxide. *Escherichia coli* was routinely grown in LB broth or on LB agar. *E. coli* strain CA434 (HB101 carrying R702) was used as the conjugation donor throughout. NEB5α (New England Biolabs) was used for cloning and plasmid propagation. Growth media was supplemented with chloramphenicol (15 μg/ml), thiamphenicol (15 μg/ml) or cycloserine (250 μg/ml) as appropriate.

2.2. Viability of C. difficile after heat treatment

C. difficile strain R20291 was grown overnight in TY broth. 200 μl

---

Table 1
C. difficile strains used in this study including their respective clade, ribotype, and S-layer type.

C. difficile strain	Clade	Ribotype	S-layer type
Ox858	1	029	2
Ox1437a	1	026	5
Ox1896	1	018	6
630	1	012	7
Ox1396	1	012	8
Ox1533	1	129	10
R20291	2	027	4
R7404	4	017	7b

samples were incubated at 44, 46, 48, 50, and 52 °C for 5, 15, 30, 45, or 60 min. An unheated control was included. Samples were then serially diluted and 10 μl of each dilution was spotted onto well-dried BHI plates and incubated overnight. Enumerations were performed in triplicate on biological duplicates. Colonies of *C. difficile* were counted and viability calculated as CFU/ml.

2.3. Plasmid conjugations

A previously described and widely used conjugation protocol was used as the starting point for development of our improved method [2]. 200 μl samples of *C. difficile* overnight cultures were heated, as above, and incubated at 37 °C for 2 min. 1 ml of overnight *E. coli* conjugant donor (CA434) culture was harvested by centrifugation at 4000g for 2 min and transferred into the anaerobic workstation. *E. coli* cell pellets were then gently resuspended in 200 μl of heat treated or untreated *C. difficile* culture. This mixed cell suspension was then pipetted onto well-dried, non-selective agar plates (10 × 10 μl spots) and allowed to dry. BHI agar was used routinely but BHIS (BHI agar supplemented with 0.1% (w/v) cysteine and 0.5% (w/v) yeast extract), TY [5] and Brazier's (Brazier's CCEY media, 1% (v/v) defibrinated horse blood, 4% (v/v) egg yolk emulsion) agar were also tested. All solid media contained 1.5% agar. Conjugations were then incubated for 8—24 h following which growth was harvested using 900 μl of TY broth, serially diluted and spread on plates containing either cycloserine (for total *C. difficile* CFU counts), or cycloserine and thiamphenicol (to select for transconjugants). Approximate conjugation efficiency was then calculated as transconjugant CFU/total *C. difficile* CFU. These experiments were performed using biological duplicates with technical triplicates. Statistical significance of these results was determined using either individual student t-tests or in combination with one way analysis of variance (ANOVA), performed using GraphPad Prism 6. *P* values of <0.05 were considered statistically significant.

2.4. Creation of a chromosomal divIVA-SNAP fusion mutant and fluorescence microscopy

1.2 kb upstream and downstream of the 3′ region of *divIVA* were amplified using RF439 (CGTAGAAA-TACGGTGTTTTTTGTTACCCTACTGTAGCAATATTAAATTCTACAAATG) with RF440 (CAGCAGCTGCCTCGAGTTCTAAAGTTGTAGCAGCTTC) and RF443 (CCAGGACTTGGGTAAGCTTGTAATTTGTTTATTTTTTATG) with RF444 (GGGATTTTGGTCATGAGATTATCAAAAAGGA-CATTTGATGGTAAAGTCCATG) respectively. A *C. difficile* codon-optimised SNAP-tag coding sequence was cloned from pFT46 [6] using RF441 (GCTACAACTTTAGAACTCGAGGCAGCTGCTGAT) and RF442 (AAACAAATTACAAGCTTACCCAAGTCCTGGTTTCC). pMTL-SC7215 [7] was linearized using RF311 (TAGGGTAACAAAAAA-CACCG) and RF312 (CCTTTTTGATAATCTCATGACC). These fragments

were then combined using a combination of SOEing PCR [8] and Gibson assembly [9] using the HiFi DNA assembly kit (New England Biolabs). The resulting plasmid (pJAK050) was then conjugated into *C. difficile* strain R20291 and mutants selected using the method previously described [7]. Correct insertion of the SNAP-tag coding sequence into the chromosome was confirmed by PCR using RF445 (GGTTTAAGAGGGTATAGAGATG) and RF545 (CGAGTTA-TAAATCGCGTTACCACC). For fluorescent labelling of DivIVA-SNAP, 1 ml of exponentially growing R20291 and R20291 *divIVA-SNAP* was incubated anaerobically with 250 nM SNAP TMR-star (New England Biolabs) for 30 min before slides were prepared for standard fluorescence microscopy, or lysates prepared. Proteins in the lysates were resolved on 12% SDS-PAGE gels and visualised using a BioRad ChemiDoc MP imaging system. Slides were visualised using a Nikon Ti Eclipse inverted epifluorescence microscope.

3. Results and discussion

3.1. Heat treatment of R20291 increases conjugation efficiency

C. difficile strain R20291 was chosen for these experiments due to its clinical relevance and relatively low conjugation efficiency. The effect of heat treatment on conjugation efficiency was tested. R20291 cultures were heated to 44—52 °C for 15 min prior to conjugation with the *E. coli* donor CA434 bearing the plasmid pRPF185 [10]. pRPF185 was chosen as it is a commonly used plasmid in *C. difficile* research and contains the native *C. difficile* replicon pCD6. A clear increase in conjugation efficiency was observed with increasing temperatures, from 46 °C to 50 °C (Fig. 1A). However, no transconjugants were observed when treated at 52 °C for 15 min. Additionally, although conjugation efficiency was increased at 50 °C relative to 48 °C, lower total CFUs and transconjugant CFUs were recorded. To test the effect of the heat treatment on *C. difficile* viability, cultures were heated to 44—52 °C for 5—60 min and viable cells enumerated (Fig. 1B). Viability was unaffected by incubation at 44 °C but there was a significant reduction in viability at temperatures above this and viability further decreased with increasing incubation time. Based on these data it is clear that the total number of transconjugants recovered is a compromise between increasing conjugation efficiency and decreasing viability. Although heat treatment at 52 °C for 15 min resulted in recovery of no transconjugants, this treatment did reduce viability by more than four orders of magnitude. As shorter heat treatments minimised the impact on viability, conjugation was tested following a 52 °C heat treatment for 5 min. This treatment almost doubled the recovery of transconjugants compared to the next best condition (approx. 2800 CFU/ml following 52 °C, 5 min compared with 1500 CFU/ml following 50 °C, 15min). These conditions were used as standard for the remainder of the experiments described below unless otherwise stated.

3.2. Conjugation efficiency is affected by media choice

There appears to be no standard media choice for *C. difficile* conjugations in the current literature. For this reason, the effect of media choice on conjugation efficiency, after heat treatment, was investigated. Conjugations were performed on BHI, BHIS, TY, and Braziers agar (Fig. 2). The highest conjugation efficiency was recorded when conjugations were performed using BHI agar. Although conjugation efficiency was lower when using BHIS relative to TY, higher numbers of transconjugants were recovered when using TY due to a higher total CFU count recorded when using this growth medium. Total *C. difficile* CFUs recovered from Brazier's was the lowest of all tested media, and resulted in the recovery of no transconjugants. While this observation is unlikely to be due to the
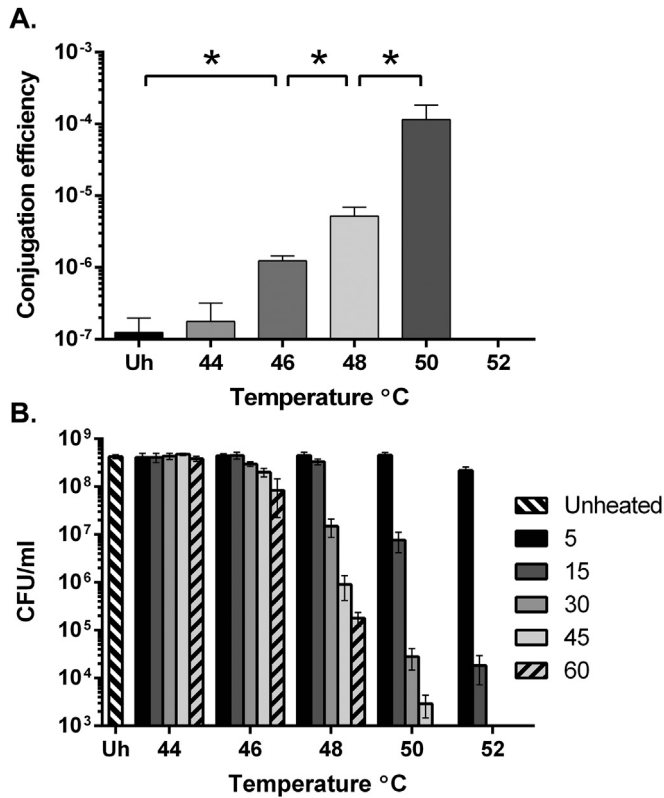
Fig. 1. Effect of heat treatment on conjugation efficiency and viability of *C. difficile* strain R20291. **(A)** Conjugation efficiency of unheated and heat treated (44—52 °C for 15 min) *C. difficile* strain R20291. Asterisks (*) denote statistical significance between heat treatment conditions ($P < 0.05$). Each bar represents the mean and standard deviation of data collected from experiments performed in triplicate using biological duplicates. **(B)** Viability of R20291 after heat treatment at varying temperatures (44—52 °C) and incubation times (5—60 min). Each bar represents the mean and standard deviation of data collected from experiments performed in triplicate using biological duplicates.
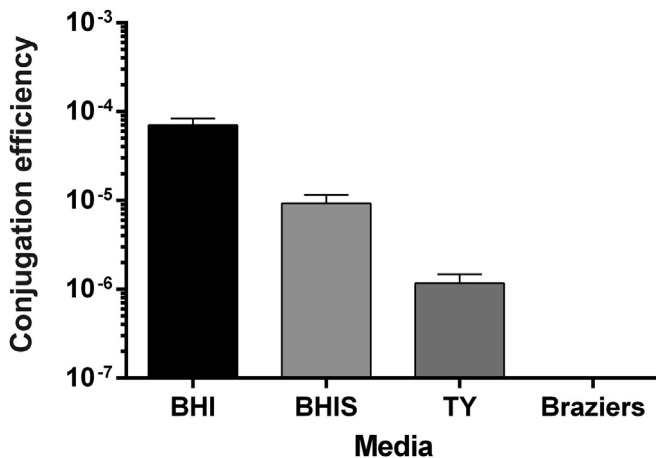

Fig. 2. Effect of media choice on conjugation efficiency of *C. difficile* strain R20291 after heat treatment. BHI, BHIS, TY and Brazier's agar were tested in 8 h conjugations following a 52 °C, 5 min heat treatment. Use of Brazier's agar resulted in the recovery of no transconjugants. Each bar represents the mean and standard deviation of data collected from experiments performed in triplicate using biological duplicates.

media having a direct effect on the actually conjugation process, it is likely that population equilibrium between *C. difficile* and *E. coli* is affected, which may in turn impact overall conjugation efficiency.

3.3. Heat treatment improves conjugation efficiency of a plasmid containing a non-native replicon

The precise manipulation of the genome of R20291 can be achieved through the use of the allele-exchange vector pMTL-SC7215. This pseudosuicide vector contains the pBP1 replicon, which replicates at a rate lower than that of the host cell chromosome. The result is non-exponential growth, limited by the rate of plasmid replication, when under antibiotic selection [7]. However, plasmids based upon this vector are difficult to conjugate into strain R20291 and the number of publications describing the use of this plasmid remains low. For this reason a pMTL-SC7215 based vector was created to add the SNAP-tag coding sequence to the chromosomal *divIVA* gene. This plasmid, pJAK050, was then used in heat treated and unheated conjugations. Due to the extremely low conjugation efficiency when using this plasmid, 24 h conjugations were performed following a 50 °C for 15 min heat treatment. No transconjugants were recovered using the standard unheated conjugation protocol. However, using our optimised heat treatment protocol large numbers of transconjugants were recovered, allowing a *divIVA-SNAP* mutant to be isolated in little over 2 weeks after the plasmid was constructed. Correct insertion of the SNAP-tag coding region into the *C. difficile* genome was confirmed by PCR (Fig. 3A). Fluorescence imaging of lysates of SNAP TMR-star treated R20291 *divIVA-SNAP* resolved by SDS-PAGE shows a band at approximately 40 kDa, the estimated size of DivIVA-SNAP (Fig. 3B). SNAP TMR-star treated R20291 *divIVA-SNAP* displayed mostly septal and some polar localisation of fluorescence (Fig. 3C and D), similar to DivIVA localisation observed in other bacterial species [11]. Taken together, these results suggest correct insertion of the SNAP-tag coding sequence into the *C. difficile* chromosome at the appropriate genetic locus.

3.4. Heat treatment increases conjugation efficiency in a panel of C. difficile strains

In order to determine whether the improved conjugation protocol was more widely applicable, conjugations were performed with pRPF185 on a panel of *C. difficile* strains. The strains tested represented a further 6 ribotypes and the majority of *C. difficile* clades and S-layer types [12]. All strains tested displayed an increase in conjugation efficiency after heat treatment, although the observed improvement was greatest for strain R20291 (Fig. 4). Although efficiency was improved, heat treatment did not result in a significant increase in the number of transconjugants recovered for strains Ox1396 and Ox1437a. This is likely due to a decrease in viability following heat treatment and may be improved by extending the conjugation time. Using the standard 52 °C for 5 min heat treatment, conjugation efficiency into strain R7404 was too low to accurately measure after an 8 h conjugation and no significant increase in conjugation efficiency was observed when conjugations were extended to 24 h. R7404 conjugation was then tested after incubation at 50 °C for 15 min followed by a 24 h incubation. This resulted in a significant increase in both conjugation efficiency (Fig. 4) and in the number of transconjugants recovered. These results indicate that our heat treatment conjugation protocol is broadly applicable to the majority of *C. difficile* strains and may be further optimised for specific applications.

4. Conclusions

C. difficile shows considerable strain-to-strain variation in conjugation efficiency. Conjugation efficiency is improved in a panel of strains, representing the majority of clades and S-layer types, through heat treatment of the recipient *C. difficile* prior to
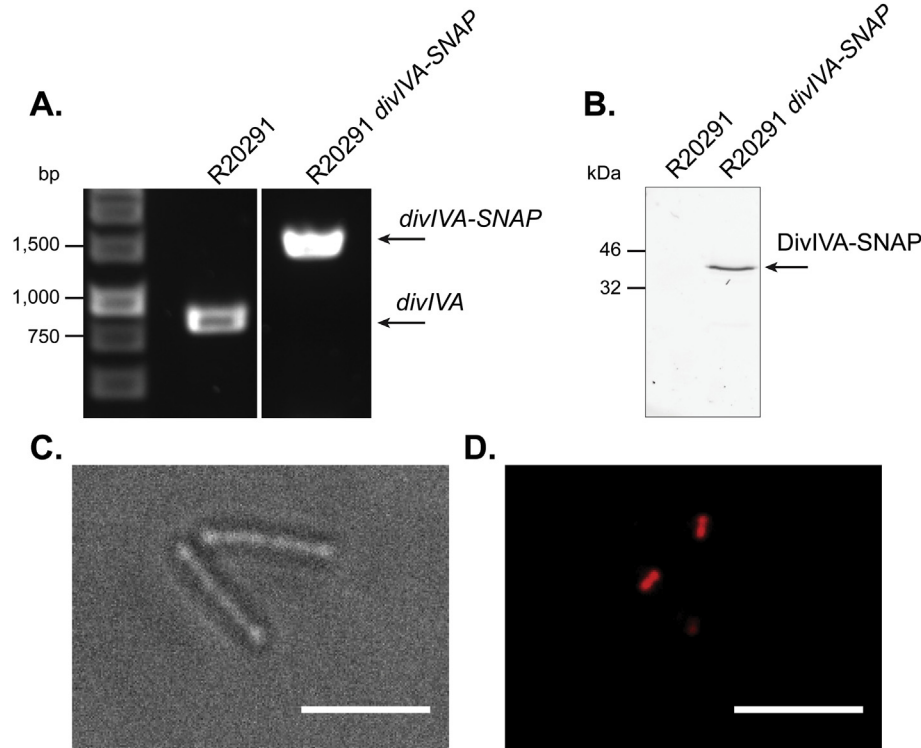
Fig. 3. Precise manipulation of the R20291 genome accelerated by the use of optimised conjugation. **(A)** 0.8% Agarose gel showing PCR fragments amplified using primers flanking the SNAP-tag coding sequence insertion site in the chromosomal *divIVA* gene. The increase of approximately 550bp from 885 bp (R20291) to 1443 bp (R20291 *divIVA-SNAP)* suggests the correct insertion of the SNAP-tag coding sequence. **(B)** 12% SDS-PAGE gel imaged using a fluorescence imager, showing resolved lysates from SNAP TMR-star treated R20291 and R20291 *divIVA-SNAP*. A band at approximately 40 kDa in the mutant corresponds to the addition of a SNAP-tag on DivIVA. **(C)** Brightfield and **(D)** fluorescence microscopy of exponentially growing R20291 *divIVA-SNAP* stained with 250 nM SNAP TMR-star for 30 min, showing mostly septal and some polar localisation of fluorescence. This demonstrates successful modification of the R20291 genome. Scale bar represents 5 μm.
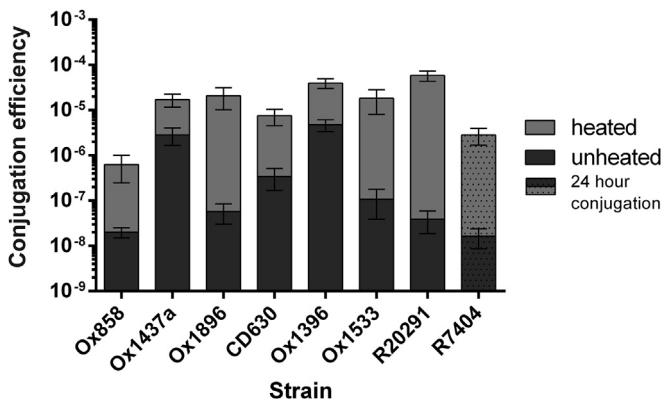

Fig. 4. Effect of heat treatment on the conjugation efficiency of a panel of *C. difficile* strains. A panel of *C. difficile* strains was used in unheated (dark grey) and heat treated (light grey, 52 °C for 5 min) 8 h conjugations. Strain R7404 showed no significant increase in conjugation efficiency with this treatment and was therefore used in a 24 h conjugation after a more stressful heat treatment (50 °C for 15 min). Each bar represents the mean and standard deviation of data collected from experiments performed in triplicate using biological duplicates. All strains displayed a statistically significant (P < 0.05) increase in conjugation efficiency after heat treatment.

conjugation. Conjugation efficiency is dependent on the media choice on which the conjugation is performed, with the highest number of transconjugants being recovered from BHI agar in this study. Using an optimised heat treatment conjugation protocol significantly improved the ease in which a pMTL-SC7215 based plasmid was introduced into R20291, facilitating chromosomal manipulation of the R20291 genome by homologous

recombination. Heating the recipient *C. difficile* to 52 °C for 5 min prior to conjugation proved to be optimum for an 8 h conjugation, although conjugation efficiency can be increased by coupling more stressful heat treatments with longer conjugation times.

Acknowledgements

References

[1] W.K. Smits, D. Lyras, D.B. Lacy, M.H. Wilcox, E.J. Kuijper, *Clostridium difficile* infection, Nat. Rev. Dis. Prim. 2 (2016) 16020.

[2] D. Purdy, T.A. O'Keeffe, M. Elmore, M. Herbert, A. McLeod, M. Bokori-Brown, et al., Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier, Mol. Microbiol. 46 (2002) 439–452.

[3] X. Zeng, D. Ardeshna, J. Lin, Heat shock-enhanced conjugation efficiency in standard *Campylobacter jejuni* strains, Appl. Environ. Microbiol. 81 (2015) 4546–4552.

[4] A. Schafer, J. Kalinowski, A. Puhler, Increased fertility of *Corynebacterium glutamicum* recipients in intergeneric matings with *Escherichia coli* after stress exposure, Appl. Environ. Microbiol. 60 (1994) 756–759.

[5] B. Dupuy, A.L. Sonenshein, Regulated transcription of *Clostridium difficile* toxin genes, Mol. Microbiol. 27 (1998) 107–120.

[6] F.C. Pereira, L. Saujet, A.R. Tome, M. Serrano, M. Monot, E. Couture-Tosi, et al., The spore differentiation pathway in the enteric pathogen *Clostridium difficile*, PLoS Genet. 9 (2013) e1003782.

[7] S.T. Cartman, M.L. Kelly, D. Heeg, J.T. Heap, N.P. Minton, Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between

the *tcdC* genotype and toxin production, Appl. Environ. Microbiol. 78 (2012) 4683–4690.

[8] R.M. Horton, H.D. Hunt, S.N. Ho, J.K. Pullen, L.R. Pease, Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension, Gene 77 (1989) 61–68.

[9] D.G. Gibson, L. Young, R.Y. Chuang, J.C. Venter, C.A. Hutchison 3rd, H.O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases, Nat. methods 6 (2009) 343–345.

[10] R.P. Fagan, N.F. Fairweather, *Clostridium difficile* has two parallel and essential Sec secretion systems, J. Biol. Chem. 286 (2011) 27483–27493.

[11] D.H. Edwards, J. Errington, The *Bacillus subtilis* DivIVA protein targets to the division septum and controls the site specificity of cell division, Mol. Microbiol. 24 (1997) 905–915.

[12] K.E. Dingle, X. Didelot, M.A. Ansari, D.W. Eyre, A. Vaughan, D. Griffiths, et al., Recombinational switching of the *Clostridium difficile* S-layer and a novel glycosylation gene cluster revealed by large-scale whole-genome sequencing, J. Infect. Dis. 207 (2013) 675–686.