

**Phenotypic characterisation of *Clostridium difficile*
strains defective in lipoprotein biosynthesis**

Submitted by Edward Cunningham Farries to the University of Exeter as a dissertation for the degree of Masters by Research in Biosciences, May 2014.

This thesis is available for library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: 

Edward C. Farries, May 2014

Acknowledgements

I would firstly like to thank my supervisors Dr. Stephen L. Michell and Dr. Nicholas Harmer for their excellent supervision and support throughout this project. I would also like to thank all of the others, particularly those in the Bacterial Pathogenesis Research Group at the University of Exeter, who gave their much valued advice.

Special thanks to my parents, Mark and Angela, for your support with funding, housing and your assistance in proof-reading this dissertation.

Finally, thank you to all of the others, especially Toni Hunt, who supported me along the way.

Abstract

Clostridium difficile is regarded as the primary etiological agent of antibiotic-associated diarrhoea, posing a significant challenge to healthcare facilities. The changing nature of *C. difficile* infection is causing an increase in associated disease occurrence outside of the healthcare setting and a gradual move away from the historical association with antimicrobial treatment. Adhesion of spores and vegetative cells to host gut epithelium is thought to be a key aspect of *C. difficile* virulence; disruption of this process may significantly reduce the impact of an infection and the likelihood of infection spread. Lipoproteins are involved in adhesion of *C. difficile* to host tissues and may have roles in other key aspects of virulence. Lipoproteins undergo a specific biosynthesis process within the bacterial cell involving addition of an acyl-glycerol moiety by lipoprotein glycerol transferase (Lgt) followed by signal peptide cleavage by lipoprotein signal peptidase (LspA); disruption of this process may cause attenuation of virulence and a reduction in adhesion to host tissue. *C. difficile* has been shown to encode two functional and homologous lipoprotein signal peptidases: LspA and LspA2. The novel antimicrobials globomycin and myxovirescin directly target lipoprotein signal peptidases and therefore may have potential for use in treatment of *C. difficile* infection. Evaluation of their efficacy against LspA and LspA2 can be determined by protection assays using *Escherichia coli* strains expressing LspA or LspA2 from *C. difficile*.

In this study, both LspA and LspA2 from *C. difficile* are shown to contain the consensus sequences, domains and *in silico* predicted tertiary structure expected of lipoprotein signal peptidases. Characterisation of *C. difficile* strains with silencing mutations in either *lspA* or *lspA2*, in comparison to a wild type, reveals that the absence of either lipoprotein signal peptidase causes an increased survivability in hydrogen peroxide and may affect protein localisation within the bacterium. Finally, successful cloning of *C. difficile* *lspA* and *lspA2* and subsequent expression of LspA and LspA2 via auto-induction in *E. coli* is reported, paving the way for further investigation into the effect of globomycin or myxovirescin treatment.

Contents

Acknowledgements.....	2
Abstract.....	3
Contents.....	4
List of Figures.....	6
List of Tables.....	8
List of Appendices.....	9
Chapter 1. Introduction.....	11
1.1. <i>Clostridium difficile</i> and lipoprotein biosynthesis.....	11
1.2. Pathogenicity of <i>C. difficile</i>	12
1.3. Treatment of <i>C. difficile</i> Infection.....	18
1.4. <i>C. difficile</i> toxins.....	20
1.5. Sporulation and germination in <i>C. difficile</i>	24
1.6. Bacterial Adhesion.....	26
1.7. Lipoprotein biosynthesis.....	28
1.8. Project Aim.....	35
Chapter 2. Materials and Methods.....	40
2.1. Bacterial strains and growth conditions.....	40
2.2. Plasmids used and generated in this study.....	42
2.3. Oligonucleotide primers used in this study.....	43
2.4. Investigation of bacterial lipoprotein signal peptidase proteins.....	44
2.5. Characterisation of growth of <i>Clostridium difficile</i> strains.....	45
2.6. Bacterial protein extractions.....	46
2.7. SDS-PAGE.....	47
2.8. Hydrogen peroxide survival assays.....	47
2.9. Statistical analysis using GraphPad Prism 6.....	48
2.10. <i>C. difficile</i> CD2597 and CD1903 cloning.....	49
2.11. IPTG induction of protein expression.....	54
2.12. Auto-induction of protein expression.....	54
2.13. Western Blotting.....	55
Chapter 3. Bioinformatic analysis of lipoprotein signal peptidases in bacteria.....	58
3.1. Introduction.....	58

3.2.	Multiple alignment of annotated lipoprotein signal peptidase genes .	58
3.3.	Phylogenetic tree of lipoprotein signal peptidas es.....	61
3.4.	Potential sources of LspA2	62
3.5.	Analysis of the predicted structures of LspA and LspA2 in <i>C. difficile</i>	65
3.6.	Discussion.....	67
Chapter 4. Phenotypic characterisation of <i>C. difficile</i> strains defective in lipoprotein biosynthesis and analysis of their response to antimicrobial stresses		
		68
4.1.	Introduction	68
4.2.	<i>C. difficile</i> strains defective in lipoprotein biosynthesis	68
4.3.	Analysis of the growth of <i>C. difficile</i> strains	70
4.4.	Protein profiles of <i>C. difficile</i> strains.....	72
4.5.	Response to hydrogen peroxide oxidative stress.....	75
4.6.	Discussion.....	81
Chapter 5. Expression of <i>C. difficile</i> LspA and LspA2 in <i>E. coli</i>		
		84
5.1.	Introduction	84
5.2.	Cloning Methodology	86
5.3.	Discussion.....	103
Chapter 6. Discussion.....		
		105
Concluding Comments		
		108
Appendices		
		109
Bibliography		
		138

List of Figures

Figure 1.1 Pathogenesis of <i>C. difficile</i>	17
Figure 1.2 The <i>C. difficile</i> 630 PaLoc.....	23
Figure 1.3 Lipoprotein biosynthesis in Gram-positive bacteria.....	29
Figure 1.4 Schematic of type II signal peptidase structure and proteolytic mechanism.....	32
Figure 1.5 Alignment of the amino acid sequences of <i>C. difficile</i> LspA and a functional homolog LspA2 to homologs in <i>E. coli</i> and <i>S. pneumoniae</i> , demonstrating the presence of conserved regions	37
Figure 1.6 Sensitivity of LspA and LspA2 mutants of <i>C. difficile</i> to malachite green	38
Figure 3.1 WebLogo of aligned lipoprotein signal peptidase from various bacteria, highlighting the conserved functional amino acids.....	60
Figure 3.2 Phylogenetic tree of lipoprotein signal peptidase genes from various bacteria.....	63
Figure 3.3 Amino acid sequence alignment of <i>C. difficile</i> 630 LspA2 (CD1903) to peptidase A8 from <i>Lachnospiraceae</i> bacterium.....	64
Figure 3.4 Domain predictions for both copies of lipoprotein signal peptidase in <i>C. difficile</i> 630.....	66
Figure 4.1 PCR confirmation of ClosTron mutants in LspA and LspA2 of <i>C. difficile</i>	69
Figure 4.2 Growth curves of wild type and LspA mutant strains of <i>C. difficile</i>	71
Figure 4.3 SDS-PAGE of protein extractions from <i>C. difficile</i> 630 Δ erm, ECF1, ECF2 and 630 Δ erm Δ lgt	74
Figure 4.4 Effect of hydrogen peroxide concentration on survival of <i>C. difficile</i> strains	76
Figure 4.5 Further investigation of the effect of hydrogen peroxide concentration on survival of <i>C. difficile</i> strains.....	78
Figure 4.6 Effect of incubation time with hydrogen peroxide on survival of <i>C. difficile</i> strains	80
Figure 5.1 Plasmids used for the expression of <i>C. difficile</i> LspA (CD2597) and LspA2 (CD1903) in <i>E. coli</i>	85

Figure 5.2 Anti-His Western blot of protein extractions from non-induced and IPTG-induced cultures of <i>E. coli</i> C43 (DE3) (BL21) harbouring expression plasmids for <i>C. difficile</i> LspA and LspA2 using pQE80 as a backbone	87
Figure 5.3 Cloning methodology and workflow for the expression of LspA and LspA2 from <i>C. difficile</i> in <i>E. coli</i>	89
Figure 5.4 Anti-His Western blot of protein extractions from non-induced and auto-induced cultures of <i>E. coli</i> harbouring expression plasmids for <i>C. difficile</i> LspA or LspA2 using pBAT4 as a backbone	92
Figure 5.5 Restriction endonuclease digests and colony PCRs to demonstrate cloning of CD2597 and CD1903 into pGEM-T Easy (A) and pET28a (B & C)..	95
Figure 5.6 Alignment of the DNA sequences of pEF14 and pEF15 to in silico references to check for sequence errors	96
Figure 5.7 Colony PCR to demonstrate transformation of pET28a, pEF14 and pEF15 into <i>E. coli</i> Rosetta2 (DE3).....	98
Figure 5.8 Analysis of the DNA sequences of pEF14 and pEF15 to in silico references to check for sequence errors post auto-induction.....	99
Figure 5.9 Safe-stain-treated SDS-PAGE and anti-His western blot of crude protein extractions of <i>E. coli</i> Rosetta2 (DE3) harbouring pET28a, pEF14 or pEF15 and <i>E. coli</i> C43 (DE3) (BL21) harbouring pNIC-KSA1 pre- and post-auto-induction.....	101
Figure 5.10 Ponceau-S-stained nitrocellulose membrane and anti-His western blot of low pH glycine protein extractions of <i>E. coli</i> Rosetta2 (DE3) harbouring pET28a, pEF14 or pEF15 and <i>E. coli</i> C43 (DE3) (BL21) harbouring pNIC-KSA1 post-auto-induction.....	102

List of Tables

Table 1 Lipoprotein signal peptidase or Signal peptidase II conserved domain expected sequences.....	31
Table 2 Bacterial strains used in this study	41
Table 3 Plasmids used in this study.....	42
Table 4 Oligonucleotide primers used in this study, all 5' to 3'.....	43
Table 5 Polymerase chain reaction cycle conditions.....	51
Table 6 Polymerase chain reaction constituents	51
Table 7 Ligation reaction mixes and volume calculations	53
Table 8 Media composition for auto-induction of protein expression.....	57
Table 9 Comparison of the expected amino acid sequences of lipoprotein signal peptidase conserved domains given in the literature (Figure 1.4 and Table 1) to the sequences generated in this study showing considerable similarities.....	61

List of Appendices

APPENDIX 1 Multiple alignment of lipoprotein signal peptidase proteins from various bacterial species indicating conserved regions. Accession numbers are in the left-most column.	109
APPENDIX 2 Protein accession numbers for bacterial lipoprotein signal peptidases used in multiple alignments for the generation of Figure 3.1 WebLogo of aligned lipoprotein signal peptidase from various bacteria, highlighting the conserved functional amino acids and Figure 3.2 Phylogenetic tree of lipoprotein signal peptidase genes from various bacteria. A number of proteins have been removed from online databases or combined with similar entries thus these proteins are listed here with identical accession numbers.	110
APPENDIX 3 ClosTron plasmids used to generate <i>C. difficile</i> LspA and LspA2 mutants.....	113
APPENDIX 4 Statistical analysis of the <i>C. difficile</i> growth curves in Figure 4.2 Growth curves of wild type and LspA mutant strains of <i>C. difficile</i> , using GraphPad Prism 6.....	114
APPENDIX 5 Statistical analysis of <i>C. difficile</i> survival under hydrogen peroxide stress using GraphPad Prism 6	122
APPENDIX 6 Cloning strategy for expression of <i>C. difficile</i> LspA and LspA2 in <i>E. coli</i> using vectors pQE80 and pBAT4.....	127
APPENDIX 7 Plasmids generated for expression of <i>C. difficile</i> LspA and LspA2 in <i>E. coli</i> that did not result in detectable expression	128
APPENDIX 8 DNA sequencing data analysis for pEF10.....	129
APPENDIX 9 Rare Codon Caltor analysis of CD2597 and CD1903 genes as found in plasmids pEF10, pEF11, pEF12, pEF13, pEF14 and pEF15, showing considerable rare codon usage in all cases.....	130
APPENDIX 10 Alignments of DNA sequencing results for pEF14 and pEF15 to in silico references, further to the information in Figure 5.6 Alignment of the DNA sequences of pEF14 and pEF15 to in silico references to check for sequence errors	133
APPENDIX 11 Alignments of DNA sequencing results for pEF14 and pEF15 to in silico references using DNA extracted from post-auto-induction culture, further to the information in Figure 5.8 Analysis of the DNA sequences of pEF14	

and pEF15 to in silico references to check for sequence errors post auto-
induction..... 135

APPENDIX 12 Permission to reuse material from Paetzel M, Karla A, Strynadka
NC, Dalbey RE. Signal peptidases. Chemical reviews. 2002 Dec;102(12):4549-
80 (29), as reprinted for Figure 1.4 Schematic of type II signal peptidase
structure and proteolytic mechanism 137

Chapter 1. Introduction

1.1. *Clostridium difficile* and lipoprotein biosynthesis

Clostridium difficile is a Gram-positive, spore-forming, anaerobic bacillus, considered to be the leading cause of antibiotic-associated diarrhoea (AAD). Pathogenic strains of *C. difficile* are recognised as major nosocomial pathogens (1) and increasingly, community-acquired pathogens (2). In recent years a decrease in incidence of *Clostridium difficile* infection (CDI) has been associated with an increase in infection control measures in hospitals (3, 4). Despite this, *Clostridium difficile* disease (CDD) is still a major factor in healthcare, with new strains exhibiting novel phenotypes constantly emerging, such as ribotype 332 which emerged in Scotland in May 2013 with a mortality rate of 75% (3/4) (5).

While the usual clinical association with *C. difficile* is AAD, there is mounting evidence that CDD, or at least colonisation of the human gastrointestinal tract, can occur without pre-treatment with antibiotics (6), with as few as 36% of total CDI patients having been exposed to antibiotics in some studies (2, 7, 8). Initially, patients are taken off their existing courses of antimicrobials to examine potential for recovery and requirement for additional treatment (9, 10). There has been a shift towards automatic treatment with antimicrobials such as fidaxomicin and vancomycin because of the increased proportion of cases involving hypervirulent strains of *C. difficile* (10) such as the ribotype 027, toxinotype III, restriction endonuclease analysis BI and pulsed-field gel electrophoresis NAP1 strains (BI/NAP1/027).

When a patient presents with CDD, symptoms can range from self-limiting diarrhoea (assumed to be AAD), through to pseudomembranous colitis which is fatal in up to 30% of cases (11). The clinical outcome is generally worsened by pre-existing conditions such as inflammatory bowel disease (IBD) or renal failure or administration of anti-peristaltic and immunosuppressant medications (12). These factors, combined with an ageing-related increased incidence (13) results in a much higher mortality rate among those aged over 60 years. Conversely, there is little evidence or theory to explain the increased isolation of *C. difficile* in infants where there is no evidence of disease, with debate covering type of feed (14) and a simple lack of colonisation resistance (15).

Recently, attempts have been made to reduce the impact of severe cases of CDI. A valuable tool for reducing infection impact is predicting the severity and mortality of disease (16). Further, improving knowledge of the epidemiology and pathogenesis of *C. difficile* strains generally (17) and specifically contributes greatly towards understanding CDI, enabling better management of infection and disease.

Lipoproteins are important virulence factors for bacteria (18, 19), and are involved in secretion, the spore cycle, antibiotic resistance (20) and adhesion (21) and form component parts of cellular machinery such as ATP-binding cassette transporters (22-25). As a result, lipoproteins and their biosynthesis are putative targets for antimicrobial therapy. Lipoprotein biosynthesis is a process common to all bacteria and involves a processing cascade of two enzymes in Gram-positive bacteria and three enzymes in Gram-negative bacteria. Inactivation of the enzymes involved has been shown to attenuate virulence in Gram-positive pathogens (26-28). One of the proteins involved in lipoprotein biosynthesis is lipoprotein signal peptidase (Lsp), a type II signal peptidase (SPaseII) (29). Lsp has been shown to contain five conserved amino acid sequence domains (30) which are in conserved locations within the tertiary structure which includes four transmembrane domains. It is thought that Lsp is a valid target for novel antimicrobials; the compound globomycin (31) and its derivatives (32) have been shown to target Lsp. Additionally, myxovirescin (33) also targets Lsp and is potentially better suited for use. *C. difficile* possesses two functional lipoprotein signal peptidases, LspA and LspA2, encoded by CD2597 (lspA) and CD1903 (lspA2) respectively (34-36). It is thought that targeting of one or both of these proteins with antimicrobials would significantly reduce the virulence of *C. difficile*, providing a novel route for treatment of infection and control of diseases caused by this pathogen.

1.2. Pathogenicity of *C. difficile*

1.2.1. Prevalence of infection

Infection with *C. difficile* has frequently been described as epidemic due to the sudden increase in infection rate observed through the last decade. Much of this increase is attributed to the emergence of BI/NAP1/027 strains (37-40) which reportedly cause the most severe form of CDD and are thought to be the most infectious, with strains belonging to other ribotypes, including 078, also

showing prevalence among identified cases (17). It must be noted that ribotype 027 is not the sole cause of severe and widespread cases of CDI. For example, in south east Scotland, ribotype 027 does not seem to be present, with ribotype 001 the most prevalent (41) and in Spain ribotypes 014/020, 001 and 126/078 are the most frequently reported (42). The total number of CDI cases globally has fallen in recent years, in particular in the United Kingdom, where cases have fallen steadily from 55,498 April 2007 to March 2008, to 14,687 April 2012 to March 2013 (43). This fall is attributable to the implementation of mandatory reporting of cases with the initial aim to reduce national cases in the UK by 30% by 2011. In addition, the introduction of significant infection control measures have greatly reduced the likelihood of infection spread at least in the healthcare setting (3, 4).

The community setting, outside of the healthcare environment, has an important role to play in the future of *C. difficile* infection, with indications that CDI in the community is more infectious but causes less serious symptoms. Surprisingly large numbers of cases of CDI are reported as community acquired (42, 44) and it has been suggested that many more go unreported with evidence that CDI is increasing in frequency (7, 45). Intriguingly, the common association of CDI with antibiotic treatment falls away when community acquired disease is examined, with reports of 49% having had no exposure to antibiotics in the previous 3 months (2, 6-8, 46). Also of note, due to implications for healthcare infection control. is an emerging relationship between community-acquired CDI and outpatient care (2, 45, 47), though most patients encounter only low-level outpatient care and are exposed to other potential sources of infection (46).

C. difficile infection is not limited to humans; ribotype 078 commonly affects calves and pigs (48), and horses (49). This, coupled with the frequent isolation of *C. difficile* spores in foods (7, 50, 51), is causing some concern that some incidences of CDI may be acquired from food (52). There is also evidence that infective *C. difficile* can be waterborne (53) through environmental contamination with sewage effluent.

1.2.2. Symptoms of infection

Antibiotic-associated diarrhoea resulting from CDI is described as ranging from brief episodes of loose stools to diarrhoea similar to that caused by cholera, often with over 20 watery stools per day (54). The presence of colitis is

suggestive of the presence of *C. difficile* and can result in serious complications including toxic megacolon and leukocytosis which are parts of pseudomembranous colitis, which is fatal in up to 30% of cases (11). Other symptoms are outlined in the NHS Map of Medicine for *Clostridium difficile* (55) and include fever, bowel perforation and sepsis, with emphasis on the correlation of increasing white cell count with increasing severity, the emergence of a high temperature combined with evidence of severe colitis at the severe stage of disease and the emergence of hypotension, ileus and toxic megacolon at the life-threatening stage.

1.2.3. Association with antibiotic treatment and other predispositions

CDI is primarily associated with antibiotic use and other factors contribute as either predispositions or co-morbidities. The principal antibiotics that are thought to facilitate CDI are clindamycin, broad-spectrum cephalosporins, fluoroquinolones and β -lactams (8, 11, 56) though any antimicrobial has the potential to do so. The association between CDI and antibiotic use is that antimicrobials disrupt the normal gut flora of commensal bacteria (57), increasing the likelihood of infection by removing the inherent protective characteristics of the gut flora (58, 59). The actions of the commensals include providing a simple barrier to *C. difficile* vegetative cell adhesion to intestinal epithelia and remarkably complex signals to the host immune system (59). The nature of these signals may change during infection with *C. difficile* due to increased translocation of commensals through the epithelial barrier, increasing neutrophil recruitment (60). Thus a loss of commensals allows *C. difficile* adhesion to mucosa and epithelia and reduces the induction of an immune response. The association with antimicrobials runs deeper with some strains of the BI/NAP1/027 ribotype, where it has been demonstrated that low levels of ciprofloxacin, a fluoroquinolone, may induce higher levels of toxin production by strains of this ribotype(61). Other medication-based associations include non-steroidal anti-inflammatories (62) and gastric acid suppressors, in particular proton pump inhibitors (46, 62).

There is a varied list of co-morbidities associated with CDI, including heart failure and gastroesophageal reflux disease (2), chronic kidney disease, chronic obstructive pulmonary disease and diabetes (63), immunosuppression, liver

failure and aspects of patient care such as the presence of a nasogastric tube (55). Many of these diseases are also strongly correlated with increasing patient age, indeed there is a high presence of culturable *C. difficile* in those aged over 60 years (13). The effects of ageing, including decreased intestinal mobility and dietary changes combine with the increased rate of disease to elevate both the rate of development of CDI and the mortality rate. Individuals may be colonised with *C. difficile* naturally and asymptotically, infants in particular (13). The most serious pre-existing disease that acts both as a predisposing factor and co-morbidity is inflammatory bowel disease (IBD). Patients with IBD are more likely to contract CDI, resulting in a more severe course of disease (64, 65). Previous or concurrent gastrointestinal infections can also predispose a patient to CDI due to their disruptive effect upon the commensal gut flora, for example a fatal case that occurred after *Salmonella* serotype *Saintpaul* gastroenteritis (66).

1.2.4. Pathogenesis pathway

In general, *C. difficile* infections that lead to a manifesting disease follow a common pathogenesis pathway (67), shown in Figure 1.1. Initially a set of factors, as discussed previously, are involved in disrupting the protective normal gut flora, allowing contaminating *C. difficile* to occupy the available niche and bind to enterocytes and the gut mucosa. The contaminating *C. difficile* could be exogenous spores, from the external environment, or endogenous, from an internal reservoir of infection, or self-contamination, including vegetative cells. Germination of spores follows contamination, a process that is not yet well understood, though bile salts appear to play a key role in stimulation (68, 69). Post-germination, colonisation factors such as surface layer proteins (70), lipoproteins (71) including the novel adhesin CD0873 (21) and other adhesins are expressed by the vegetative cell and promote adhesion followed by the establishment of a colony. Toxin production also begins, which causes a manifestation of symptoms. While some vegetative cells undergo normal cellular division, other cells produce spores which are then released to the surroundings to complete the cycle. The extent to which these steps are followed by a given strain varies greatly. Some strains are non-toxigenic and thus will not cause symptomatic disease, while others such as BI/NAP1/027 strains overproduce toxins and thus cause a more severe disease. Other strains have a different response to germinants (72) and thus may cause a slower

onset of disease while others sporulate less readily, therefore a patient will release less spores into the environment (73).

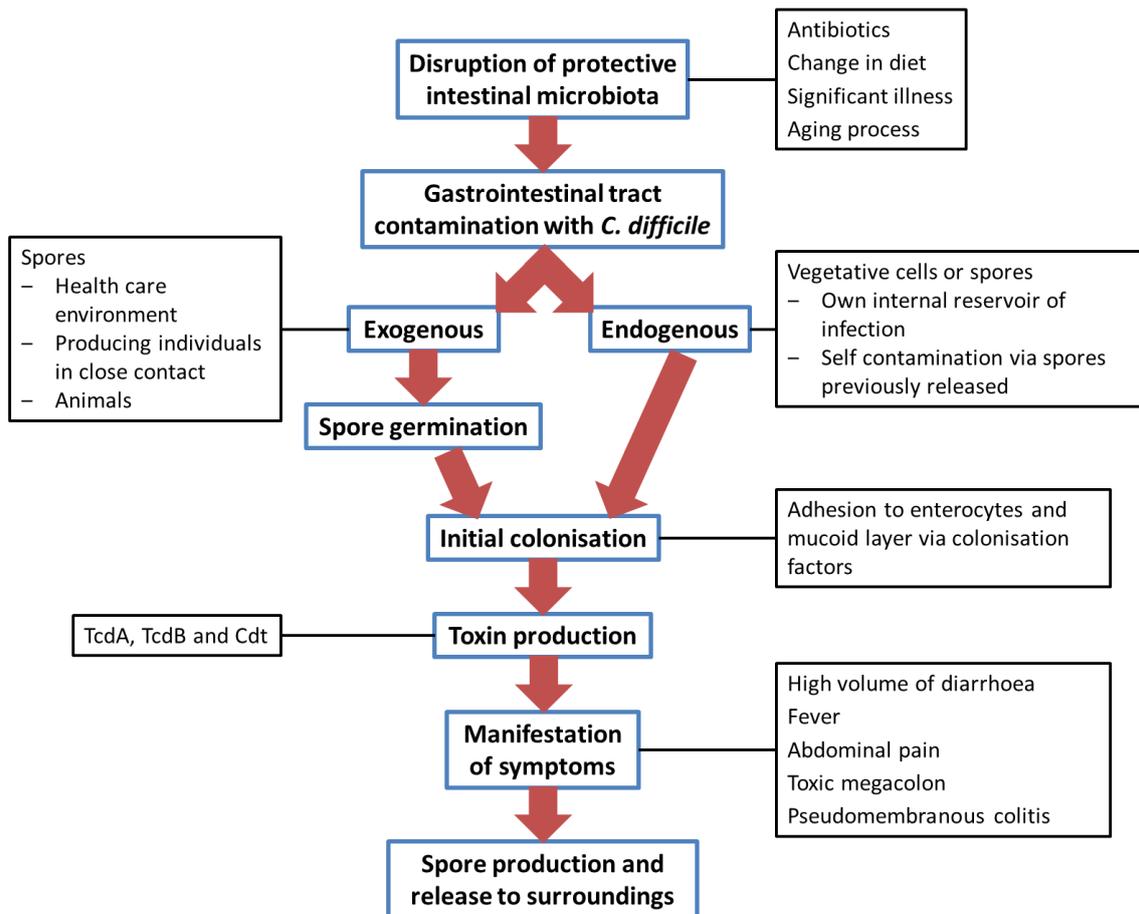


Figure 1.1 Pathogenesis of *C. difficile*

The pathway that *C. difficile* pathogenesis takes from initial predisposition to infection through to manifestation of symptoms and a potential for infection to other individuals is thought to be well conserved. Individuals usually become pre-disposed to infection by disruption of their normal intestinal microflora, allowing the development of a suitable niche for the germination of *C. difficile* spores and the attachment of vegetative cells via adhesins. This leads to an initial colonisation phase, immediately followed by production of toxins TcdA, TcdB and in some strains CDT. These toxins severely damage the gut wall, leading to manifestation of symptoms from a high volume of diarrhoea to pseudomembranous colitis. The infecting *C. difficile* population will start to form spores which will be released to the environment with the potential to infect other individuals or to re-infect the same individual.

1.3. Treatment of *C. difficile* Infection

1.3.1. Current treatments

Currently, there is a typical and standard approach to the treatment of CDI, *C. difficile*-associated diarrhoea (CDAD) and CDD. Initially, any existing treatment with antibiotics is halted allowing for recovery of the normal gut flora (55, 74). This has proven effective for *C. difficile* strains of lower virulence but severe cases require a more significant intervention. As such, therapy has been recommended in all cases, especially due to the increased prevalence of unknown or so called hypervirulent strains (9, 10). Initial therapy involves treatment with metronidazole for moderate cases, vancomycin for severe cases (55, 74) and also fidaxomicin where there is a high risk of recurrence (75); some of the few antibiotics that *C. difficile* does not show resistance to (9, 76). Likelihood of infection recurrence increases with each episode of disease; treatment strategies remain the same however fidaxomicin is the antimicrobial of choice (75, 77). In very severe cases where the patient is presenting with megacolon, perforation or septic shock, it can be necessary to perform a colectomy (75, 78) or other surgical procedures. In addition, the diarrhoeal element of the disease requires management by keeping the patient hydrated and administering electrolytes.

1.3.2. Emerging treatments

Alongside the standard methods described above, new methods for the treatment of disease caused by *C. difficile* are steadily being approved and adopted into the guidelines. Many of these methods are complementary to the primary antimicrobial treatment including administration of probiotics (79), biotherapy with yeasts such as *Saccharomyces boulardii* (75) and administration of other antimicrobial compounds such as teicoplanin, fusidic acid (76) and tirapazamine (80, 81). One method which is gaining considerable recognition is the use of donor stool in faecal transplant (82-85), a method which has been shown to cause rapid restoration of the patient's microbiota in striking similarity to that of the donor, restoring both homeostatic and protective function (86). Similarly, there is evidence to suggest that deliberate patient colonisation with non-toxigenic strains of *C. difficile* can offer protection against challenge with toxigenic strains (87-89). Many of these methods, in particular those involving the administration of antimicrobial compounds, require

significant further testing before they will be recommended for use in the clinical setting (75).

1.3.3. Future treatments

Many more putative treatments being are yet to enter the clinical trial stage. There is a set of experimental antimicrobial compounds that have potential to be applied to *C. difficile*, including NVB302, a novel type B lantibiotic that inhibits cell wall biosynthesis (90), SMT19969, a novel, narrow-spectrum antibiotic being developed specifically for CDI (91) and globomycin (92). Approaches such as treatment with globomycin or its derivatives (32, 93) or the similar myxovirescin (33) aim to inhibit key aspects of virulence, preventing functions such as adhesion by lipoproteins (21, 31, 94) and are discussed in more detail in Section 1.7.3. Other future methods include targeting the primary toxins with antibodies (95), using equine serum, particularly in horses (96), and investigating the protective effects of the food glycome (97, 98).

1.3.4. Prevention of infection and infection control measures

The most significant impact on the number of reported cases of *C. difficile* infection is arguably the implementation of new, stricter regimes for the prevention and control of infection in the healthcare setting. There is evidence to suggest a great potential for infection of otherwise uncolonised patients with *C. difficile* spores from other patients (99), thus hand washing regimes for patients, visitors and staff have helped to reduce spread of disease. In addition, much research has been undertaken to evaluate methods for disinfecting wards or rooms (100-103), including the use of helium (104) and, in particular, hydrogen peroxide (105-107). Antimicrobials such as vancomycin and fidaxomicin have been shown to inhibit outgrowth of spores (77) and efforts to isolate patients immediately on suspicion of infectious diarrhoea (75) combine to reduce environmental contamination with spores.

C. difficile is not restricted to the healthcare setting; the community acquired aspect of the disease cannot be ignored. This is a difficult aspect of the disease to investigate with many cases passing unreported (2). Work in this area is not limited to humans; with significant issues with CDI in horses (17, 49) and the bacterium has been isolated in meat products (50, 51). The reduction in disease incidence by redefining and maintaining standards in the healthcare setting has had a great impact on the disease however, more work is required to reduce

incidence in the community and to protect against the constantly changing nature of the infection.

1.4. *C. difficile* toxins

Typically, pathogenic *C. difficile* strains produce two toxins; TcdA and TcdB, and are thus termed toxigenic strains. TcdA and TcdB are encoded on the pathogenicity locus (*PaLoc*), formerly the toxigenicity locus (108), in conjunction with a set of genes thought to have toxin-related regulatory or secretory functions (109). The *PaLoc* is highly stable and conserved in toxigenic strains with non-toxigenic strains lacking the locus and strains with a defective *PaLoc* still able to cause disease (109). Some strains, notably BI/NAP1/027, also produce a binary toxin, CDT, encoded elsewhere in the genome, which functions similarly to the two primary toxins and may also contribute to virulence (110, 111). The effect and requirement of both primary toxins are unclear; evidence suggests that they act synergistically (112) with TcdB acting secondarily to TcdA but exploiting TcdA-induced tissue damage. Despite this, there is conflicting evidence that suggests that symptoms of *C. difficile* infection can be produced with TcdB alone (113-115) while strains producing TcdA alone have been shown to be avirulent (113). It has since been shown that mutant strains producing only one of the pair can cause disease and that a double mutant exhibits completely attenuated virulence (110, 116). In any case, it is accepted that the primary toxins, TcdA and TcdB, are responsible for causing the damage to the gut epithelium associated with CDD.

1.4.1. Mechanism of action

TcdA and TcdB display significant homology and could result from a gene duplication event. They are modular in structure (117) with three key domains: enzymatic, translocation and receptor binding (118); a structure which enables the toxins to function as Rho-glucosyltransferases (119), targeting the actin cytoskeleton of the host cell. After secretion from *C. difficile*, the toxins are internalised into the host cell via receptor-mediated endocytosis. The low pH environment within endosomes enables a conformational change within the toxin, leading to insertion into the endosomal membrane and thus translocation of the N-terminal enzymatic domain into the host cell cytosol (118). This domain then targets the small GTPase Rho as well as Rac and Cdc42 which are involved in regulation of the cell cycle, catalysing irreversible glucosylation and

inactivation. This halts the host cell's ability to regulate its actin cytoskeleton, causing actin depolymerisation, loss of cell integrity and cell death. CDT has a similar end result; it is also translocated into the host cell where it acts as an actin-specific ADP-ribosyltransferase, again disrupting the cytoskeleton (111). CDT enters the host cell via receptor-mediated endocytosis after binding the lipolysis-stimulated lipoprotein receptor (LSR) and may cause clustering of this receptor into rafts (120).

1.4.2. Regulation of expression and activity

The primary toxins TcdA and TcdB are both encoded on the *PaLoc*, alongside genes *tcdC*, *tcdE* and *tcdR*, thought to encode regulatory functions (121). The arrangement of these genes in the *PaLoc* is shown in Figure 1.2.

TcdE is thought to encode a holin protein that is required for secretion of the primary toxins (122, 123) though this is debatable as toxins have been shown to be released from a *tcdE* inactivation mutant (124).

TcdR is thought to act as an alternative sigma factor, potentially allowing differential expression of the *PaLoc* by binding to RNA polymerase, providing a putative regulation step. There is evidence to suggest that *tcdR* expression is activated by SigD, a flagellar alternative sigma factor, and thus expression of *tcdA* and *tcdB* is activated (125).

TcdC is a putative inhibitory regulator of transcription, however there is confusion about its role in the control of TcdA and TcdB synthesis. It is considered to be an anti-sigma factor; mutation has been shown to play an important role in the differentiation of hypervirulent strains, possibly through a loss of negative regulation (126). Conversely, measuring the transcription levels of genes within the *PaLoc* and expression of TcdA and TcdB in a *tcdC* mutant of the 630 Δ *erm* strain of *C. difficile* revealed only minor differences when compared to the wild type (127). A similar result was achieved using the R20291 strain, from ribotype 027, which has both a frame shift mutation and a deletion within the *tcdC* gene; these mutations were restored using an allelic exchange system, demonstrating that *tcdC* genotype bears no association with toxin production (128).

A wide variety of factors have been implicated in regulation of toxin expression, including flagellar components (129) and the AgrA transcriptional regulator

acting through c-di-GMP (125, 130). Recently a single toxin-antitoxin system, MazEF-cd, has been found in *C. difficile* and affects many parts of pathogenesis, possibly including toxin production (131).

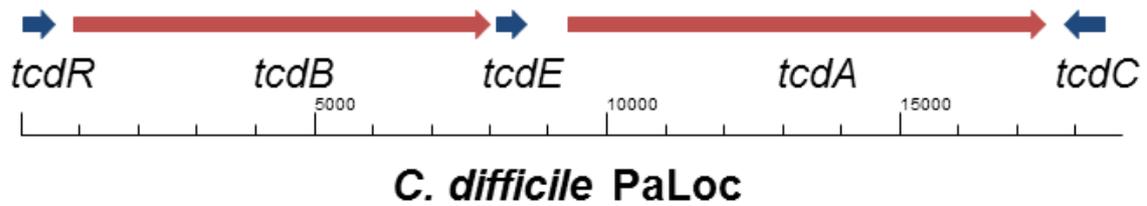


Figure 1.2 The *C. difficile* 630 PaLoc

The two primary toxins of *C. difficile*, TcdA and TcdB, are encoded together on a genomic region named the PaLoc. This region also encodes genes with the following putative regulatory functions: *tcdE* a secretion-related holin; *tcdR* an alternative sigma factor and *tcdC* an inhibitory regulator of transcription.

1.4.3. Toxin variation among strains

There is variation in toxin expression between strains of *C. difficile* and as such a toxinotyping system has been devised which improves upon the previously established serotyping system (132, 133). There are currently 36 distinguishable toxinotypes with designations from 0 through to XXXI (134). These toxinotypes are determined by the expression of TcdA, TcdB and CDT and also by the particular variant of each gene; there are natural variations in the *PaLoc* and CDT, resulting in differing profiles. Such variations can affect the production and properties of each toxin and arise from events such as insertions and deletions (133). For example, an epidemic isolate of BI/NAP1/027, isolated in Quebec, Canada, has been characterised as toxinotype III, with production of TcdA and TcdB 16 and 23 times higher than a toxinotype 0 reference (135). This suggests a relationship between toxinotype and disease severity however there is no direct correlation (136).

1.5. Sporulation and germination in *C. difficile*

Spores are the infectious agent of *C. difficile* and thus the key vehicle of pathogenesis for *C. difficile*. They are produced when vegetative cells enter a period of high stress, usually severe nutrient deficiency, and are a very resilient form of the organism which can withstand most cleaning and sterilisation efforts (17). Most understanding of sporulation and germination comes from studies of *Bacillus subtilis* however recent work has shown that there are considerable differences within the *Firmicutes* (137, 138), revealing the need for specific studies in *C. difficile*.

1.5.1. Regulation of sporulation and germination

Regulation of sporulation is a complex process that is becoming more understood. There are at least 225 genes in the sporulation pathway, possibly representing a simpler ancestral version of sporulation in *Firmicutes* (138). Expression of these genes is controlled by a set of alternative sigma factors: σ^F , σ^E , σ^G and σ^K plus Spo0A, the master sporulation regulator (137), which is itself controlled by the transition phase sigma factor SigH (139). These sigma factors are activated in a cascade of conserved periods of activity that controls the progression of endospore formation (140). Following initiation of asymmetric division controlled by Spo0A, σ^F and σ^E control early stages and σ^G and σ^K control late stages in the mother cell and the forespore respectively. After spore

maturation the mother cell autolyses, releasing the spore.

CDI cannot occur without germination and outgrowth of spores. Therefore there are a number of crucial processes that allow correct germination. At least 511 genes are differentially regulated during germination with co-regulation of functional groups of proteins (141). Key processes include lysis of the spore cortex by SleC (142) in conjunction with the serine protease CspB (143). The spore coat and surface layer carries out other functions both during assembly and into infection with proteins identified as a superoxide dismutase, a manganese catalase and a bifunctional peroxiredoxin and chitinase (144, 145). After germination to vegetative growth, virulence factors such as adhesins (70, 71) and toxins are produced.

1.5.2. Response to bile salts

While bacteria such as *B. subtilis* germinate in response to nutrient stimuli, *C. difficile* does not appear to possess versions of germinant receptors that are found in other bacteria (146). It has been shown that components of bile, particularly the cholate derivative taurocholate, with glycine as a cogerminant, can cause germination initiation both *in vitro* (68, 69) and *in vivo* (147). It has been proposed that these germinants are recognised by a specific protease: CspC, and that bile acid-mediated germination is important for disease in a hamster model of infection (148). The mechanism of taurocholate and glycine binding to a receptor is complex, especially with the addition of a putative germination inhibitor, chenodeoxycholate (149). Chenodeoxycholate is a cholate derivative metabolite produced by normal gut flora, proposed as a natural inhibitor of *C. difficile* germination (150), revealing a mechanism for preferred germination in antibiotic-treated individuals (149).

The situation is complicated by substantial evidence that strains vary in their response to these three compounds (72). This diversity runs deeper than chemical responses; many strains display differing rates and extents of germination (73, 151). It was thought that such diversity explained the “hypervirulence” of certain strains, but there is more variation from strain to strain than type to type (152, 153).

In the *Bacillales*, the signals from germinant receptors enter a transduction pathway that passes the signal to downstream effectors. The first key step in

this pathway is performed by the spore germination protein GerD (154, 155), which functions in both signal transduction and in clustering of germinant receptors into germinosomes (154). In *Clostridiales*, there is no obvious homolog of GerD however, the overall signal transduction process is the same indicating the presence of an alternative (155). GerD has been shown to localise to the inner membrane of *Bacillus subtilis* spores and analysis has shown it to be a lipoprotein (156). It may also localise to other parts of the spore however, because the protein appears to be shed during germination (156), where GerD is associated with the inner membrane it is most likely to be functional and this localisation is the most appropriate for any interaction with germination receptors and downstream effectors (157). This evidence suggests that the protein that the *Clostridiales* order uses to perform the function of GerD is a lipoprotein, localised to a similar part of the spore.

1.6. Bacterial Adhesion

The first step in CDI is thought to be adhesion of live cells to exposed colonic epithelia. This process is mediated by components of the bacterial surface layer called adhesins; in *C. difficile* these include the surface layer protein SlpA (158), members of the cell wall protein family (70), components of the flagella (159, 160) and a lipoprotein called CD0873 (21).

1.6.1. Surface layer proteins and adhesins

The outermost layer of bacteria is a proteinacious network which has important roles in growth and survival (161) and is termed the surface or S-layer. The S-layer has two primary constituents; the monomeric high- and low-molecular weight surface layer proteins (SLPs) which bind non-covalently together to form the H/L complex (162), a regularly spaced 2 dimensional array (161). These proteins follow dedicated pathways of secretion and anchoring to the bacterial cell wall; the majority of translocation of these and other proteins across the cell membrane is performed by the Sec secretion system, of which *C. difficile* has two (162).

Once anchored to the cell wall SLPs play an important role in virulence of bacterial pathogens including *C. difficile*, specifically in adhesion to host tissues and interaction with the host immune system (161). For example, the human toll-like receptor 4 (TLR4) found on the surface of immature dendritic cells, has been shown to recognise *C. difficile* SLPs, causing the maturation of these

dendritic cells, generation of T-cell helper cells thus activating both innate and active immune systems (163). Of the two SLPs, the high-molecular weight protein has been shown to be the most involved in *C. difficile* binding to host cells via two major polypeptides and ligands in the extracellular matrix including collagen I, thrombospondin and vitronectin (70). Other proteins in the S-layer have been shown to be involved in adhesion to the host including CbpA which has high binding affinity to collagen I and collagen V (164) and members of the cell wall protein (CWP) family which has significant homology to the SLPs (70, 161). CWP proteins have other roles besides host adhesion; for example CwpV has a conserved aggregation-promoting function by which it may assist in anchoring neighbouring cells (165). One major contribution to adherence of *C. difficile* to the host appears to come from SlpA (158) with variations in the encoding gene allowing identification of strains though this does not alter the adhesion properties of the particular strain (166). Another significant contribution comes from flagellar components, as specifically demonstrated in the 027/BI/NAP1 R20291 strain, in a study which also demonstrated inter-strain variation; the 630 Δ *erm* strain does not require flagellar components for colonisation (160).

1.6.2. Lipoproteins

Lipoproteins are a set of proteins grouped by their method of secretion and anchoring which perform many important functions in bacteria. These functions include roles in protein secretion, the spore cycle, sensory systems, antibiotic resistance and adhesion (20, 21), biogenesis of other systems (167) and protection from superoxide (168). Lipoproteins appear to be particularly involved in transport systems, ATP-binding cassette (ABC) transporters in particular (23-25, 169), and are frequently shown to be required for virulence (18, 19, 24). In Gram-negative bacteria lipoprotein biosynthesis appears to be essential for viability while the process may be dispensable in Gram-positive bacteria (22). Of particular interest to this study is the potential for a role of lipoproteins in the adhesion of *C. difficile* spores and vegetative cells to host tissue with the lipoprotein CD0873 having been recently identified as a definitive adhesin, the first such identification in *C. difficile* (21).

1.7. Lipoprotein biosynthesis

Lipoprotein biosynthesis in its simplest form is common to both Gram-positive (Figure 1.3) and Gram-negative bacteria. Initially, the translated pre-prolipoprotein undergoes modification with a diacylglycerol moiety by lipoprotein diacylglycerol transferase (Lgt) at a conserved cysteine residue within the so called lipobox of the protein (22, 170). This prelipoprotein then undergoes signal peptide cleavage at the cysteine residue by lipoprotein signal peptidase (LspA), a type II signal peptidase, to form what constitutes a mature lipoprotein in Gram-positive bacteria. Gram-negative bacteria and some Gram-positive bacteria have a final step in which a third acyl moiety is added to the N-terminus of the cysteine by lipoprotein N-acyl transferase (Lnt). The two or three acyl moieties enable insertion of the lipoprotein into the cell membrane and it is suggested that the third moiety in Gram-negative bacteria could have a role in localisation.

Inactivation of the enzymes involved in lipoprotein biosynthesis have been shown to attenuate virulence in Gram-positive pathogens such as *Mycobacterium tuberculosis* (26, 27) and *Bacillus anthracis* (28) however this does not seem to hold true for all, *Streptococcus suis* for example (171). Full virulence studies have not been conducted in other species where lipoprotein biosynthesis has been disrupted but these bacteria do display substantial differences in growth and other phenotypes (172-176).

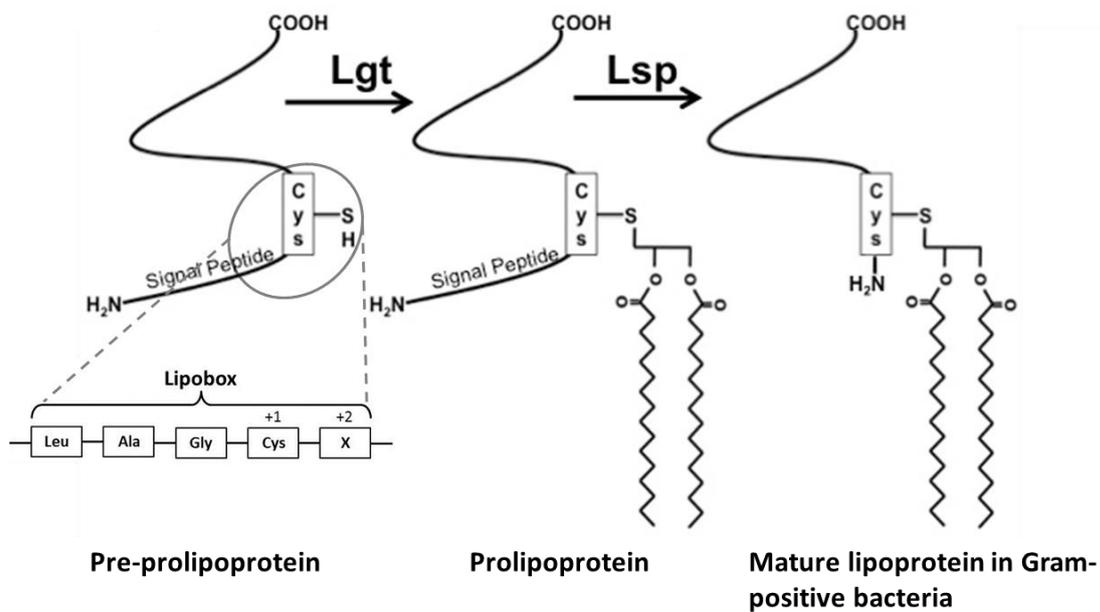


Figure 1.3 Lipoprotein biosynthesis in Gram-positive bacteria

All types of bacteria follow the same conserved mechanism for the biosynthesis of lipoproteins except that Gram-negative and mycobacteria perform additional steps in accordance with their additional protein-localisation requirements. Gram-positive bacteria perform the minima of steps: initially a freshly translated pre-lipoprotein is modified with a diacyl-glycerol moiety at a conserved cysteine residue within the lipobox region of the signal peptide by lipoprotein glyceryl transferase (Lgt). The resultant prolipoprotein is then modified via the cleavage of the signal peptide immediately upstream of the conserved cysteine residue by lipoprotein signal peptidase (Lsp), producing a mature lipoprotein.

1.7.1. Lipoprotein signal peptidase

The structure of a protein is closely related to its function; primary structure places the required functional and structural amino acids in the desired order, secondary structure folds the protein into helices and sheets that allow the tertiary structure to fold the protein into a conformation that brings the amino acids for the active site into correct alignment and permits interaction by the protein with the cell. A lipoprotein signal peptidase (Lsp) or signal peptidase II (SPasell) has a specific predicted structure, first reported in Tjalsma *et al* (30), based upon studies in *B. subtilis*. Tjalsma *et al* (30) also reports on analysis of the amino acid sequences of other bacteria and concludes that there are five conserved sequence domains between all proteins. This conclusion is supported by later work in Rahman *et al* (177) and Paetzel *et al* (29), in which the nomenclature of the domains were changed from I through V to A through E. The tertiary structure of lipoprotein signal peptidases, including the approximate expected locations for the five conserved amino acid sequence domains is shown in Figure 1.4, reproduced from Paetzel *et al* (29), and has been predicted by various membrane topology predicting algorithms in that study, Tjalsma *et al* (30) and, more recently, in Dalbey *et al* (178). All agree that the protein should take on a four transmembrane domain structure which allows localisation of the conserved domains to important positions. Specifically domains C and D, which contain the two aspartic acid residues thought to form the active site. A model mechanism for the proteolytic cleavage of signal peptides is also displayed in Figure 1.4. While the five conserved amino acid domains are thought to contain conserved sequences, the exact sequence location of each domain within each protein is subject to variation. They do not occur at a given amino acid number though they do always occur in order. The predicted amino acid sequence of each domain has changed between studies while the positioning and general composition of the domains has remained the same. The sequences of each domain given in each study are shown in Table 1.

Table 1 Lipoprotein signal peptidase or Signal peptidase II conserved domain expected sequences

The amino acid sequences of the five conserved domains predicted across lipoprotein signal peptidases are given according to the study that reported them. The relative frequency of each amino acid in consensus sequences is denoted by the type of lettering: uppercase indicates a very high frequency; lowercase indicates a low frequency and 'x' indicates no evidence of conservation at that residue. Amino acid frequency or level of conservation was not analysed by Rahman *et al* (177).

Domain	Tjalsma <i>et al</i> (30)	Paetzel <i>et al</i> (29)	Rahman <i>et al</i> (177)
A	dqxxk	dxxtk	DQLSK
B	NxGaaf	NxGaaf	NYGISF
C	ixggalgNxxDr	iiggaxlgNxxDr	VIGGAVGNLIDR
D	VvD	vvd	VFD
E	FNxAD	FNxAD	FNLAD

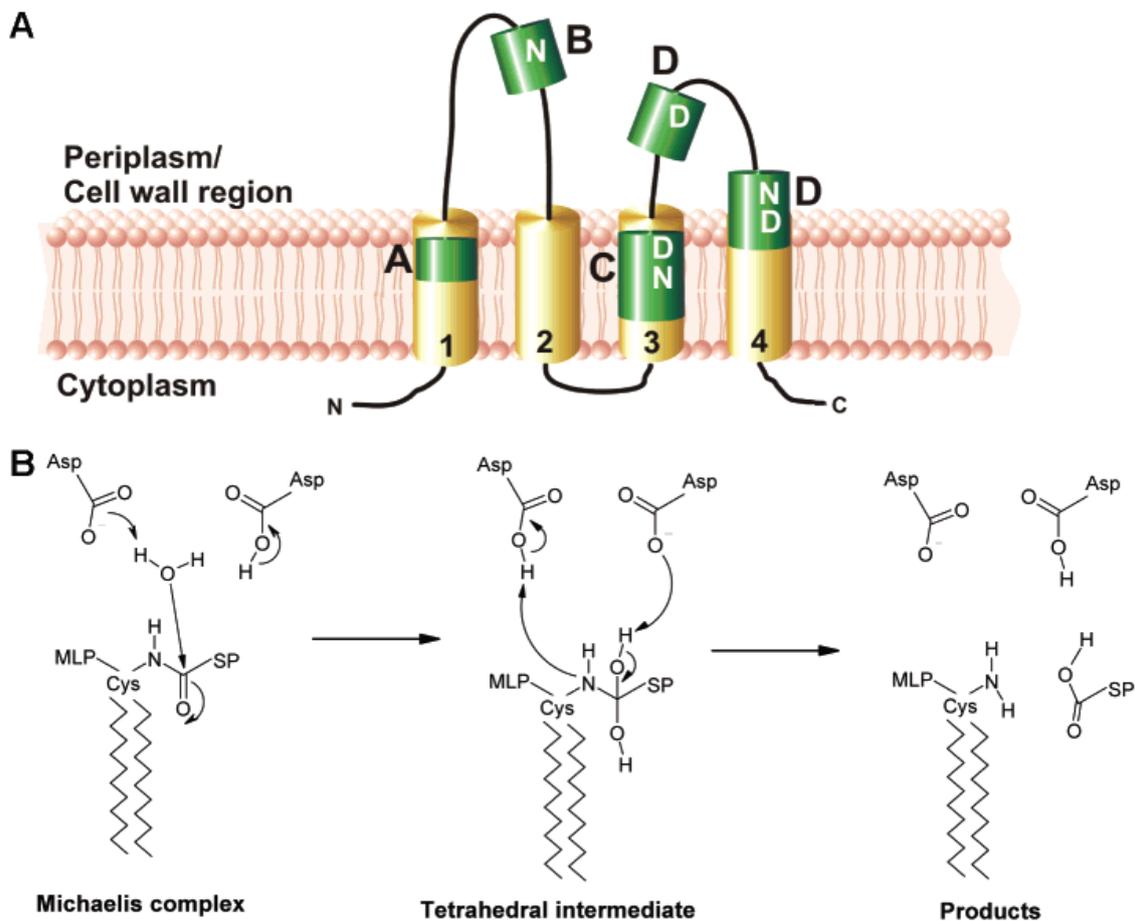


Figure 1.4 Schematic of type II signal peptidase structure and proteolytic mechanism

Figure reprinted from Paetzl M, Karla A, Strynadka NC, Dalbey RE. Signal peptidases. *Chemical reviews*. 2002 Dec;102(12):4549-80.(29). Copyright 2014 American Chemical Society. Showing, **A**: the expected membrane topology of type II signal peptidases as first reported in Tjalsma *et al* (30), showing the approximate positions of active amino acids (Aspartic acid, Asp, D in white letters) and the conserved amino acid domains (A through E; labelling error carried over from the original) [green boxes]. **B**: the predicted proteolytic mechanism of signal peptide cleavage performed by this type of enzyme.

There is a number of reasons for the variation in reported sequences. The study by Rahman *et al* (177) focussed on proteins from bacteria closely related to *Rickettsia typhi* and only examined 10 genes. The original study, Tjalsma *et al* (30), used 19 genes, representing a retrospective improvement in terms of representation of the majority of bacteria. Paetzel *et al* (29) examined 54 genes, representing the best set of all studies. However, due to the small numbers of species sampled in each study, not one of these studies should be considered the superior source for the sequence of these domains. Despite the potential for variation in the absolute common sequence of each domain their general form remains the same, providing an indicator for the functionality of these proteins. As a result of the data presented in these studies, it is clear that a lipoprotein signal peptidase has a number of features; the presence of the five conserved domains approximately located to allow function by the tertiary structure which includes the four transmembrane domains.

1.7.2. Targeting of the lipoprotein biosynthesis pathway

Lipoprotein biosynthesis is clearly a potential target for the development of novel antimicrobials due to the importance of this process for Gram-negative viability and Gram-positive virulence. It is thought that, even if these antimicrobials were not completely bactericidal and instead decreased virulence and in particular prevented bacterial adhesion, they would be highly effective against CDI. One emerging antimicrobial that has specific action against LspA is globomycin. It was initially shown that globomycin treatment caused an increase in levels of prolipoprotein in *Escherichia coli* associated with an inhibition of growth and that lipoprotein-negative mutants were resistant to globomycin (179). Subsequently, it was shown that the action was specific to LspA, and that it was non-competitive and tighter than the binding of the prolipoprotein substrate (31). Globomycin may have additional actions on other aspects of the bacteria that cause lethality as it was shown to kill *M. tuberculosis* independently of LspA (180) however this does not detract from its potential for use as an antimicrobial. It has been shown that analogues of globomycin can, in fact, have more potent activity than globomycin itself (32) and that the variation in activity between analogues is structure-dependent (181). Experiments to evaluate the antimicrobial activity of globomycin against *C. difficile* would shed valuable light on its suitability for targeting lipoprotein signal peptidases in this organism. Protection assays with *E. coli* expressing *C.*

difficile LspA or LspA2 such as those performed by Rahman *et al* (177) and Khandavilli *et al* (24) would provide evidence that these enzymes are targets of globomycin in *C. difficile* and that other factors are not involved.

1.7.3. Globomycin alternatives

Until recently, there has been a very low availability of globomycin, partly due to the challenges posed by synthesis. Alternatives exist that allow feasibility testing while other compounds such as the aforementioned analogues may prove to be better candidates. The common antimicrobial and spore stain malachite green that has been used to treat infections in fish and humans prior to emerging evidence on its toxic properties has been shown to interfere with the growth of bacteria such as *M. tuberculosis* (182). The antimicrobial activity of malachite green stems from its action as a respiratory enzyme poison (183). It was subsequently shown that lipoprotein biosynthesis is essential for the resistance of *M. tuberculosis* to malachite green via the generation of an *lspA* mutant strain (184). A strain mutant in the P27-P55 operon which encodes P55, an efflux pump, was shown to be hypersensitive to malachite green and also caused more rapid decolourisation of the compound (185). Decolourisation of malachite green is generally caused by reduction to leucomalachite green, a colourless and water-insoluble compound, thought to occur very slowly in the absence of air (183). Intestinal bacteria have been shown to perform this reduction (186), typically using an enzyme called triphenylmethane reductase (Tmr) (187). The latter study also demonstrated that lipoproteins are involved in the effective sequestration of malachite green away from Tmr and it is thought that such sequestration could reduce the toxicity of malachite green to bacteria by sequestration away from respiratory enzymes. It is suggested that malachite green could be used to assay the requirement of lipoprotein biosynthesis in resistance of a bacterial species to this toxic stress and thus shed light on the likely outcome of the use of globomycin as an antimicrobial against that bacterium.

As a potential alternative to rare globomycin, the *Myxobacterium* secondary metabolite myxovirescin, or TA had been shown to inhibit LspA (33, 188). This study also compared the effects of myxovirescin to those of globomycin and found that myxovirescin has a better whole-cell potency and thus may, in fact, be a better option for treatment of CDI and associated diseases than

globomycin. *Myxococcus xanthus*, the myxovirescin producer strain, has four different lipoprotein signal peptidases which have recently been shown to have potential roles in protection against myxovirescin (33, 188), indicating that the previously mentioned protection assays could be used with this compound. Similar to globomycin, myxovirescin is difficult to obtain and purify (189-191) however the paper that initially identified myxovirescin as an inhibitor of LspA utilised a lawn of an *M. xanthus* strain that overproduces the compound; an approach that would be difficult to implement in *C. difficile* due to the need for incubation under anaerobic conditions.

1.8. Project Aim

In previous work, it was shown that *C. difficile* encodes two potential lipoprotein signal peptidases, LspA and LspA2 (34-36). LspA2 was initially considered as a lipoprotein signal peptidase as the gene encoding this protein, CD1903, was potentially paralogous to *LspA*, CD2597, in *C. difficile* and orthologous to LspA-encoding genes in other bacteria (Figure 1.5). LspA2 is a putative functional homolog of LspA and contains the key set of functional amino acids found in type II signal peptidases. Previous work used targeted insertional mutagenesis via the ClosTron system (192-194) to generate mutant strains in either *LspA* (*C. difficile* ECF1) or *LspA2* (*C. difficile* ECF2) and demonstrated that both strains display increased sensitivity to malachite green (Figure 1.6) thus indicating their involvement in lipoprotein processing.

Another method for characterising the functionality of LspA and LspA2 is a globomycin protection assay, as used in Rahman *et al* (177) and Khandavilli *et al* (24). This assay relies upon the expression of a non-native lipoprotein signal peptidase in *E. coli* to cause a decrease in globomycin susceptibility. To achieve this, *C. difficile* *LspA* and *LspA2* would need to be separately cloned into an expression vector that would allow induction of expression of each protein once the vector was transformed into *E. coli*. The strains generated could then be subjected to globomycin treatment with and without induction of expression and the survival of each strain calculated.

There are other aspects of lipoprotein biosynthesis in *C. difficile* and the effects of its disruption on bacterial phenotype that need to be investigated. There is evidence that disruption of certain lipoproteins causes increased susceptibility to oxidative stress (168) and that there is a link between susceptibility to

oxidative stress and adhesion (195). Knowledge of the effect of these mutations on the sensitivity of *C. difficile* to other existing antimicrobials may also shed light on the best course of action to pursue with regards to the development of a novel antimicrobial that exploits lipoprotein biosynthesis.

```

E. coli LspA      1 ----MKIKALL-ILFIFLPLIGCDRYTKEKAIVS LKQEPASF-----FN
C. diff LspA     1 -----LL-YILIIILLIGLDQLSKIWVLNNIVDVSTIPI-----IN
C. diff LspA2    1 LQGGVNIRQVKSFVFPVISLIFLDQISKVLIGLFLMDFEIDIIGKFLRFN
S. pneum LspA    1 ----MKKRAIV-AVIVLL-LIGLDQLVKSIVQQIPLGEVRSW-----IP

E. coli LspA     41 GI--FTLTYHENTGMLSLGADLPENVRHIIF TLMVGAVLLSGLAYILIK
C. diff LspA     36 NV--FHLTYVENRGAAGFL---LQNN--QWIFIIVALLATVFGLYYLINTR
C. diff LspA2    51 PVQNTNLSYGGNFIGILSNLWVL-----VLENILVILVIISGYAFYKSK
S. pneum LspA    40 NF--VSLTYLQNRGAAFSILQD-----QQLLEAVITLVVVIGAIWYIHKH

E. coli LspA     89 PMNKLSFSVGL-IM---LSGGLGNIYDRVLNEGRVVD FMLLQIGPLRTGV
C. diff LspA     79 KVHIFGRLGII-LI---ISGALGNLIDRV-RLGFVVDYDFDRI--IWEYV
C. diff LspA2    95 -NEQTSYSVKV-IMSCGLAGTICSLIDK-LFWGGSLDF--LQIPSF--FI
S. pneum LspA    83 MEDSFWMVLGLTLI---IAGGLGNFIDRV-SQGFVVD MFHLDLDF--INFAI

E. coli LspA     135 FNVADVAIMAGLFGF----IFISSKSGKQLTNQSN-----
C. diff LspA     122 FNIADV FVVVGTVFLCIYVLF FESKSR-----
C. diff LspA     138 FDLKDCYLTVAEIIIF----VVIGILHNREISMKEYIYFCYRQFKR
S. pneum LspA    127 FNVADSYLTVGVI-----ILLIAMLKEEINGN-----

```

Figure 1.5 Alignment of the amino acid sequences of *C. difficile* LspA and a functional homolog LspA2 to homologs in *E. coli* and *S. pneumoniae*, demonstrating the presence of conserved regions

An alignment of the amino acid sequences of the *C. difficile* proteins encoded by genes CD25870 (LspA) and CD19030 (LspA2) to orthologous proteins in *Escherichia coli* and *Streptococcus pneumoniae*, as shown in Farries, 2012 (34). There are clear conserved residues across all 4 proteins indicating a potential for a common function, suggesting that LspA2 may be a paralog of LspA.

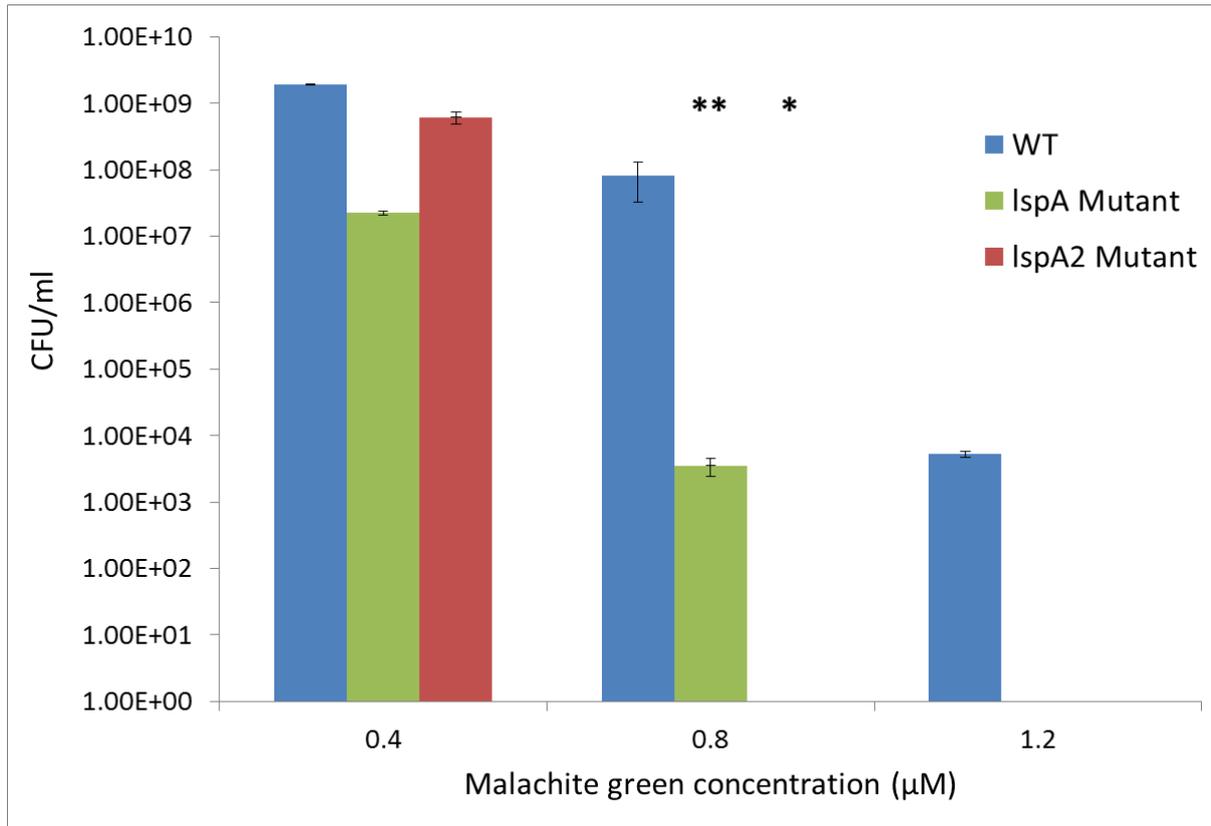


Figure 1.6 Sensitivity of IspA and IspA2 mutants of *C. difficile* to malachite green

C. difficile strains mutant in either *IspA* or *IspA2* were grown alongside the *C. difficile* 630Δ*erm* wild type strain in the presence of malachite green, an antimicrobial. Colony forming units (CFU) of each were enumerated and plotted graphically, indicating that both mutant strains are more susceptible than the wild type and that the *IspA2* mutant strain is the more susceptible of the two. Statistics displayed: * indicates a P-value ≤ 0.05, ** indicates a P-value ≤ 0.01. Statistical P-values were calculated in Microsoft Excel using a Bonferroni correction comparing mutant strains to the wild type after a two-tailed ANOVA performed by the Data Analysis Toolpack. Figure adapted from Farries *et al*, 2013 (36).

This study aims to further characterise the phenotypes of *C. difficile* strains ECF1 and ECF2 by analysis of growth characteristics and response to hydrogen peroxide. This study also aims to analyse the two *C. difficile* lipoprotein signal peptidases, LspA and LspA2, bioinformatically and by generating *E. coli* strains capable of expressing *C. difficile* 630 LspA or LspA2 for use in globomycin protection assays.

Chapter 2. Materials and Methods

2.1. Bacterial strains and growth conditions

2.1.1. General bacterial growth conditions

The strains used in this study are listed in Table 2. Liquid cultures of *E. coli* strains were grown in Luria Bertani (LB) broth [Miller] at 37°C with 200 rpm orbital shaking unless otherwise stated. Plate cultures of *E. coli* were grown on Luria Bertani broth [Miller], supplemented with No2 Bacteriological Agar [Fisher] at 37°C. Liquid cultures of *C. difficile* were grown in Brain Heart Infusion (BHI) Broth [Sigma, or Oxoid] at 37°C and 70% humidity in a Don Whitley DG 500 Anaerobic Workstation supplied with BOC Anaerobic Mix gas (10% CO₂, 10% H₂, N₂). Plate cultures of *C. difficile* were grown on BHI Agar [Oxoid] under the same conditions. Supplements and antibiotics appropriate to the strain or experiment being performed were added to the growth media where required. *E. coli* Rosetta2 (DE3) cultures were always grown in media supplemented with 2% glucose to inhibit induction of protein expression unless desired and chloramphenicol at a concentration of 34 µg ml⁻¹ to select for the pRARE2 plasmid, with the addition of other antibiotics as required. Where used, other antibiotics were used at the following concentrations: 75 µg ml⁻¹ kanamycin; 5 µg ml⁻¹ erythromycin; 100 µg ml⁻¹ ampicillin. Agar plates used in transformations of pGEM-T Easy plasmids were also supplemented with 0.5 mM IPTG and 80 µg ml⁻¹ X-Gal.

2.1.2. Chemically competent *Escherichia coli*

Chemically competent cultures of *E. coli* strains DH5α, TOP10 and Rosetta 2(DE3) were generated by growing cultures in 100 ml LB medium at 37°C and 200 rpm shaking to an OD₅₉₅ of 0.5 – 0.7. Culture was then incubated on ice for 10 minutes followed by centrifugation at 4,000 x g and 4°C for 10 minutes. Cells were resuspended in 50 ml sterile, ice cold 0.1 M CaCl₂ solution and incubated on ice for 30 minutes. Cells were centrifuged as before and resuspended in 600 µl 0.1 M CaCl₂ followed by the addition of 300 µl 50% glycerol, aliquoting 200 µl to 1.5 ml Eppendorf tubes and storage at -80°C or immediate use in transformation experiments.

E. coli HB101 chemically competent cells were purchased from Promega, UK and used according to the manufacturer's specifications.

Table 2 Bacterial strains used in this study

Strain	Relevant Genotype	Description	Source
<i>C. difficile</i> 630 Δerm		Laboratory strain, used as wild type	Hussain <i>et al</i> (196)
<i>C. difficile</i> ECF1	$\Delta lspA$ (CD25970:: <i>ermB</i> ::321 322)	<i>C. difficile</i> 630 Δerm <i>lspA</i> ClosTron mutant. <i>Erm</i> ^R	Previous work (34)
<i>C. difficile</i> ECF2	$\Delta lspA2$ (CD19030:: <i>ermB</i> ::317 318)	<i>C. difficile</i> 630 Δerm <i>lspA2</i> ClosTron mutant. <i>Erm</i> ^R	Previous work (34)
<i>C. difficile</i> 630 Δerm Δlgt	Δlgt :: <i>ermB</i>	<i>C. difficile</i> 630 Δerm <i>lgt</i> ClosTron mutant. <i>Erm</i> ^R	Gift from A. Kovacs-Simon
<i>C. difficile</i> KSA1	CD0873:: <i>ermB</i> ::317 318	<i>C. difficile</i> 630 Δerm CD0873 ClosTron mutant. <i>Erm</i> ^R	A. Kovacs-Simon <i>et al</i> (21)
<i>E. coli</i> DH5 α		High-efficiency transformation strain	Invitrogen
<i>E. coli</i> TOP10		High-efficiency transformation strain	Invitrogen
<i>E. coli</i> Rosetta2(DE3)	pRARE2 <i>Cam</i> ^R	Expression strain able to produce tRNAs for rare codons	Novagen
<i>E. coli</i> HB101		High-efficiency transformation strain	Promega
<i>E. coli</i> C43 (DE3) (BL21)		Expression strain, tolerant to toxic proteins	(197)

2.2. Plasmids used and generated in this study

The plasmids used in this study are listed in Table 3. The construction of plasmids produced by this study is explained in Section 2.10.2.

Table 3 Plasmids used in this study

Plasmid	Characteristics	Source
pQE80	Expression plasmid: Amp ^R Cam ^R N-terminal 6xHis	Quiagen
pBAT4	Expression plasmid: Amp ^R <i>lacI</i> T7 <i>lac</i>	Peränen <i>et al</i> (198)
pET28a	Expression plasmid: Kan ^R <i>lacI</i> C-terminal 6xHis	Novagen
pGEM-T Easy	TA cloning vector, Amp ^R <i>lacZ</i>	Promega
pRL4A	pQE80/(His-)CD1903	R. Lawrence, unpublished
pRL6A	pQE80/(His-)CD2597	R. Lawrence, unpublished
pEF10	pBAT4/mCD2597-His	This study
pEF11	pBAT4/iCD2597-His	This study
pEF12	pBAT4/mCD1903-His	This study
pEF13	pBAT4/iCD1903-His	This study
pGEM-T Easy::CD2597pET	Intermediate plasmid for pEF14	This study
pGEM-T Easy::CD1903pET	Intermediate plasmid for pEF15	This study
pEF14	pET28a/CD2597(-His)	This study
pEF15	pET28a/CD1903(-His)	This study
pNIC-KSA1	pNIC28 carrying CD0873	A. Kovacs-Simon <i>et al</i> (21)

2.3. Oligonucleotide primers used in this study

The oligonucleotide primers used for PCR in this study are shown in Table 4.

Table 4 Oligonucleotide primers used in this study, all 5' to 3'

Name	Sequence	Characteristics
CD2597_pBAT4_F	TCCATGGCTATGCTATATATATTAATAATA	PCR for pEF10 insert with base insertion and Nco1 site
CD2597_pBAT4_F2	CCCATGGCTTATATATTAATAATAATTCTA	PCR for pEF11 insert with base deletion and Nco1 site
CD2597_pBAT4_R	CCCATGGTTAGTGGTGGTGGTGGTGGTG CCTACTTTTACTTTCAAAAAAT	PCR for pEF10 and pEF11 with His tag and Nco1 site
CD1903_pBAT4_F	TCCATGGCTATGCAAGGAGGTGTTAATA TC	PCR for pEF12 with base insertion and Nco1 site
CD1903_pBAT4_F2	TCCATGGCTGGAGGTGTTAATATCAGGC AA	PCR for pEF13 with base deletion and Nco1 site
CD1903_pBAT4_R	ACCATGGTTAGTGGTGGTGGTGGTGGTG TCTTTTAAACTGACGATAGC	PCR for pEF12 and pEF13 with His tag and Nco1 site
EFTC_pET28a_CD6 3025970_F	GATCCCATGGGCCTATATATATTAATAAT AATTCTACTC	PCR for pEF14 with Nco1 site
EFTC_pET28a_CD6 3025970_R	GATCCTCGAGCCTACTTTTACTTTCAAAA AATAAAAC	PCR for pEF14 with Xho1 site
EFTC_pET28a_CD6 3019030_F	GATCCCATGGGCCAAGGAGGTGTTAATA TCAG	PCR for pEF15 with Nco1 site
EFTC_pET28a_CD6 3019030_R	GATCCTCGAGTCTTTTAAACTGACGATAG	PCR for pEF15 with Xho1 site
CD2597_pCheck_L	ATAGAGGTGCAGCATTTGG	Sequencing of CD2597 insertions
CD2597_pCheck_R	AACCTAATCGTACTCTATC	Sequencing of CD2597 insertions
CD1903_pCheck_L	TTTATGGGTATTGGTATTG	Sequencing of CD1903 insertions
CD1903_pCheck_R	TTCTGCAACAGTAAGGTAG	Sequencing of CD1903 insertions
T7 Terminator	TAATACGACTCACTATAGGG	PCR screening of pET28a ligations
T7 Promoter	TAGTTATTGCTCAGCGGTGG	PCR screening of pET28a ligations

2.3.1. **Electrocompetent *Escherichia coli***

Aliquots of *E. coli* DH5 α , TOP10 and Rosetta2 (DE3) cultures for electroporation were generated by growing in 100 ml of LB to OD_{595 nm} = 0.7 at 37°C with shaking at 200 rpm. Cells were then harvested in 50 ml aliquots at 3,500 rpm for 10 minutes at 4°C followed by three repeated washes in 50 ml ice cold 10% glycerol. Cells were finally resuspended in 1 ml ice cold 10% Glycerol, aliquoted to 1.5 ml Eppendorf tubes in 50 μ l volumes and frozen at -80°C or used immediately.

2.4. **Investigation of bacterial lipoprotein signal peptidase proteins**

2.4.1. **Sourcing of lipoprotein signal peptidase sequences**

The amino acid sequences of bacterial lipoprotein signal peptidases were retrieved from online repositories by conducting specific searches. For proteins analysed by Paetzel *et al* (29) or Rahman *et al* (177), the accession numbers given were used as search terms. In some cases more than one result was returned therefore the protein with the most similar description (gene name, source bacterium) to that in the literature was selected. For other proteins, such as those belonging to *Clostridia* or *Myxococcus* species, searches were conducted for the species and anticipated gene name; *IspA*. Amino acid sequences were imported into Clone Manager Professional Suite and stored for later use.

2.4.2. ***In silico* alignments of lipoprotein signal peptidase amino acid sequences**

To allow comparison of the amino acid sequences of lipoprotein signal peptidases from a variety of bacteria, amino acid sequences sourced in 2.4 were aligned using the Align Multiple Sequences tool in Clone Manager Professional Suite, selecting the Multi-Way alignment type with the BLOSUM 62 scoring matrix.

2.4.3. **Generation of WebLogo of lipoprotein signal peptidases**

The initial amino acid sequence alignment generated using 25 proteins as described in 2.4.2 was entered into the WebLogo server (199) which generated the output image.

2.4.4. **Phylogenetic tree of lipoprotein signal peptidases**

A phylogenetic tree of bacterial signal peptidases was generated by selecting the Picture output option for the *in silico* alignment produced as described in 2.4.2. The result was exported to Microsoft PowerPoint where it was modified to highlight proteins of interest.

2.4.5. **LspA2 BLASTp search and sequence alignment**

To identify proteins with similar sequences to *C. difficile* LspA, its amino acid sequence was entered into the BLASTp internet database search tool, choosing to omit any matches from *C. difficile*. The results were analysed and the protein that displayed the greatest sequence identity with the highest sequence coverage was selected. The amino acid sequence of this protein was imported into Clone Manager Professional Suite and the two proteins were aligned using the Compare Two Sequences tool.

2.4.6. **PSIPRED server predictions**

Predictions of the secondary structure and membrane helix topology of *C. difficile* LspA and LspA2 were performed by entering the amino acid sequences of each protein into the PSIPRED protein sequence analysis workbench (200), selecting PSIPRED v3.3 and MEMSAT3 & MEMSAT-SVM prediction methods. The outputs were exported as image files.

2.5. **Characterisation of growth of *Clostridium difficile* strains**

A scrape from a streak plate for each particular strain was inoculated into 10 ml sterile, pre-reduced BHI broth supplemented with antibiotics where appropriate and grown overnight as a start culture. 100 ml sterile pre-reduced BHI broth was then inoculated with 1 ml of overnight culture. At 1 hour intervals OD₅₉₅ of the culture was measured in triplicate in a 1 cm path length plastic cuvette. Where required, colony forming units per millilitre (CFU ml⁻¹) of the culture was determined by the technique first described by Miles and Misra (201). Briefly, a serial dilution of culture in PBS was performed and plated in triplicate 20 µl drops on agar supplemented with appropriate antibiotics; numbers of countable colonies were to calculate the CFU ml⁻¹. Values were plotted graphically and analysed for statistical significance as described in 2.9.1.

2.6. Bacterial protein extractions

2.6.1. Crude protein extractions

Crude protein extractions of *E. coli* strains were prepared by mixing culture samples 4:1 in a screw-capped Eppendorf with SDS Sample Loading Buffer (41.5% v/v dH₂O; 12.5% v/v 0.5 M Tris, pH6.8; 10% v/v Glycerol; 2% v/v SDS; 5% 2-βmercaptoethanol; 0.01% bromophenol blue). Samples were then vortexed vigorously and boiled for 10 minutes. Finally, samples were pulsed in a bench-top microcentrifuge to settle cellular debris and were either used directly in SDS-PAGE or frozen for later analysis.

2.6.2. Preparation of whole cell lysate and culture filtrate

Whole cell lysates and culture filtrate samples from *C. difficile* strains were prepared as described previously (162). Briefly, the OD_{595 nm} of 50 ml overnight cultures was recorded before harvesting of cells at 5,000 g for 10 minutes at 4°C. The supernatant was retained as the culture filtrate sample. The cell pellet was then frozen at -20°C. Cell pellets were then thawed on ice and resuspended in phosphate-buffered saline (PBS) to OD_{595 nm} = 20 based upon the OD of the initial culture followed by incubation at 37°C for one hour. Samples were either used directly in SDS-PAGE or frozen to allow later analysis.

2.6.3. Low pH glycine extraction

To prepare surface layer protein extractions of *C. difficile* and *E. coli* strains 25 ml overnight cultures were treated as in Wright *et al* (202). Briefly, 25 ml of culture was harvested at 2,700 g for 15 minutes at 4°C then washed in 2.5 ml PBS followed by resuspension in 200 µl 0.2 M glycine at pH 2.2. Samples were then incubated at 37°C with shaking for 30 minutes before centrifugation at 12,000 g for 15 minutes at 4°C. Supernatant was retained and neutralised with 3.5 µl 0.25 M NaOH and either used directly in SDS-PAGE or frozen for later analysis.

2.6.4. Triton X-114 Extraction

Lipoprotein-enriched Triton X-114 extractions of *C. difficile* strains were prepared as described previously (203). Briefly, whole cell lysate samples were treated with pre-condensed Triton X-114 prior to freezing. Briefly, ice cold precondensed Triton X-114 was added to samples to a final concentration of

20% v/v and vortexed vigorously. This solution was then incubated on ice for 45 minutes with frequent vortexing followed by centrifugation at 16,000 g for 15 minutes at 4°C. The supernatant was transferred to a fresh Eppendorf tube and incubated at 37°C in a water bath for 10 minutes. This was then layered over an equal volume of warm sucrose cushion (6% sucrose w/v, 0.06% Triton X-114, pH 7.4) and incubated at 37°C in a water bath for 10 minutes. The sample was then centrifuged at 500 g for 3 minutes causing the sample to fractionate into three parts; the upper aqueous layer was removed and retained and the middle sucrose cushion layer was discarded. 9 X sample volume of ice-cold acetone was then added to the detergent-phase sample followed by incubation overnight at -20°C. Proteins were then pelleted at 16,000 g for 15 minutes at 4°C followed by washing with 80% v/v acetone and air drying before resuspension in 100 µl PBS. Samples were either used directly in SDS-PAGE or frozen for later analysis.

2.7. SDS-PAGE

To analyse the molecular weights of proteins in a sample aliquots were subjected to sodium dodecylsulfate – polyacrylamide gel electrophoresis (SDS-PAGE). Unless crude protein extractions were being examined, protein extractions samples were mixed 4:1 in a screw-capped Eppendorf with NuPAGE 4X SDS Sample Buffer (Invitrogen), boiled for 10 minutes in a heating block and debris settled by pulsing in a bench-top microcentrifuge; crude protein extract samples were used directly. Initially, 15 µl of prepared samples were loaded into NuPAGE Bis-Tris 4 – 12% Pre-Cast gels (Life Technologies) alongside 5 µl of Perfect Protein Marker (Novagen) and gels were run in NuPAGE MES SDS Running Buffer (Life Technologies) at 150V for 40 minutes. To visualise protein bands, gels were stained with SimplyBlue SafeStain (Life Technologies) using the manufacturer’s microwave protocol followed by imaging using a Li-Cor Odyssey CLx platform and ImageStudio software. Where a sample did not produce a quality trace with good band separation and limited smearing, dilutions and alterations in volume were loaded.

2.8. Hydrogen peroxide survival assays

To assess the survival of *C. difficile* strains in hydrogen peroxide (H₂O₂), 100 ml overnight cultures were split according to the number of H₂O₂ concentrations being tested and individually diluted to give 20 ml of culture at OD_{595 nm} = 0.5.

These cultures were then split into three 5 ml cultures, one per technical replicate. Colony forming units (CFU) of these untreated cultures were enumerated via the Miles and Misra technique (201) as described in Section 2.5. H₂O₂ was then added sequentially to each culture to achieve the desired final concentration, allowing processing time between each aliquot to enable consistent timing. After 30 or 45 minutes incubation with H₂O₂, CFU for each culture were enumerated. CFU counts for cultures before addition of H₂O₂ (CFU_{Initial}) and after the implemented incubation time (CFU_{Final}) were used to calculate percent survival using the formula $\frac{CFU_{Final}}{CFU_{Initial}} \times 100$. Results were combined and analysed for statistical significance using GraphPad Prism 6 software as described in 2.9.2.

2.9. Statistical analysis using GraphPad Prism 6

2.9.1. Growth of strains of *C. difficile*

To allow analysis of statistical difference between the growth of the tested strains of *C. difficile* it was first necessary to determine the most accurate model of curve fitting for the data points of each strain. This was done by comparison of the two most suitable models: the Exponential growth equation and the Sigmoidal dose-response (variable slope), anticipating the Exponential growth equation to be the best suited. Analysis revealed that the Sigmoidal dose-response (variable slope) model was preferred therefore this model was applied to the data points. Next, the suitability of one curve for all data sets was assessed; if this was the case then there would be no difference between the data sets. This was done by posing the null hypothesis that one curve would fit all data sets and testing the fit using the Extra sum-of-squares F test which showed that a different curve for each data set was the preferred model. Finally, the statistical difference between all data points for a given time point was calculated using a two-way ANOVA test with multiple comparisons between the data sets of each strain using the Tukey multiple comparisons test to calculate confidence intervals and significance.

2.9.2. Effect of Hydrogen Peroxide on strains of *C. difficile*

To analyse the statistical differences between the percent survival of strains of *C. difficile* when subjected to different concentrations of hydrogen peroxide, two-way ANOVA analysis was conducted with multiple comparisons between

each strain using the Tukey test to calculate confidence intervals and significance.

To analyse the statistical differences between the percent survival of strains of *C. difficile* at different incubation times with hydrogen peroxide, two-way ANOVA analysis was conducted with multiple comparisons between the mean percent survival of a strain at both time periods using the Sidak test to calculate confidence intervals and significance.

2.10. *C. difficile* CD2597 and CD1903 cloning

2.10.1. Cloning Methodology

The cloning methodology for generating plasmids encoding *C. difficile* *lspA* and *lspA2* for expression in *E. coli* is shown in Figure 5.3. Initially an expression plasmid would be selected based upon its suitability for use in *E. coli*, allowing design of PCR primers to amplify each gene, with desired alterations to flanking regions for the inclusion of correct restriction sites and modifications of the gene. Once amplified, PCR products were ligated into the TA-cloning vector pGEM-T Easy (Promega) according to manufacturer's instructions and transformed into *E. coli* TOP10, whereby transformants were selected for by blue-white selection and colony PCR confirmation. The intended protein-encoding inserts were produced by plasmid DNA extraction from *E. coli* holding pGEM-T Easy plasmids followed by restriction digest, size selection using agarose gel electrophoresis and gel extraction. Expression vectors were digested and processed concurrently, then used in ligations which were then transformed into *E. coli* TOP10 or HB101. Once the DNA sequence of each plasmid had been checked for accuracy, plasmids were transformed into expression strains of *E. coli*, ready for use in protein expression experiments.

2.10.2. Generation of plasmids

2.10.2.1. pRL4 and pRL6

Plasmids pRL4 and pRL6 (*lspA* and *lspA2* in pQE80) had been previously generated in Dr. S. L. Michell's laboratory at the University of Exeter by R. Lawrence (unpublished).

2.10.2.2. pEF10, pEF11, pEF12 and pEF13

Plasmids pEF10, pEF11, pEF12 and pEF13 were generated by PCR amplification of *C. difficile* genes CD2597 (pEF10 and pEF11) and CD1903 (pEF12 and pEF13) using the primers described in Table 4 which introduced *Nco*1 restriction sites flanking each gene, followed by conduction of the cloning methodology described above using pBAT4 as the expression vector. The start codons of each gene were modified during PCR amplification of insert fragments from TTG to ATG using either a 10 base insertion (pEF10 and pEF12) or force cloning base changes (pEF11 and pEF13) as detailed in APPENDIX6. Each insert also had a C-terminus His-tag (GTG GTG GTG GTG GTG GTG) added during PCR amplification. Post-digestion and pre-ligation, the pBAT4 expression vector was treated with alkaline phosphatase (New England Biolabs).

2.10.2.3. pEF14 and pEF15

Plasmids pEF14 and pEF15 were generated by PCR amplification of *C. difficile* genes CD2597 and CD1903 respectively using the primers described in Table 4, which introduced an upstream *Nco*1 restriction site and a downstream *Xho*1 restriction site, followed by conduction of the cloning methodology described above using pET28a as the expression vector. The start codons of each gene were modified by overwriting base changes, mediated by the forward primers, during PCR amplification from TTG to ATG GGC.

2.10.3. Bacterial DNA extractions

2.10.3.1. Preparation of plasmid DNA via minipreps

Plasmid DNA from strains of *E. coli* was extracted using either the QIAprep Spin Miniprep kit (Qiagen) or the PureYield Plasmid Midiprep System (Promega) depending on the desired output volume and concentration according to the manufacturers' instructions.

2.10.3.2. Preparation of DNA boilates for colony PCR

To allow rapid assessment of plasmid transformation success, antibiotic-resistant colonies were picked and suspended in 100 µl sterile nuclease-free deionised water (Life Technologies), followed by streaking as individual colonies to selective grid plates. Suspensions were then boiled in a heating block followed by pelleting of cellular debris in a bench-top microcentrifuge.

Supernatant was then used as template DNA in PCR reactions designed to amplify fragments specific to the transformed plasmid.

2.10.4. Polymerase chain reaction conditions

Table 5 shows the PCR cycle conditions used in this study.

Table 5 Polymerase chain reaction cycle conditions

Reaction Step	Temperature	Time	Cycles
Enzyme activation	95 °C	10'	
Denaturation	94 °C	1'	35
Primer Annealing	56 °C	1' 30"	
Extension	72 °C	4'	
Final Extension	72 °C	10'	
Hold	4 °C	∞	

All PCRs were performed using HotStarTaq DNA Polymerase (Quiagen) with associated 10X PCR Buffer and PCR Grade dNTP Mix. Each reaction used constituents in the proportions shown in Table 6, prepared as a master mix of all except DNA template in a quantity suitable for 1.1 times the number of reactions to be performed.

Table 6 Polymerase chain reaction constituents

Constituent	Volume (µl) for 10 µl total
Forward primer (20 mM)	1
Reverse primer (20 mM)	1
10X PCR Buffer	1
dNTPs (10 mM)	0.5
HotStarTaq	0.1
Nuclease-free water	3.4
DNA template / nuclease-free water	3

2.10.5. **DNA digestion by restriction endonucleases**

Digestion of DNA samples was performed using restriction endonucleases supplied by New England Biosciences according to the manufacturer's specifications. Briefly, 10 μ l of DNA sample was mixed with 1 μ l of each enzyme selected, 1.5 μ l appropriate enzyme buffer and made up to 15 μ l with nuclease-free water. Reactions were mixed thoroughly then incubated at 37 °C before inactivation of the enzymes at manufacturer-advised temperatures. For preparatory digests to produce large quantities of DNA fragments for cloning the reaction volume was increased to 80 μ l by mixing 42.5 μ l DNA sample, 4 μ l of each selected enzyme, 8 μ l of appropriate enzyme buffer and made up to the total volume with nuclease free water. These reactions were incubated at 37 °C for at least one hour to allow complete digestion followed by heat-inactivation.

2.10.6. **Agarose gel electrophoresis**

To allow size selection of DNA samples for analysis of either PCR products or restriction digest products, samples were mixed 1:5 with 6X DNA Loading Dye (Thermo Scientific) and loaded into 1 % w/v Agarose (Melford) in Tris-acetate-EDTA buffer. Gel electrophoresis was typically performed at 120 V until loading dye bands had reached past halfway across the gel or at 90 V for size selection of preparatory DNA digests. DNA markers were included for fragment size comparison: 1 kb Plus DNA ladder and 100 bp plus DNA ladder (Thermo Scientific)

2.10.7. **Ligation of DNA fragments**

Ligations of DNA fragments except those performed with pGEM-T Easy were performed with T4 DNA Ligase (New England Biolabs) according to the manufacturer's specifications. Briefly, reaction mixes including controls were prepared according to Table 7 and incubated overnight at 16°C in a heating block. Volumes of cut vector (x) and cut insert (y) were calculated according to the formulae displayed in the table, assuming an ideal ratio of 1:3 and reaction concentrations of 50 fmol vector and 150 fmol insert. DNA concentrations in samples were determined by analysis using a Nanodrop 1000 (Thermo Scientific).

Table 7 Ligation reaction mixes and volume calculations

Reaction	Cut Vector (µl)	Cut Insert (µl)	10X T4 DNA ligase buffer (µl)	T4 DNA ligase (µl)	Nuclease-free water (µl)	Total volume (µl)
Control	x	0	3	1	To Total	30
Ligation	x	y	3	1	To Total	30

$$\text{ng of Cut Vector required} = \text{length (bp)} \times 0.033$$

$$\text{ng of Cut Insert required} = \text{length (bp)} \times 0.099$$

$$\text{Volume required (}\mu\text{l)} = \frac{\text{ng required}}{\text{concentration } \left(\frac{\text{ng}}{\mu\text{l}}\right)}$$

2.10.8. Transformation of plasmids into bacteria

2.10.8.1. Heat Shock

Plasmid DNA was transformed into chemically competent *E. coli* by heat shock. Briefly, one 200 µl aliquot of competent cells per sample was thawed on ice. 10 µl of DNA sample was then added to the cells and gently mixed before incubation on ice for 30 minutes. Samples were then heat shocked at 42°C for 30 – 60 seconds before further incubation on ice for 10 minutes. 1 ml pre-warmed LB broth was then added to samples which were then incubated at 37°C for one hour with shaking. Samples were then plated to LB agar supplemented with appropriate antibiotics in multiple 100 µl aliquots before incubation overnight at 37°C.

2.10.8.2. Electroporation

Plasmid DNA was transformed into electro-competent *E. coli* by electroporation. Briefly, one 50 µl aliquot of competent cells per sample was thawed on ice while 6 – 10 µl DNA samples were dialysed on nitrocellulose membranes (Millipore) over 20 ml of deionised water for 30 minutes. DNA samples were then added to the competent cells and mixed gently. Samples were then transferred to the bottom of 0.2 cm electroporation cuvettes (Bio-Rad) followed by pulsing in a Bio-Rad MicroPulser Electroporator set to Bacteria Ec2. Cells were then

washed from the cuvette by five repeated flushings with 200 µl pre-warmed LB. Flushings were pooled for each sample and incubated in an Eppendorf tube at 37°C for one hour with shaking. Finally, samples were serially diluted to 10⁻⁵ and 100 µl aliquots spread plated to LB agar supplemented with appropriate antibiotics and incubated overnight at 37°C.

2.10.9. DNA sequencing and analysis

To determine the DNA sequence of a region of a plasmid, samples of Miniprep elutions were sent for Sanger Sequencing by Source BioScience. The DNA concentration of Miniprep elutions was measured by analysis using a Nanodrop 1000 (Thermo Scientific) and adjusted to 100 ng µl⁻¹ as per sample specifications. Sequencing runs were either performed with stock primers or specific primers were designed and provided at a concentration of 3.2 pmol µl⁻¹. DNA sequence files were analysed using Clone Manager Professional Suite.

2.10.10. Rare Codon Caltor analysis of *C. difficile* CD1903 and CD2597

To determine the rare codon usage within *C. difficile* genes CD2597 and CD1903 the DNA sequences of each gene was entered into the Rare Codon Caltor (204) internet form for analysis.

2.11. IPTG induction of protein expression

Expression of proteins cloned into expression vectors containing T7/*lac* promoters in suitable *E. coli* strains was initially conducted using Isopropyl β-D1-thiogalactopyranoside (IPTG). Overnight cultures of strains were used to provide 250 µl inoculums which were added to 10 ml pre-warmed LB broth containing appropriate antibiotics. These cultures were then incubated at 37 °C with shaking until the OD_{595 nm} reached between 0.6 and 1. 800 µl was removed and preserved as a non-induced sample. IPTG was added to the remainder to a final concentration of 1 mM. After 3 hours further incubation, 800 µl induced samples were taken and prepared for SDS-PAGE analysis.

2.12. Auto-induction of protein expression

Protein expression was auto-induced following the protocol set out in Studier, 2005 (205). The general compositions of the media used are given in Table 8 and the composition of specific solutions as follows: 50X “M” solution: 1.25 M Na₂HPO₄, 1.25 M KH₂PO₄, 2.5 M NH₄Cl, 0.25 M Na₂SO₄; 50X “5052” solution:

25 % (v/v) glycerol, 2.5 % (w/v) glucose, 10 % (w/v) α -lactose monohydrate; 100X "505" solution: as for "5052" solution with no α -lactose monohydrate. 1000X Trace Metals solution was prepared as follows: 20 mM CaCl_2 , 10 mM MnCl_2 , 10 mM ZnSO_4 , 2 mM CoSO_4 , 2 mM $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, 2 mM $\text{NiSO}_4 \cdot 6\text{H}_2\text{O}$, 2 mM $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 2 mM Na_2SeO_3 , 2 mM H_3BO_3 made up with water and autoclaved plus 50 mM $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ made up in 1 ml concentrated HCl and 100 ml water before filter-sterilization. Briefly, *E. coli* strains for the experiment were grown overnight at 37°C and 200 rpm shaking in MDG media supplemented with the appropriate antibiotics, typically 100 $\mu\text{g ml}^{-1}$ Ampicillin. 1 ml of overnight culture was then inoculated to 150 ml or 100 ml respectively of ZYM-5052 and ZYM-505 media. These cultures were then grown for 48 hours at 20°C and 300 rpm shaking. MDG media is a minimal media suitable for growing working stocks, ZYM-505 media is suitable for the growth of high-density cultures (non-induced samples) and ZYM-5052 media is suitable for the growth of high-density cultures and is auto-inducing due to the presence of α -lactose monohydrate in the "5052" solution (auto-induced samples).

2.13. Western Blotting

To demonstrate that His-tagged proteins were present in DNA samples that had been subjected to SDS-PAGE analysis, non-stained SDS-PAGE gels were blotted to nitrocellulose membranes in iBlot Transfer Stacks (Invitrogen) using the iBlot system (Invitrogen). Membranes were then stained by washing with Ponceau-S stain (0.1% w/v Ponceau-S in 5% v/v glacial acetic acid) to confirm transfer of proteins. The Ponceau-S stain was removed by washing with 0.1 M sodium hydroxide and deionised water. Membranes were then blocked in tris buffered saline with 1% Tween 20 (TBS-T) 5% w/v skim milk powder at 4°C overnight. Next, membranes were washed three times in TBS-T with room temperature incubation with shaking for 10 minutes between each wash. His-Tag monoclonal antibody HRP-conjugate (Novagen) was diluted 1:4,000 in TBS-T plus 3% w/v skim milk powder; this solution was made up to a 10 ml volume and poured over the membrane before incubation for one hour at room temperature. Membranes were then washed three times in TBS-T before developing using 200 μl each of Lumigo Reagent A and Peroxide Reagent B (Cell Signalling Technology) made up to 4 ml with deionised water then poured over the membrane before 5 minutes incubation. Membranes were then

wrapped in cling film and imaged by chemiluminescence illumination and sequential imaging using a Bio-Rad ChemiDoc XRS+ System and ImageLab software.

Table 8 Media composition for auto-induction of protein expression.

Media	Tryptone (g)	Yeast Extract (g)	De- mineralised Water (ml)	60X “M” Solution (ml)	100X “505” Solution (ml)	50X “5052” Solution (ml)	2 M MgSO₄ (ml)	1000X Trace Metals Solution (μl)	40% (w/v) Glucose (ml)	5% (w/v) L- Aspartate Solution (ml)
MDG			920	20			1	200	12.5	50
ZYM-505	10	5	970	20	10		1	200		
ZYM-5052	10	5	960	20		20	1	200		

Chapter 3. Bioinformatic analysis of lipoprotein signal peptidases in bacteria

3.1. Introduction

To date, few species of bacteria have been shown to contain two versions of a lipoprotein signal peptidase gene; while both CD2597 and CD1907 may be putative lipoprotein signal peptidases in *C. difficile*. Both genes are shown to be functional later herein and in previous work using strains of *C. difficile* 630 mutated for each gene, with very similar behaviour which may indicate a similar function. Much can be elucidated about the function of a protein from its primary, secondary and tertiary structures, from the sequence location of functional amino acids to the presence of structural domains. There is considerable evidence of a conserved tertiary structure and membrane topology within lipoprotein signal peptidases in conjunction with conserved domains containing functional amino acids (29, 30, 177, 178); these properties may be used to indicate the functionality of a putative lipoprotein signal peptidase. This chapter aims to show the properties of *C. difficile* 630 LspA (CD2597) and LspA2 (CD1903) elucidated through a selection of bioinformatics analyses. These analyses include amino acid sequence comparison to known lipoprotein signal peptidases to find conserved domains, the similarity between LspA and LspA2 and of LspA2 to other lipoprotein signal peptidases.

3.2. Multiple alignment of annotated lipoprotein signal peptidase genes

The amino acid sequences of twenty five annotated lipoprotein signal peptidase proteins, including both LspA and LspA2 from *C. difficile* 630 were aligned using the Align Multiple Sequences tool in Clone Manager Professional Suite. These genes were selected from those used in similar alignments in Rahman *et al* (177) and Paetzel *et al* (29) where the same gene sequence was available in online repositories with the addition of LspAs from some *Clostridia* and the four LspAs from *Myxococcus xanthus*. The initial alignment with protein accession numbers is shown at APPENDIX 1 and includes indication of the location of the 5 conserved amino acid sequence homology domains A through E with regions of homology highlighted in green. This alignment shows that all of the genes selected contain all of the expected domains. There are regions of significant

variation outside of these domains. A second alignment was therefore performed which included as many Gram positive lipoprotein signal peptidase proteins from Paetzel *et al* (29) and Rahman *et al* (177) as possible. This alignment was entered into the WebLogo 3.3 tool (199) to produce Figure 3.1, in which the conserved domains have again been identified. Here, the domains are made much clearer by the greater frequency of the expected amino acids in each domain, as represented by the larger letter at the given position. Both of these views of the alignments support the evidence that lipoprotein signal peptidases contain these 5 conserved amino acid sequence domains and they show that both LspA and LspA2 in *C. difficile* 630 contain them. The highly represented amino acids in the sequence alignment (APPENDIX 1) and the WebLogo (Figure 3.1) were then used to produce consensus sequences for each conserved amino acid domain (Table 9), allowing comparison to the sequences shown in Table 1.

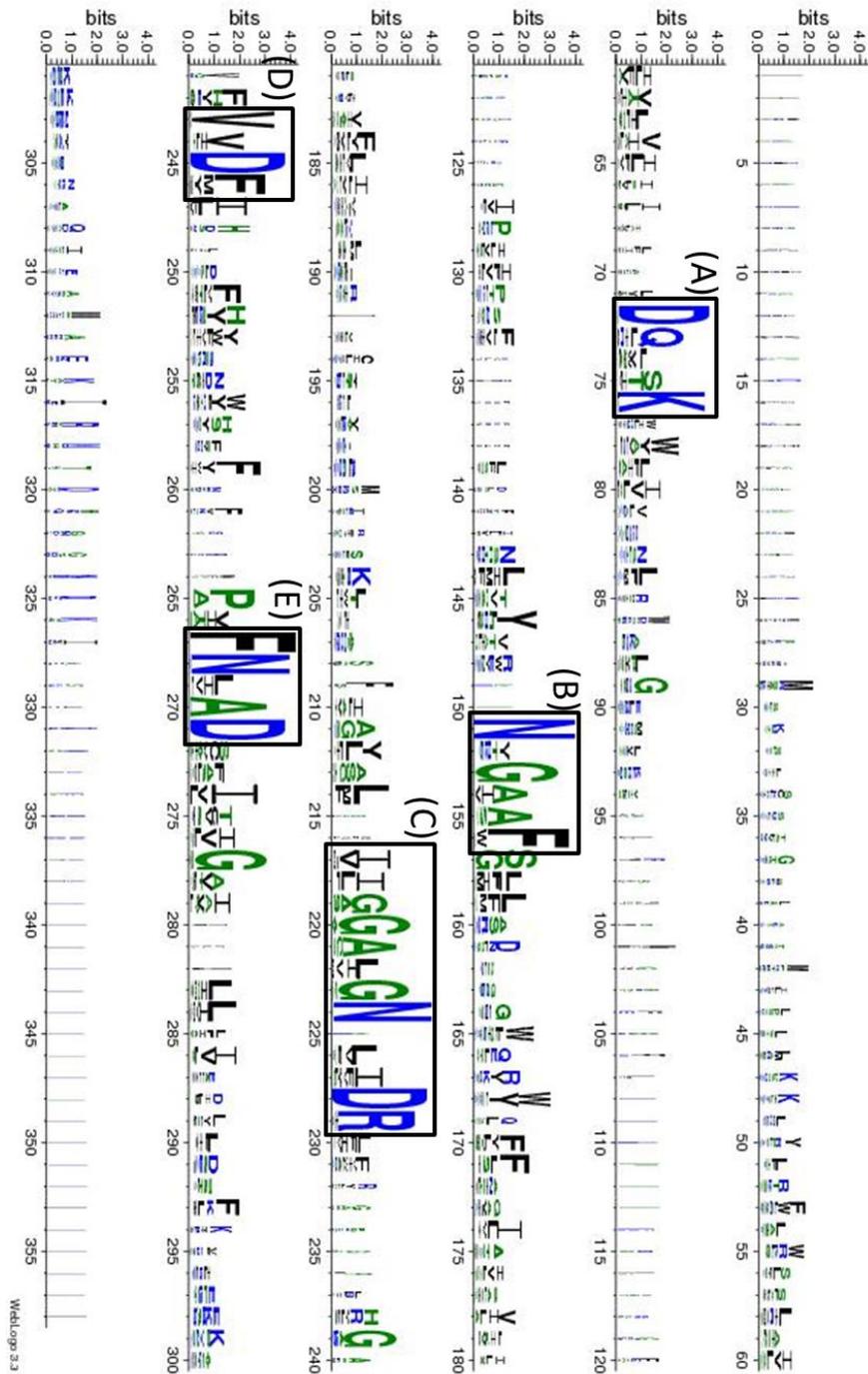


Figure 3.1 WebLogo of aligned lipoprotein signal peptidase from various bacteria, highlighting the conserved functional amino acids

The WebLogo 3.3 tool (199) was used to generate a representative image of the amino acid sequences of twenty five lipoprotein signal peptidase genes from a variety of bacteria. Relative amino acid frequency at each locus is represented by letter height. The conserved amino acid sequence domains reported in Rahman *et al* (177) and Paetzel *et al* (29) are clearly demonstrated as regions with several amino acids at high relative frequencies and are identified in boxes, using the nomenclature A through E.

Table 9 Comparison of the expected amino acid sequences of lipoprotein signal peptidase conserved domains given in the literature (Figure 1.4 and Table 1) to the sequences generated in this study showing considerable similarities

The conserved domain sequences for lipoprotein signal peptidases shown in Table 1 are presented alongside those derived in this study. Comparison of the domain sequences shows considerable similarity, with most differences arising in amino acid relative frequencies. The relative frequency of each amino acid in consensus sequences is denoted by the type of lettering: uppercase indicates a very high frequency; lowercase indicates a low frequency and 'x' indicates no evidence of conservation at that residue. Amino acid frequency or level of conservation was not analysed by Rahman *et al* (177).

Domain	Tjalsma <i>et al</i> (30)	Paetzel <i>et al</i> (29)	Rahman <i>et al</i> (177)	This study
A	dqxxk	dxxtk	DQLSK	DqxxK
B	NxGaaf	NxGaaf	NYGISF	NxGaaF
C	ixggalgNxxDr	iiggaxlgNxxDr	VIGGAVGNLIDR	iigGalGNxxx DR
D	VvD	vvd	VFD	VvDf
E	FNxAD	FNxAD	FNLAD	FNxAD

3.3. Phylogenetic tree of lipoprotein signal peptidases

The Align Multiple Sequences tool was then used, with the second alignment produced in Section 3.1, to produce a phylogenetic tree, shown in Figure 3.2; the protein accession numbers are shown in APPENDIX 2. Lipoprotein signal peptidase genes from either the same species or species from the same genus cluster together, indicating that they are more similar to each other than to the genes from the other bacteria. This suggests that the regions of variation outside of the domains and the variation within them may be associated with the source species or genus. Of particular note however, is that the *C. difficile* 630

LspA2 sequence does not group to the same sub-clade or even clade as the sequences of *C. difficile* 630 LspA or other *Clostridia*. In fact, it does not appear to group with any of the other genes in the alignment. This indicates that this gene may have origins in a species of bacteria whose lipoprotein signal peptidase proteins have not been investigated here.

3.4. Potential sources of LspA2

Because the amino acid sequence of LspA2 shares very little homology to any of the lipoprotein signal peptidases used in the alignment and phylogenetic tree, this sequence was entered into the BLASTp internet-based tool and the database of protein sequences was searched excluding proteins from *C. difficile*. The results demonstrate that the two most similar protein sequences in the database belong to unclassified *Lachnospiraceae* bacterium, which also belong to the order *Clostridiales*. Both of these proteins are annotated as peptidase A8 with either 73% identity at 80% coverage or 96% at 61% coverage. An amino acid sequence alignment of the result (80% coverage) to LspA2 reveals many amino acid differences and similar conserved domain sequences, except domain A which does not appear in the peptidase A8 sequence. The top 10 BLASTp results and the amino acid sequence alignment of LspA2 and peptidase A8 are shown in Figure 3.3.



Figure 3.2 Phylogenetic tree of lipoprotein signal peptidase genes from various bacteria

An alignment of the amino acid sequences of twenty five lipoprotein signal peptidases from a variety of bacteria was generated and processed in Clone Manager Professional Suite to produce a phylogenetic tree of these proteins. Proteins from a single bacteria or from bacteria in the same genus group together into apparent clades, as demonstrated by the highlighted groups; *Myxococcus xanthus* (orange), *Clostridium* spp. (blue plus *C. difficile* 630 LspA in red) and *Rickettsia* spp. (green) however *C. difficile* 630 LspA2 (red, top) does not group into any clade and notable does not group with *C. difficile* 630 LspA.

```

LspA2          1 qggvnirqvksfvfpvislifldqiskvliglflmdfeidiigkflrfnpvqntnlsygg
peptidase A8   1 -----mldldfdiigkllrfspvqnlinsygg

LspA2          61 nfigilsnlwvlfnilvilviisgyafykskneqtsysvkvimscglacticslidkl
peptidase A8   27 nfigilsnfwvmlfnilvillvsgynfyktkrqhtsysvkviltcglsctlcslldkv

LspA2          121 fwggsldflqipsffifdlkdvcyltvaeeifvvigilhnreismkeyiyfcyrqfkr
peptidase A8   87 lwggsldflqiptvftfdlkdilyltiaeeifvligvfhskeisvkeylsfcyrllkr

```

Figure 3.3 Amino acid sequence alignment of *C. difficile* 630 LspA2 (CD1903) to peptidase A8 from *Lachnospiraceae* bacterium

Amino acid sequence alignment of the peptidase A8 from *Lachnospiraceae* bacterium, the top hit from a BLASTp search for *C. difficile* 630 LspA2 (CD1903) to LspA2 with non-matching amino acids highlighted in orange.

3.5. Analysis of the predicted structures of LspA and LspA2 in *C. difficile*

As discussed in Section 1.7 and shown in Figure 1.4, there is a particular structure associated with lipoprotein signal peptidases which includes 4 transmembrane domains and a particular localisation of the 5 conserved domains. To investigate whether this was the case for LspA and LspA2 in *C. difficile* 630, the amino acid sequences of each gene were entered into the PSIPRED server (200) using the secondary structure and membrane topology (MEMSAT3) prediction methods as described in Materials and Methods, with the results displayed in Figure 3.4. These predictions demonstrate with high confidence that both proteins possess the expected 4 transmembrane domains, composed of helices. Further, they show that the 5 conserved sequence domains, the positions of which have been overlaid on the figure based upon the annotated amino acid positions, are located very similarly to the model structure presented in Figure 1.4 and Paetzel *et al* (29). There are noticeable differences in the predicted secondary structures of the two proteins however, with different lengths of protein between the transmembrane helices and different predicted helix lengths.

3.6. Discussion

The two lipoprotein signal peptidases encoded in the *C. difficile* 630 genome share a number of common features. Both proteins contain the 5 expected amino acid domains that have been shown to be common to many lipoprotein signal peptidases from other species of bacteria (29, 30, 177), indicating their function and functionality. These conserved domains do differ in sequence between the two strains however the key amino acids that identify each domain are mostly present in both proteins with greater variation in LspA2 compared to the expected sequence. This variation does not extend to the two aspartic acid residues that are thought to form the active site of the proteins in domains C and E. Variation within the domains and in the regions between them is common across lipoprotein signal peptidases from other species of bacteria and such variations may allow grouping of bacteria. The amino acid sequences of lipoprotein signal peptidases from the same bacteria or from bacteria within the same genus appear to group together in clades however, LspA2 does not group with LspA or with proteins from other *Clostridia*, suggesting that LspA2 has a different origin to LspA. To briefly investigate this origin, a BLASTp search for the amino acid sequence of LspA2 was conducted, revealing that the most similar protein in the database is peptidase A8 from an uncharacterised *Lachnospiraceae* bacterium. This genus of bacteria belongs to the same taxonomic order (*Clostridiales*) as *C. difficile*, and it may be possible that *C. difficile* picked up the *lspA2* gene from this genus. The presence of multiple differences between the two amino acid sequences indicates that any genetic transfer between the two host species occurred a long time ago and no further conclusions can be drawn. As such, more work could be conducted to examine the origin of LspA2. Analysis of the predicted secondary and tertiary structures of LspA and LspA2 using the PSIPRED (200) server demonstrates that both proteins possess the expected four transmembrane helices and general configuration. Further, the 5 conserved domains are present at locations that are very similar to those in the model type II signal peptidase displayed in Figure 1.4, suggesting that both proteins could be functional in the expected fashion. Both LspA and LspA2 possess the expected features of lipoprotein signal peptidases and LspA2 may have an origin in other bacteria.

Chapter 4. Phenotypic characterisation of *C. difficile* strains defective in lipoprotein biosynthesis and analysis of their response to antimicrobial stresses

4.1. Introduction

C. difficile strains have been generated that are mutants in CD2597 (*lspA*) and CD1903 (*lspA2*) (34-36), named ECF1 and ECF2 respectively, to allow investigation of the roles that the two functional versions of lipoprotein signal peptidase encoded by these genes play in *C. difficile*. The strains were shown to have very similar growth characteristics to the 630 Δ *erm* wild type however ECF1 was shown to have a different protein profile to both the wild type and ECF2, with other variations seen in ECF2. The ability of each strain to withstand oxidative stress was examined, and it was shown that ECF1 was significantly less susceptible than the wild type and ECF2 behaving similarly but less significantly. Further work will be required in this area to demonstrate why disruption of lipoprotein biosynthesis appears to be beneficial to the mutant strains and to further investigate the effect of incubation time on survival.

4.2. *C. difficile* strains defective in lipoprotein biosynthesis

In previous work, two *C. difficile* strains were generated that were defective in lipoprotein biosynthesis by using the Clostron system to knock out the *lspA* and *lspA2* genes, creating strains ECF1 and ECF2 respectively, with each strain subsequently shown to differ from the wild type in tolerance to malachite green (34, 35). The plasmids used to generate these knockout strains are shown in APPENDIX 3. Confirmation of the presence of mutations in these strains was performed by PCR on genomic DNA extractions. These PCRs used primers which flanked the intron insertion site or which bind within the intron and are shown in Figure 4.1 and confirmed that both strains were correct mutants.

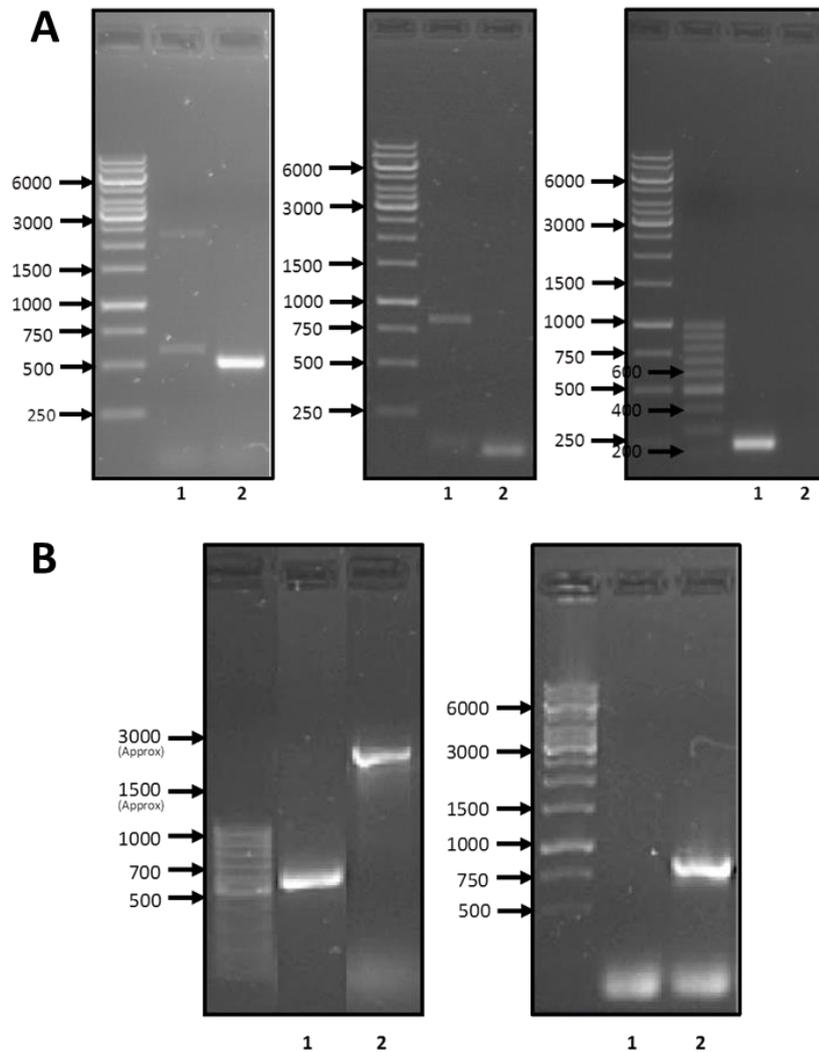


Figure 4.1 PCR confirmation of ClosTron mutants in *IspA* and *IspA2* of *C. difficile*

PCRs to confirm the generation of ClosTron mutants in *C. difficile* 630 of *IspA* (CD2597) and *IspA2* (CD1903). **A:** CD1903, left gel: insertion site flanking primers, middle gel: intron Ram primer pair, right gel: forward flanking primer with intron EBS universal primer. All gels lane 1: *C. difficile* ECF2 (CD1903 mutant), lane 2: *C. difficile* 630 Δ *erm*. **B:** CD2597, left gel: flanking primers, right gel: intron Ram primer pair. All gels lane 1: *C. difficile* ECF1 (CD2597 mutant), lane 2: *C. difficile* 630 Δ *erm*. The difference in band sizes with flanking primers for both genes indicates the presence of the intron within the gene, as does the presence of a band with the intron Ram primers. PCR with the forward flanking primer and the intron EBS universal primer was necessary for the CD1903 mutant because of the double band with the flanking primers, and the presence of a band indicates the presence of the intron in the insertion site. Figure adapted from Farries, 2012 (34).

4.3. Analysis of the growth of *C. difficile* strains

To investigate if the mutations in *C. difficile* strains ECF1 and ECF2 had a pleiotropic effect on growth of the strains, their growth was assayed alongside the *C. difficile* 630 Δ *erm* wild type strain by measuring the optical density of cultures at one hour intervals. The results of triplicate growth curves were combined and plotted using GraphPad Prism 6, which also allowed statistical analysis of the curves; the results are shown in Figure 4.2 and the statistical analysis is shown in APPENDIX 4. It was observed that, at 5h and 6h, 630 Δ *erm* had a significantly different average optical density to ECF1 ($P=0.0251$, $P=0.0223$) but not ECF2 ($P=0.2581$, $P=0.1212$). This is reflected in the graph with the fitted curve for 630 Δ *erm* appearing higher than the other two curves, particularly at these two time points. Despite this difference, all three strains appear to enter exponential growth after a very similar lag phase, followed by growth at a very similar rate, entering stationary growth after a very similar time period.

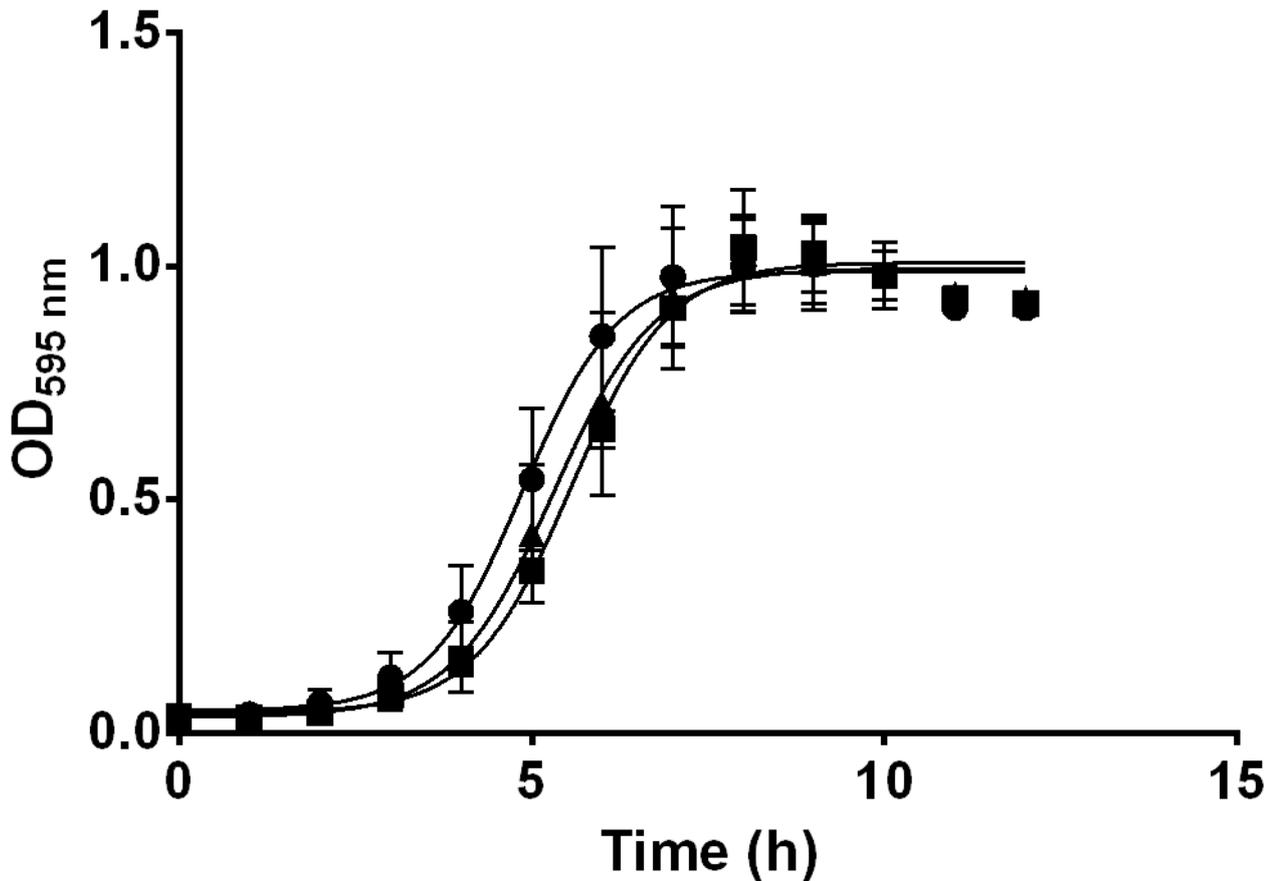


Figure 4.2 Growth curves of wild type and *lspA* mutant strains of *C. difficile*

Overnight cultures of *C. difficile* strains 630 Δ *erm* (●), ECF1 (■) and ECF2 (▲) were used to inoculate fresh media which was then monitored for growth by measuring the optical density at 595 nm every hour up to 12 hours. The experiment was performed in triplicate and the results collated in GraphPad Prism 6 which allowed statistical analysis. This analysis allowed plotting of best fit curves for each strain using a Sigmoidal dose-response model and it was found that one curve would not be suitable for all three data sets. Further analysis revealed that the OD_{595 nm} for 630 Δ *erm* at 5 h and 6 h was significantly different to that of ECF1 however the remainder of the data points were statistically similar across all three strains. Error bars represent standard deviation of the triplicate results.

4.4. Protein profiles of *C. difficile* strains

Using the extraction protocols defined in Materials and Methods, the protein profiles of *C. difficile* strains 630 Δ *erm*, ECF1, ECF2 and 630 Δ *erm* Δ *lgt* were investigated to examine whether there are any observable differences in lipoprotein processing between the strains. Lipoproteins that have not undergone signal peptide cleavage by LspA or LspA2 may be observably larger, resulting in the shift of a protein band between strains. Initially, the whole cell lysate and culture filtrate for the first three strains were compared by SDS-PAGE of prepared samples, as shown in Figure 4.3.A, to look for evidence of additional proteins being released to the culture filtrate in the two mutant strains. It was hypothesised that the lipoprotein signal peptidase mutants would be less able to anchor proteins to the cell than the wild type due to the disruption to lipoprotein maturation, causing release of lipoproteins into the culture filtrate. In this case, no such differences are directly apparent; in particular comparing strains 630 Δ *erm* to ECF2 (Figure 4.3.A lanes 1 and 3) where all bands are present in both lanes. Strain ECF1 (Lane 2) may have some differences however all bands are present with some relative intensity changes. A protein that is present in one fraction but not the other will correspond to a band at a particular molecular weight that is only visible in one fraction. As such, where a band is visible for one strain but not the others in that fraction, a protein may be localised differently between those strains. Figure 4.3.A, Culture Filtrate Lanes 1 and 3 show a clear band at approximately 26 kDa which is not as strongly present in Lane 2. No such band is visible in Lane 2 for the Whole Cell Lysate samples, suggesting that the protein represented by the band is not found in either fraction for ECF1. This protein may be localised to another cellular fraction or it may be differentially expressed in this strain.

It was then considered that incorrectly processed lipoproteins may remain associated with the cell wall due to incomplete processing, instead of the culture filtrate in the mutant strains. Therefore, extractions that enrich for the outer layers of the cell were performed for strains 630 Δ *erm*, ECF1 and ECF2. The TX-114 extraction, which enriches for lipoproteins (203), was also performed for the 630 Δ *erm* Δ *lgt* strain. Samples were run through SDS-PAGE as shown in Figure 4.3.B. In the low pH glycine extraction, which removes proteins associated with the *C. difficile* cell wall (202), one band at approximately 24 kDa

in the ECF1 lane (Lane 2) is distinctly darker than the bands for the other strains at the same molecular weight. Further, a band at the same molecular weight does not appear in Lane 2 but does appear in Lanes 1 and 3 and other bands visible in Lane 2 of the low pH Glycine extraction at approximately 15 kDa are not visible in Lanes 1 and 3 for the same extractions but may be visible in Lanes 1 and 3 for the detergent phase of the TX-114 extraction. This suggests that the proteins in this band are being located differently in the corresponding strain, ECF1. There are other differences between the two phases of the TX-114 extraction with various bands appearing at different intensities between strains though nothing as distinct as that seen with the low pH glycine extraction for ECF1 when compared to each other and to 630 Δ *erm*. There are however, more intensity differences when the TX-114 extraction of the 630 Δ *erm* Δ *lgt* strain (Lane 4) is compared to similar extractions of the other strains, supporting previous observations that protein processing is considerably different in this strain compared to 630 Δ *erm*.

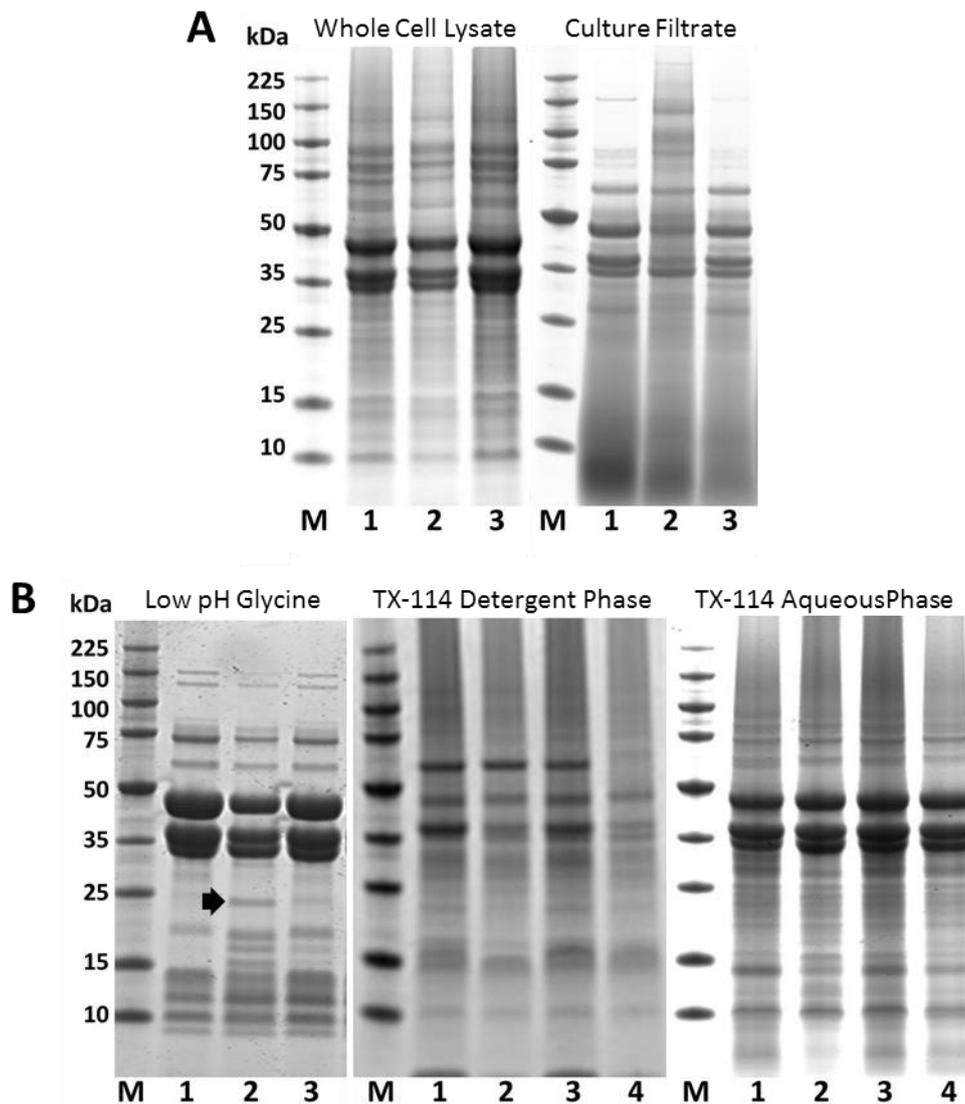


Figure 4.3 SDS-PAGE of protein extractions from *C. difficile* 630 Δ erm, ECF1, ECF2 and 630 Δ erm Δ lgt

To examine the possibility that incorrectly processed lipoproteins would locate to different parts of the cell, protein extractions were performed on *C. difficile* 630 Δ erm (Lane 1), ECF1 (Lane 2), ECF2 (Lane 3) and 630 Δ erm Δ lgt (Lane 4) where relevant. **A:** Initially, it was thought that incorrectly processed lipoproteins would be released from the cell, therefore the whole cell lysate and culture filtrate were analysed by SDS-PAGE to look for protein bands that were arranged differently in each strain; no major differences were observed. **B:** Protein extractions that correctly processed lipoproteins may locate to were then examined, including the lipoprotein-enriching TX-114 extraction. 630 Δ erm Δ lgt was also examined here. A band at approximately 24 kDa appears much brighter in a low pH glycine extraction of ECF1 (arrow) with other bands appearing at varying intensities in both extractions.

4.5. Response to hydrogen peroxide oxidative stress

4.5.1. Initial analysis

To investigate the effect of disruption of lipoprotein biosynthesis on the resistance of *C. difficile* to oxidative stress, broth cultures of 630 Δ *erm*, ECF1 and ECF2 were treated with hydrogen peroxide for 30 minutes at a variety of concentrations, ranging from 0 mM as a control to 1 mM H₂O₂. Colony forming units for each strain were enumerated before and after addition of hydrogen peroxide, allowing calculation of percent survival. As can be seen in Figure 4.4, there was a general reduction in % survival for all strains with increasing hydrogen peroxide concentration ([H₂O₂]), particularly with the higher concentrations used. It was observed that there was a higher % survival of both ECF1 and ECF2 compared to the 630 Δ *erm* wild type at 0.1 mM and no measurable growth for any strain at 1 mM. Statistical analysis was conducted using GraphPad Prism 6 using a two-way ANOVA followed by Tukey's multiple comparisons test to determine the significance of any differences in % survival between strains at each concentration. The details of this analysis can be found at APPENDIX 5, and the results are portrayed above the graph. There was statistically significant variation between ECF1 and ECF2 at 0.01 mM (P=0.0127) and between 630 Δ *erm* and ECF1 at 0.05 mM (P=0.0178). It was expected that the % survival of each strain would decrease as the hydrogen peroxide concentration increased; the 630 Δ *erm* strain does not appear to follow this trend between 0 and 0.05 mM H₂O₂, indicating that the significant difference between 630 Δ *erm* and ECF1 at 0.05 mM may be due to experimental variation. At 0.1 mM, both ECF1 and ECF2 were significantly different to 630 Δ *erm* (P=0.0012, P=0.0002), and both mutant strains appeared similarly better able to survive the imposed stress.

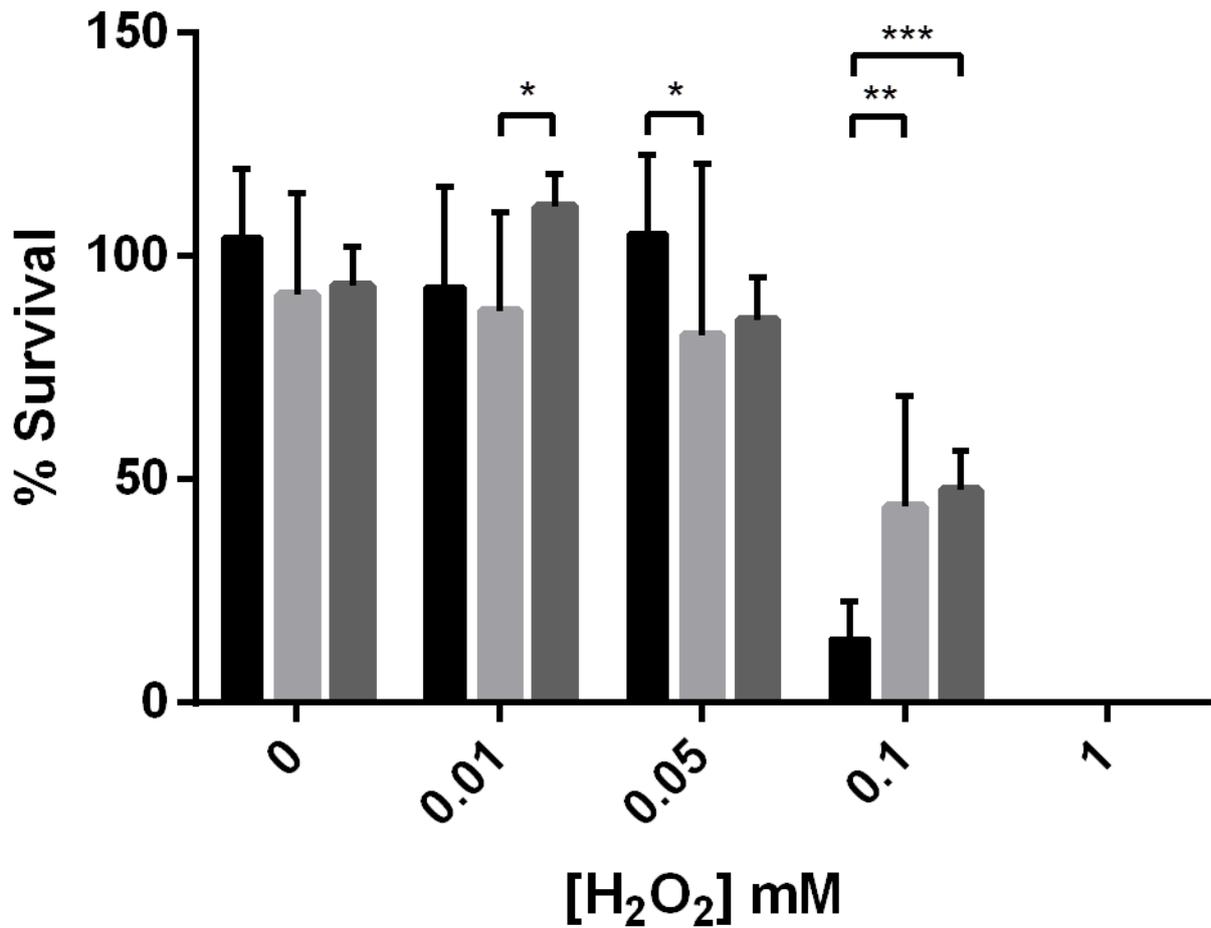


Figure 4.4 Effect of hydrogen peroxide concentration on survival of *C. difficile* strains

Broth cultures of *C. difficile* 630Δ*erm* (black bars), ECF1 (light grey bars) and ECF2 (dark grey bars) were treated with hydrogen peroxide (H₂O₂) at concentrations of 0, 0.01, 0.05, 0.1 and 1 mM for 30 minutes with three experimental and three technical replicates. Percent survival was calculated from colony forming units per millilitre of culture before and at the end of treatment, calculated via the Miles and Misra method. Statistical analysis was conducted in GraphPad Prism 6: error bars represent standard deviation; stars indicate the statistical difference between the bars indicated by the ends of the brackets. P-values: * ≤ 0.05; ** ≤ 0.01; *** ≤ 0.001.

To further analyse the effect of $[H_2O_2]$ on these three strains, a concentration that would allow zooming in on the point where % survival drops to zero was selected that lay between the two concentrations that demonstrated this effect in the previous assay. Thus, the assay was performed using concentrations of 0.1, 0.5 and 1 mM with the results treated the same as previously and displayed in Figure 4.5. In this case, the statistical analysis (APPENDIX 5) reveals significant variation between the wild type and ECF1 at 0.1 mM H_2O_2 ($P=0.0015$). The % survival of the ECF1 strain does not decrease as expected between 0.1 and 0.5 mM; experimental variation may have resulted in the value at 0.1 mM being lower than expected. At 0.5 mM H_2O_2 , it is very clear that ECF1 has a significantly higher % survival than the other two strains ($P_{630}=0.0006$, $P_{ECF2}=0.0006$).

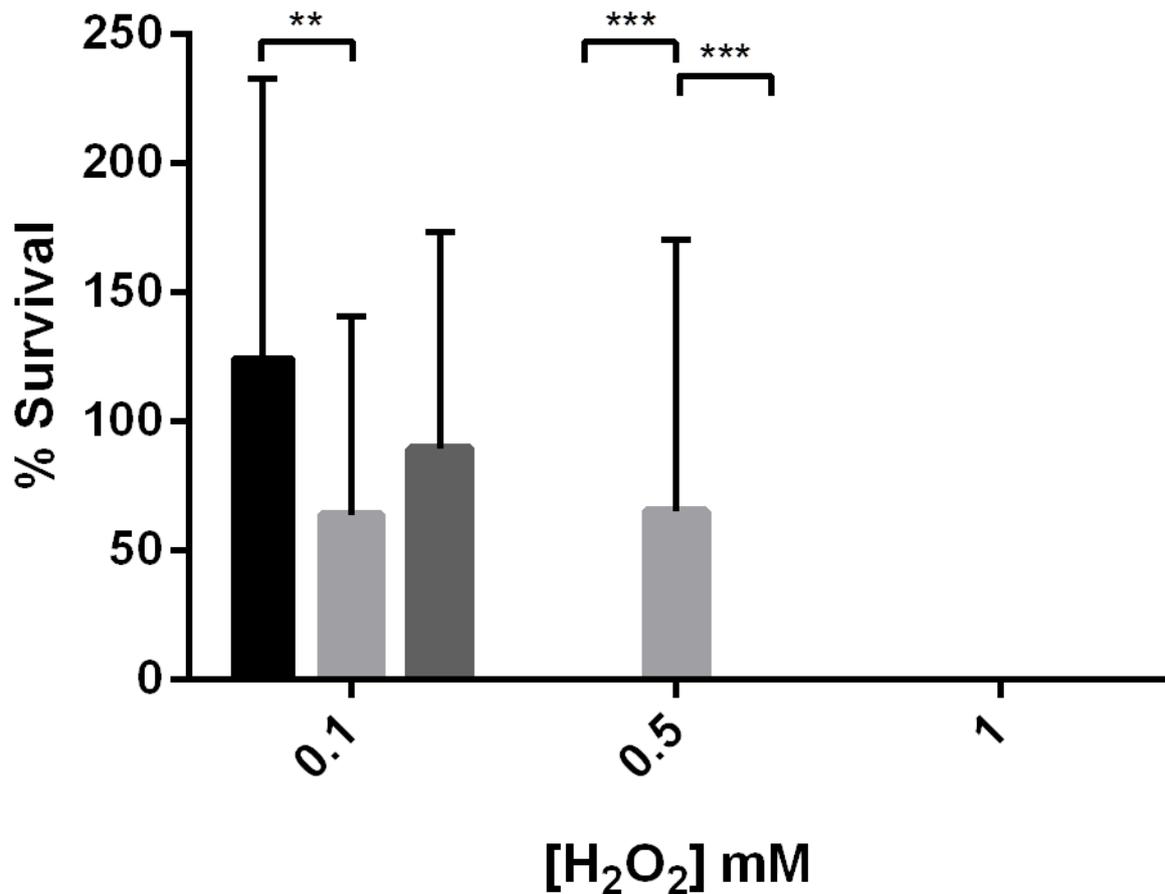


Figure 4.5 Further investigation of the effect of hydrogen peroxide concentration on survival of *C. difficile* strains

To further investigate the effect of hydrogen peroxide on *C. difficile* 630Δ*erm* (black bars), ECF1 (light grey bars) and ECF2 (dark grey bars), broth cultures of these strains were treated with 0.1, 0.5 or 1 mM hydrogen peroxide for 30 minutes in three experimental replicates with three technical replicates. Percent survival was calculated from colony forming units per millilitre of culture before and at the end of treatment, calculated via the Miles and Misra method. Statistical analysis was conducted in GraphPad Prism 6: error bars represent standard deviation; stars indicate the statistical difference between the bars indicated by the ends of the brackets. P-values: ** ≤ 0.01; *** ≤ 0.001.

4.5.2. Effect of incubation time with hydrogen peroxide

It was hypothesised that there may be a difference in response to incubation with hydrogen peroxide over time. To investigate this, 630 Δ *erm*, ECF1 and ECF2 were incubated for either 30 or 45 minutes in 1 mM H₂O₂, with % survival calculated as before, with results in Figure 4.6. It was observed that there was an increase in % survival for all three strains at 45 minutes compared to 30 minutes however statistical analysis with GraphPad Prism 6 (APPENDIX 5) revealed that this was only statistically significant for ECF1 (P=0.0065) and ECF2 (P=0.0215), indicating their greater ability to survive in the presence of hydrogen peroxide when compared to the wild type.

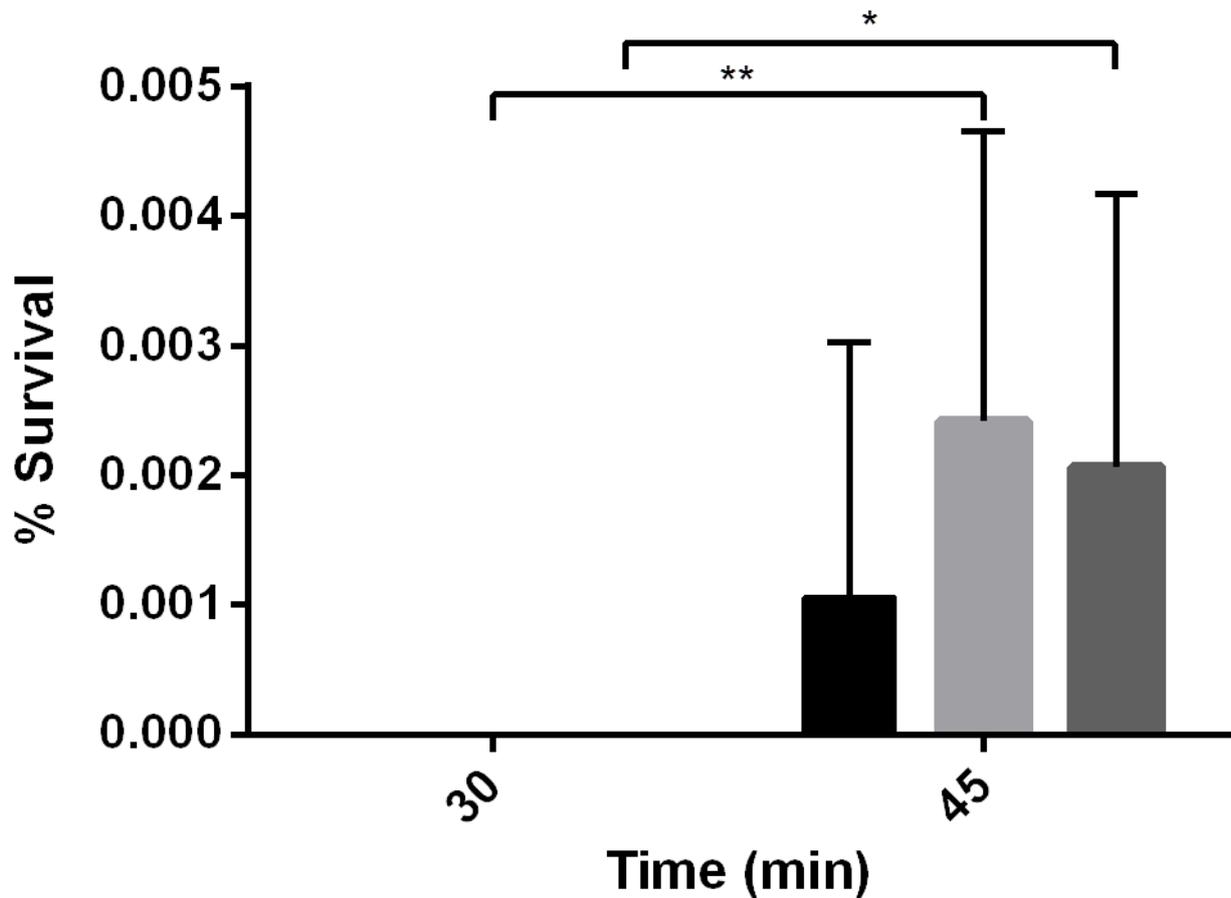


Figure 4.6 Effect of incubation time with hydrogen peroxide on survival of *C. difficile* strains

To find the effect of incubation time during hydrogen peroxide treatment on the survival of *C. difficile* 630Δ*erm* (black bars), ECF1 (light grey bars) and ECF2 (dark grey bars), broth cultures were treated with 1 mM hydrogen peroxide for 30 minutes or 45 minutes in three experimental replicates with three technical replicates. Percent survival was calculated from colony forming units per millilitre of culture before and at the end of treatment, calculated via the Miles and Misra method. Statistical analysis was conducted in GraphPad Prism 6: error bars represent standard deviation; stars indicate the statistical difference between the bars indicated by the ends of the brackets; the difference between % survival of a given strain at each time period. P-values: * ≤ 0.05; ** ≤ 0.01.

4.6. Discussion

The *C. difficile* mutant strains ECF1 and ECF2 have been shown to be phenotypically different to the *C. difficile* 630 Δ *erm* wild type. Firstly, PCR analysis of genomic DNA from each strain was used to confirm the providence of the mutants by demonstration of the insertion of the Group 2 intron. To check that neither mutation had a pleiotropic effect upon the growth of the strain, growth curves for each were generated. Statistical analysis of these curves revealed some significant differences between ECF1 and the wild type at two time points (5h and 6h). The growth rates of the three strains are very similar; the slope of each curve during the logarithmic phase of growth is very similar however, 630 Δ *erm* may have a shorter lag phase, leading to quicker entry into exponential growth, thus the statistical difference in culture optical density at 5h and 6h. Other assays rely on growing cultures of each strain to specific growth phases however, because they initially require growth to stationary phase followed by dilution to a standard culture optical density, any differences in growth between the strains used should not affect the results. Because both mutated genes are thought to be involved in lipoprotein maturation, it was thought that the mutations would have an observable effect on the observed size and localisation of certain proteins due to the lack of signal peptide cleavage. To investigate this, a combination of protein extractions was performed to allow examination of the protein profiles of each strain in various cellular locations. This revealed that there were some minor differences between the three strains, with one significantly brighter band appearing at approximately 24 kDa and a cluster of bands at approximately 15 kDa in a low pH glycine extraction of ECF1 which may correspond to bands present in the detergent phase of a TX-114 extraction of both 630 Δ *erm* and ECF2 but not ECF1, suggesting a different profile of lipoproteins associated with the cell wall. Neither strain has as different a profile relative to wild type 630 as the 630 Δ *erm* Δ *lgt* strain which previous work has identified as having a substantially different protein profile due to a lack of acyl-glyceryl moiety attachment to lipoproteins during maturation leading to many lipoproteins being lost to the culture media and no-longer being isolated in the lipoprotein-enriching TX-114 extraction, similar to observations with strains of *B. subtilis* mutated for *lgt* (206). ECF1 and ECF2 carry mutations in lipoprotein signal peptidase genes which would presumably affect sorting and localisation of prolipoproteins resulting in

accumulation of un-directed prolipoproteins as observed in a *B. subtilis* strain mutated for its SPasell (173), and not directly affect the anchoring of lipoproteins to the cell. Additionally, there was no evidence for the loss of lipoproteins into the culture filtrate for ECF1 and ECF2 (Figure 4.3.A, Culture Filtrate, Lanes 2 and 3) which would indicate a lack of attachment to the cell, reflecting the observation in another study (24) where immature lipoproteins of *S. pneumoniae* lacking Lsp were shown to still be attached to the cell surface. In previous work, both mutant strains were shown to be more susceptible to the toxic effects of malachite green, suggesting a role for correct lipoprotein processing in resistance to this stress (34-36). Additional conditions were sought that may further demonstrate survival or tolerance differences of these strains. Hydrogen peroxide is used extensively in cleaning and disinfection of healthcare environments and has shown significant activity against *C. difficile* spores (100, 105, 207) and there is evidence that *Streptococcus pneumoniae* utilises lipoproteins to protect itself from superoxide and hydrogen peroxide through regulation of Mn²⁺ transport (168), with knockout of the PsaA lipoprotein rendering the strain more susceptible to hydrogen peroxide. It is possible that disruption of lipoprotein maturation in *C. difficile* may render the bacterium more sensitive to hydrogen peroxide treatment. The ability of ECF1 and ECF2 to withstand stress from hydrogen peroxide was examined by incubation with varying concentrations of the chemical, revealing an increased ability in both strains; at 0.1 mM H₂O₂ both ECF1 and ECF2 had a significantly higher % Survival than the 630 wild type. This was not directly carried through in further testing however ECF1 did maintain a significantly higher % Survival at the higher concentration of 0.5 mM H₂O₂. It was also shown that the % Survival of all three strains depends on the length of incubation with H₂O₂; ECF1 and ECF2 exhibited a significantly increased % Survival at 45 minutes compared to 30 minutes incubation time with 1 mM H₂O₂. As these strains are thought to be defective in lipoprotein localisation, any similar system to the Mn²⁺ transport in *S. pneumoniae* which uses the internally localised lipoprotein PsaA could be beneficially disrupted in the mutant strains. Further work is required here to elucidate how such disruption is beneficial and it is suggested that the reason may involve the increased presence of lipoproteins in parts of the cell that are particularly affected by the stress, damping the effect. There may be a greater effect from a deficiency in specific lipoproteins such as a PsaA homolog than

from disruption of maturation of that lipoprotein or the entire lipoproteome. Additionally, further work into the effect of incubation with hydrogen peroxide over time may shed more light on the issue.

Chapter 5. Expression of *C. difficile* LspA and LspA2 in *E. coli*

5.1. Introduction

It has been shown that the expression of a non-native lipoprotein signal peptidase in *E. coli* induces protection of the bacterium against the antimicrobial globomycin (24, 177), the latter having been shown to directly target and inhibit the signal peptidase (31). As such, this protection assay provides a method for demonstrating the functionality of potential lipoprotein signal peptidases. Furthermore, there is potential for this assay to be exploited to utilise analogues of globomycin which may have a better activity against lipoprotein signal peptidase (32, 93) and there may be potential for modification to utilise the antimicrobial myxovirescin which has also been shown to inhibit lipoprotein signal peptidases and which has a demonstrable advantage over globomycin (33).

The globomycin protection assay was selected to investigate the functionality of the two potential lipoprotein signal peptidases of *C. difficile*. The first step in this process is to generate strains of *E. coli* able to express either LspA or LspA2 from *C. difficile* and as such, the genes encoding each protein were cloned into a range of expression vectors. Three vectors were used in this study; two proved to be ineffective while the third, pET28a, proved to be successful. Maps of the initial expression plasmids and intermediary and final pET28a plasmids are shown in Figure 5.1. Only *E. coli* strains holding pET28a expression plasmids produced the anticipated traces on Western blots, thus demonstrating the successful generation of strains for use in globomycin protection assays with *C. difficile* lipoprotein signal peptidases.

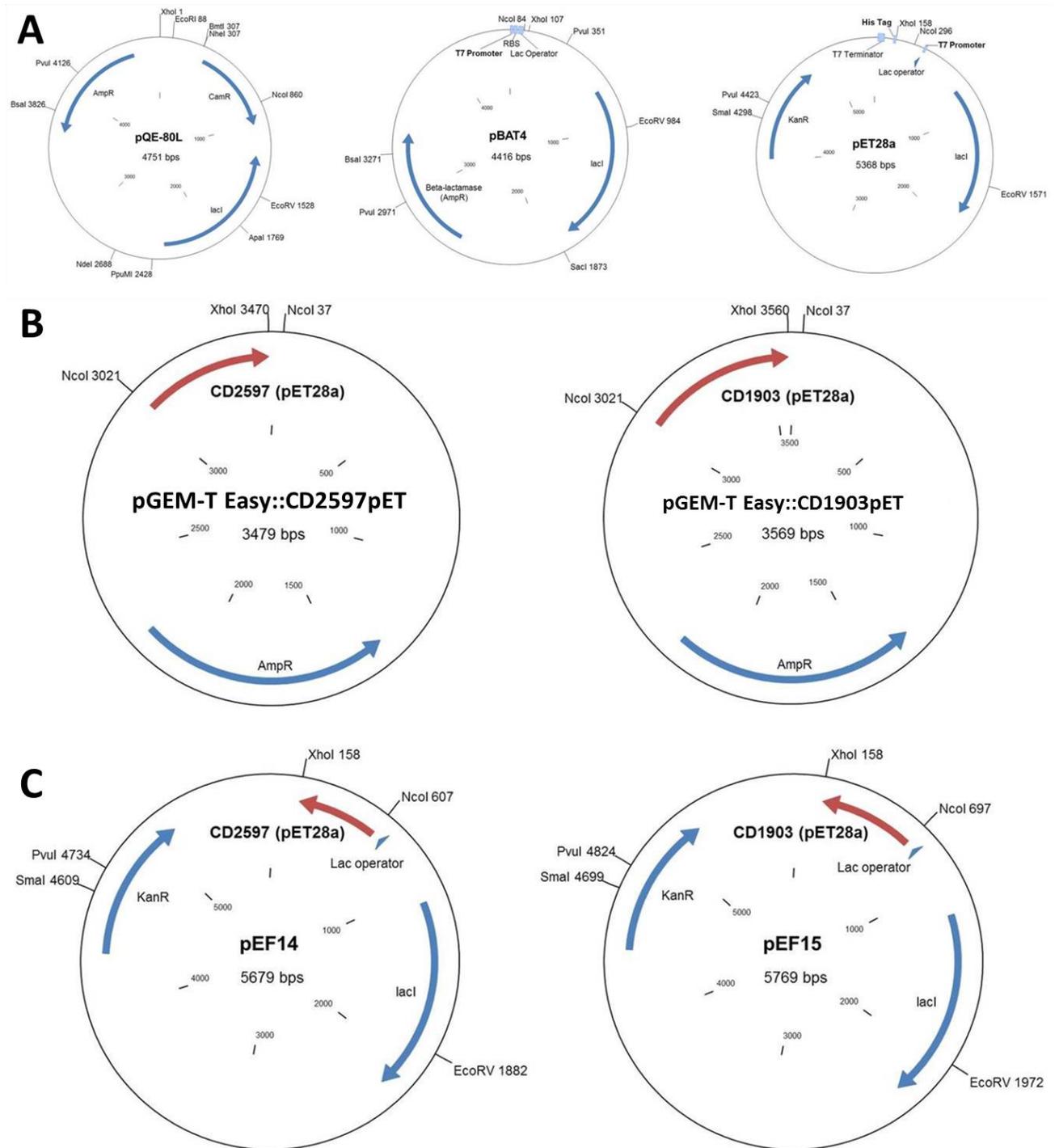


Figure 5.1 Plasmids used for the expression of *C. difficile* LspA (CD2597) and LspA2 (CD1903) in *E. coli*

A: Expression vectors selected for use in this study. **B:** pGEM-T Easy TA cloning vector holding PCR fragments of CD2597 and CD1903 for insertion into pET28a. **C:** Maps of expression plasmids pEF14 and pEF15; pET28a holding CD2597 and CD1903 respectively, cloned downstream of a Lac operator using the *Xho*1 and *Nco*1 restriction sites.

5.2. Cloning Methodology

To generate *E. coli* strains expressing either *C. difficile* LspA or LspA2, expression plasmids were selected to receive PCR products encoding either gene. Either the plasmid or the PCR product was required to encode a His-tag which would be expressed in combination with the desired *C. difficile* protein, allowing detection of expression via anti-His Western blotting. The generation of plasmids and subsequent testing of LspA or LspA2 expression in *E. coli* are described below.

5.2.1. pQE80: pRL4 and pRL6

Initially, cloning was planned using pQE80 as the expression plasmid; the cloning of both genes was performed in Dr. S. L. Michell's laboratory at the University of Exeter by R. Lawrence (unpublished) and the plasmids donated to this study, maps of which are shown in APPENDIX 7. These plasmids, pRL4 and pRL6 for expression of CD1903 and CD2597 respectively, contain an N-terminal His-tag co-expressed with the intended protein. The DNA and amino acid sequences at the termini of the expressed proteins in both plasmids are shown in APPENDIX 6.

These plasmids were transformed into *E. coli* C43 (DE3) (BL21), a strain commonly used for expression of proteins, and subsequently used in expression experiments via isopropyl β -D-1-thiogalactopyranoside (IPTG) induction at a concentration of 4 mM IPTG, added once cultures reached an optical density of 0.6 at 595 nm. *E. coli* C43 (DE3) (BL21) holding pNIC-KSA1, a plasmid designed for the expression of the CD0873 *C. difficile* lipoprotein (21) with a His-tag was included as a positive control for Western blotting. Proteins were extracted from cultures and subjected to SDS-PAGE before use in Western blotting; a Western blot of non-induced and IPTG-induced cultures of these strains is shown in Figure 5.2. There is clear expression of CD0873 from pNIC-KSA1 in both non-induced and induced cultures (lane 1) indicating some leaky expression however expression was much stronger in induced cultures indicating that induction was occurring as a response to IPTG treatment. Conversely, there is no evidence of expression in lanes 2 and 3, indicating that *C. difficile* LspA and LspA2 are not expressed from pRL4 or pRL6.

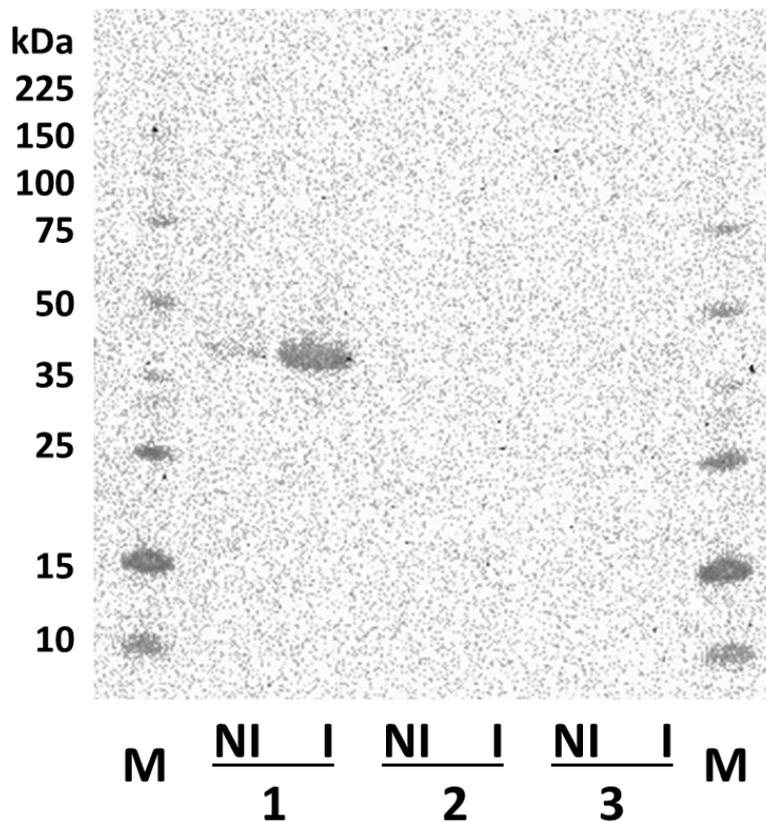


Figure 5.2 Anti-His Western blot of protein extractions from non-induced and IPTG-induced cultures of *E. coli* C43 (DE3) (BL21) harbouring expression plasmids for *C. difficile* *LspA* and *LspA2* using pQE80 as a backbone

To demonstrate expression of a protein with a His tag a Western blot was performed on protein extractions of IPTG-induced cultures of *E. coli* C43 (DE3) (BL21) holding pRL4 (2), pRL6 (3) or pNIC-KSA1 (1) as a positive control, showing a positive result in lane 1 only. Bands were very faint, resulting in high exposure of the membrane, causing specking of the background. NI – Non-Induced, I – Induced.

5.2.2. Workflow for subsequent cloning attempts

As there was no expression observed for *E. coli* C43 (DE3) (BL21) holding either pRL4 or pRL6, indicating a failure to produce CD1903 and CD2597, alternative approaches were sought. A cloning workflow was designed for the generation of new plasmids for the expression of *C. difficile* LspA and LspA2 and is shown in Figure 5.3. This workflow included check steps (red boxes) which permitted progress monitoring, allowing detection of errors prior to conducting further experimentation.

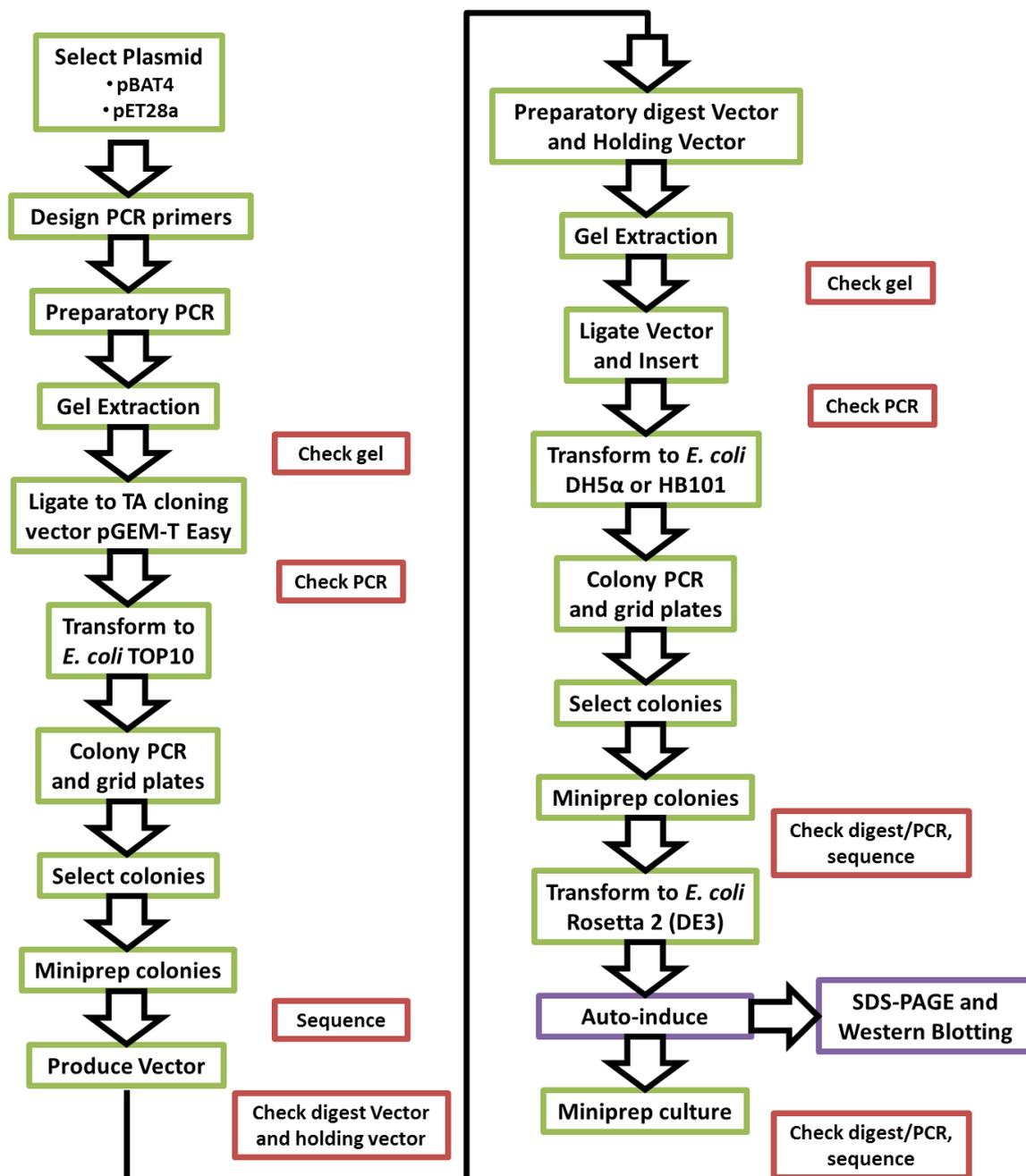


Figure 5.3 Cloning methodology and workflow for the expression of *LspA* and *LspA2* from *C. difficile* in *E. coli*

To allow the expression of a *C. difficile* protein in *E. coli*, the encoding gene must be cloned into an appropriate plasmid that will allow controlled expression. Plasmids pBAT4 and pET28a were used for cloning in this study; the *C. difficile* genes *LspA* (CD2597) and *LspA2* (CD1903) were cloned into these plasmids using the TA-cloning vector pGEM-T Easy as an intermediary for the PCR insert. Various checks were performed during the cloning process (red boxes) to allow confirmation of correct cloning and no change in DNA sequence after induction of expression.

5.2.3. pBAT4: pEF10, pEF11, pEF12 and pEF13

The second attempt at cloning *C. difficile* *lspA* and *lspA2* into an expression plasmid for expression of LspA and LspA2 in *E. coli* was performed using the vector pBAT4. This vector does not contain a His-tag; therefore it was necessary to include a His-tag in the cloning process to be present at the C-terminus of the proteins. Both *lspA* and *lspA2* possess the uncommon start codon TTG (Leucine); it was thought that this may be detrimental to expression in *E. coli*, therefore two sets of plasmids were produced that aimed to eliminate this issue, both using the *Nco1* restriction site of pBAT4. Firstly pEF10 and pEF12 were produced via PCR amplification of *lspA* and *lspA2* respectively using primers CD2597_pBAT4_F, CD2597_pBAT4_R, CD1903_pBAT4_F and CD1903_pBAT4_R (Table 4) as appropriate. The reverse primers (_R) encoded a six moiety His-tag and an *Nco1* restriction site at the 3' end of the products and the forward primers (_F) encoded both an *Nco1* site and a 10 base insertion which changed the start codon of both genes to ATG (Methionine). Secondly, pEF11 and pEF13 were produced via the same method utilising the reverse primers used for pEF12 and pEF12 with two new forward primers CD2597_pBAT4_F2 and CD1903_pBAT4_F2 (Table 4). These forward primers again contained an *Nco1* site and changed the first two codons of both genes to ATG GCT (Methionine, Alanine). These strategies are shown in APPENDIX 6 and the plasmids produced are shown in APPENDIX 7.

5.2.3.1. Rare codon usage in pBAT4-derived plasmids

Due to redundancy in genetic code with more than one codon available for each amino acid, species of bacteria can utilize a different codon bias. This is true for *E. coli* and *C. difficile* and may present a barrier to the expression of *C. difficile* proteins in *E. coli*, resulting in little or no expression (208). Genes can either be codon-optimized for the expression species, altering the gene but not the amino acid sequence (140, 209), or the expression strain can be supplemented with the tRNAs which complement rare codons. These tRNAs can be encoded on a plasmid such as pRARE2; this plasmid is held by the *E. coli* expression strain Rosetta2 (DE3). Both CD2597 and CD1903 have a lot of rare codon usage; Rare Codon Caltor (204) analysis of these genes as they appear in pBAT4-derived plasmids is shown in APPENDIX 9 and demonstrates the level of rare codon usage.

5.2.3.2. Auto-induction of protein expression from pBAT4-derived plasmids

IPTG is a molecular mimic of a lactose metabolite which activates expression from a *T7lac* promoter in plasmids also containing a *lacI* gene and a *lac* operator. An in-depth investigation into the mechanisms of this induction has led to the development of a system for auto-induction of protein expression in high-density cultures using a simplified protocol and lactose as the inducing agent (205). This system produces much higher yields of protein than IPTG induction, producing more easily-detectable protein traces during analysis and was therefore selected for use in future expression experiments.

To investigate expression of *C. difficile* LspA or LspA2 from pBAT4-derived plasmids, *E. coli* Rosetta2 (DE3) holding these plasmids were grown in 200 ml auto-inducing (ZYM-5052) and non-inducing (ZYM-505) media to an optical density of approximately 15 at 595 nm. *E. coli* C43 (DE3) (BL21) holding pNIC-KSA1 was again included as a positive control. Samples of non-induced and auto-induced cultures were subjected to crude protein extraction followed by SDS-PAGE and anti-His Western blotting; the Western blot is shown in Figure 5.4. As before, there is expression of CD0873 (lane C) with no expression of His-tagged protein in any other strain (lanes 1 through 5), indicating that there is no expression of *C. difficile* LspA or LspA2 from pEF10, pEF11, pEF12 or pEF13. Plasmid DNA was extracted from post-auto-induction cultures and sent for sequencing analysis to investigate the possibility that mutations had occurred which could prevent protein expression. Only the sample for pEF10 returned useable results which revealed potential mutations in the regions flanking the gene insertion site. This analysis is shown in APPENDIX 8 and the potential mutations revealed may explain the lack of expression.

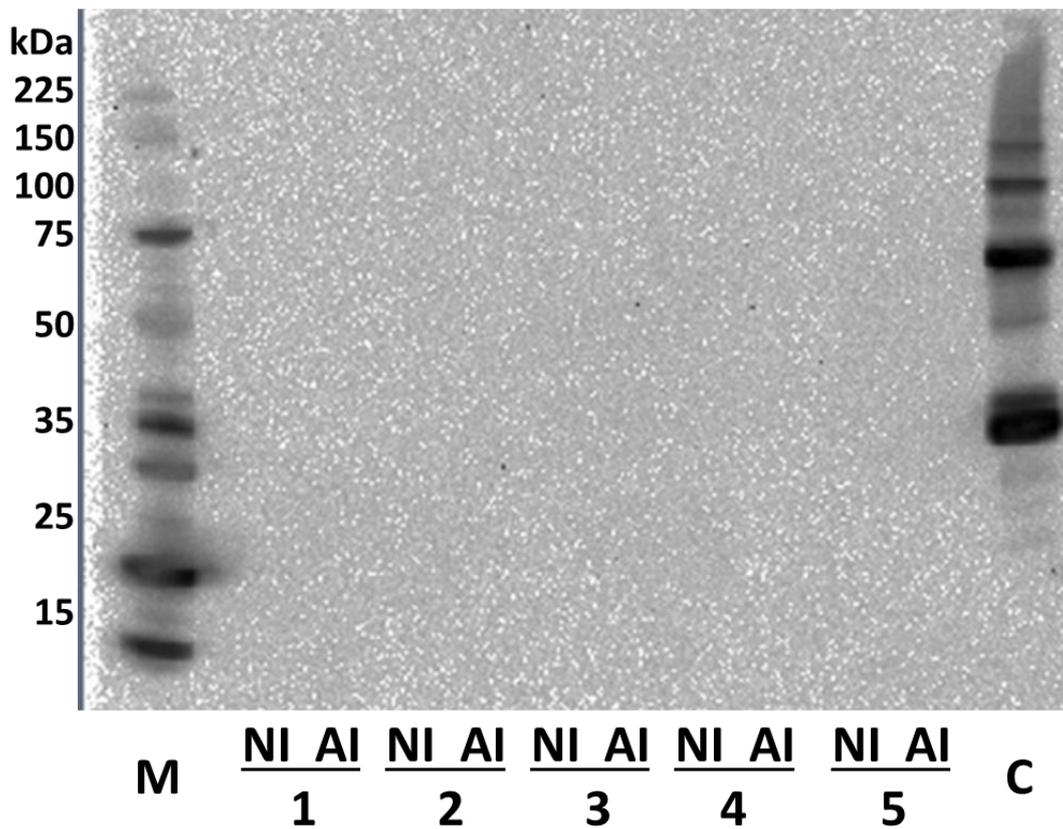


Figure 5.4 Anti-His Western blot of protein extractions from non-induced and auto-induced cultures of *E. coli* harbouring expression plasmids for *C. difficile* *LspA* or *LspA2* using *pBAT4* as a backbone

To demonstrate expression of a protein with a His-tag in auto-induced cultures of *E. coli* Rosetta2 (DE3) a Western blot was performed on crude protein extractions for 1: pEF10, 2: pEF11, 3: pEF12, 4: pEF13, C: *E. coli* C43 (DE3) (BL21) pNIC-KSA1. A positive result is shown in the control (C) lane only. NI – Non-Induced, AI – Auto-induced.

5.2.4. pET28a: pEF14 and pEF15

As no expression was obtained from any of the pBAT4 plasmids, another attempt was required. The expression plasmid pET28a was selected for this attempt; it has a His-tag downstream of its multiple cloning site, negating the need to introduce a tag via PCR. Primers were designed to produce a product with an *NcoI* restriction site at the 5' end and an *XhoI* restriction site at the 3' end to allow simpler processing. As discussed in Section 0, the presence of a TTG start codon in both genes may be disruptive to expression attempts therefore the start codon of each gene was replaced with ATG GGC (Methionine, Glycine) using PCR with primers EFTC_pET28a_CD63025970_F and EFTC_pET28a_CD63025970_R for *lspA* to produce pEF14 and EFTC_pET28a_CD63019030_F and EFTC_CD63019030_R for *lspA2* to produce pEF15 (Table 4).

5.2.4.1. Generation of pEF14 and pEF15

Following PCR amplification of the genes using the primers named above, products were ligated into the TA-cloning vector pGEM-T Easy to produce intermediate plasmids (pGEM-T Easy::CD2597pET and pGEM-T East::CD1903pET), shown in Figure 5.1.B. The presence of the insert was demonstrated by restriction digest, resulting in fragments of expected sizes (Figure 5.5). Large volume digests of these plasmids and the vector pET28a were conducted, followed by gel electrophoresis and gel extraction of each insert and vector; the inserts were then ligated separately to the pET28a vector fragment. Ligations and controls were transformed by electroporation into *E. coli* HB101 after unsuccessful attempts to transform into *E. coli* TOP10 by heat shock. Transformants were picked at random for screening by colony PCR; four putative pEF14 transformants and three putative pEF15 transformants. Two PCR screens utilised either the primers originally used to produce the insert fragments or primers for the T7 promoter and T7 terminator (Table 4), sites for which are present in pET28a, flanking the multiple cloning site, shown in Figure 5.5.B and C. All four pEF14 transformants and two pEF15 transformants (lane 7 and 9) gave a positive result with the insert amplification primers; these six transformants were also positive with the T7 primers. The colonies that produced the template DNA for the pEF14 transformant in lane 2 and the colony that produced the template DNA for the pEF15 transformant in lane 9 were taken forward as strains holding pEF14 and pEF15. Rare Codon Caltor analysis

of the CD2597 or CD1903-encoding regions of these plasmids is shown in APPENDIX 9, revealing the continued presence of high levels of rare codon usage.

5.2.4.2. Cloning error checking by sequencing

To ensure that no sequence errors were present in the region containing each gene in the two plasmids samples were sent for DNA sequencing using the primer pair for the T7 promoter and T7 terminator and a primer pair designed to bind within each gene allowing sequence reads to cover the entirety of the gene sequenced and the neighbouring regions of the plasmid (CD2507_pCheck_L and CD2597_pCheck_R for pEF14 and CD1903_pCheck_L and CD1903_pCheck_R for pEF15) (Table 4). Figure 5.6 gives a representation of the sequencing results showing the alignment of translated results to the *in silico* reference protein amino acid sequence. Sequencing results for both genes returned a few ambiguous bases resulting in unknown amino acids which are represented by an X in the alignments. Each of these occurrences was due to errors in sequencing for one or more reactions however in every case there was at least one sequencing reaction that returned the expected base; this can be seen in APPENDIX 10, where the sequencing results are shown aligned to reference DNA sequences. These results suggest that both CD2597 and CD1903 have been successfully cloned into pEF14 and pEF15 respectively and that His-tags are present in a correct form.

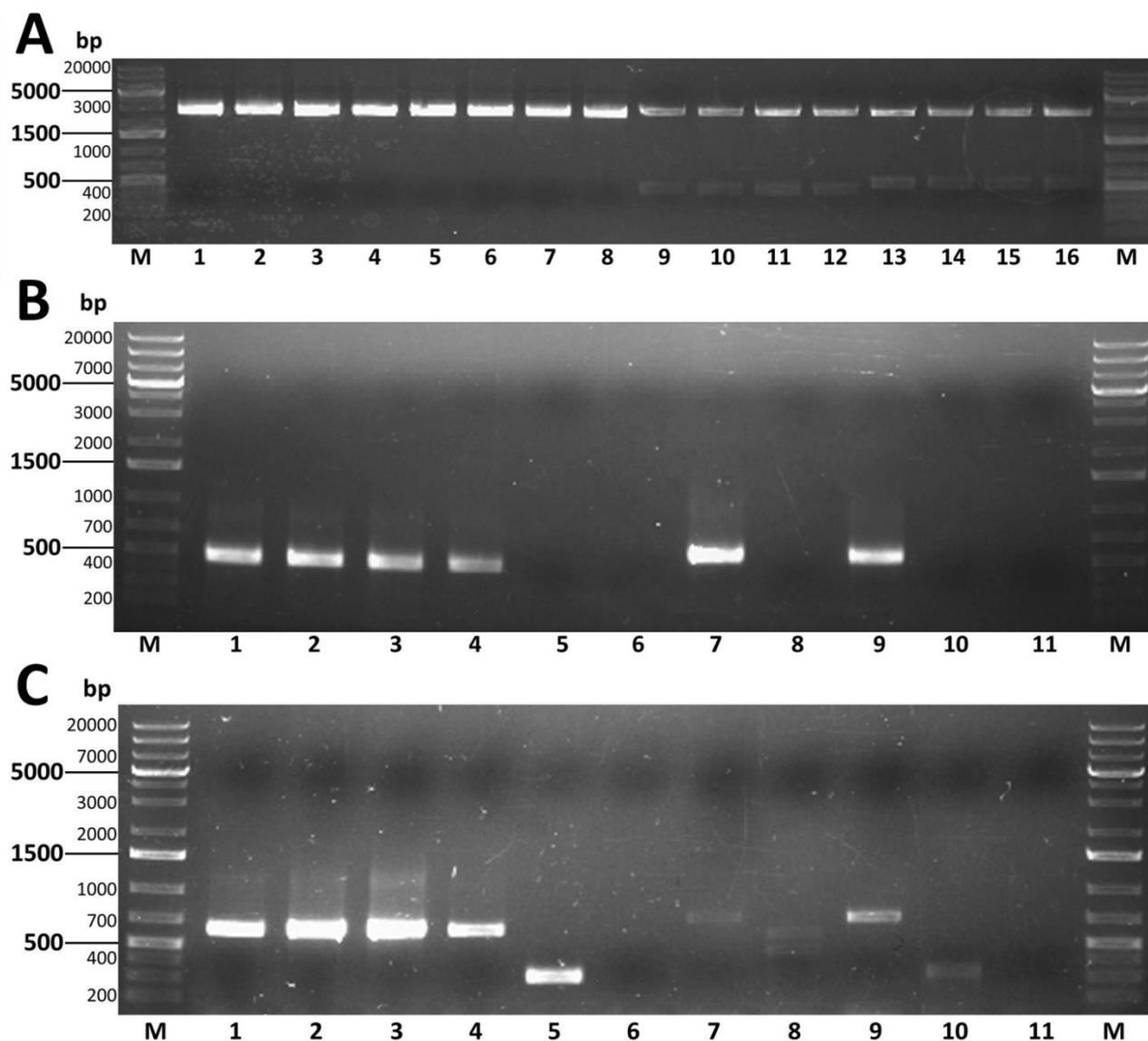


Figure 5.5 Restriction endonuclease digests and colony PCRs to demonstrate cloning of CD2597 and CD1903 into pGEM-T Easy (A) and pET28a (B & C)

A: Restriction endonuclease digests of plasmid minipreps of *E. coli* TOP10 transformation clones holding pGEM-T Easy with either CD2597 or CD1903 PCR fragments for cloning into pET28a. Lanes 1 through 4: CD2597 fragment undigested minipreps. Lanes 5 through 8: CD1903 fragment undigested minipreps. Lanes 9 through 12: CD2597 fragment Nco1/Xho1 digests. Lanes 12 through 16: CD1903 fragment Nco1/Xho1 digests. **B** and **C:** Gels of *E. coli* HB101 colony PCR from pET28a ligation transformations with fragment amplification primers (**B**) and T7 forward and reverse primers (**C**). Lanes 1 through 6: CD2597 fragment. Lanes 7 through 11: CD1903 fragment. Whole pET28a controls in lanes 5 and 10. PCR Negative controls in lanes 6 and 11.

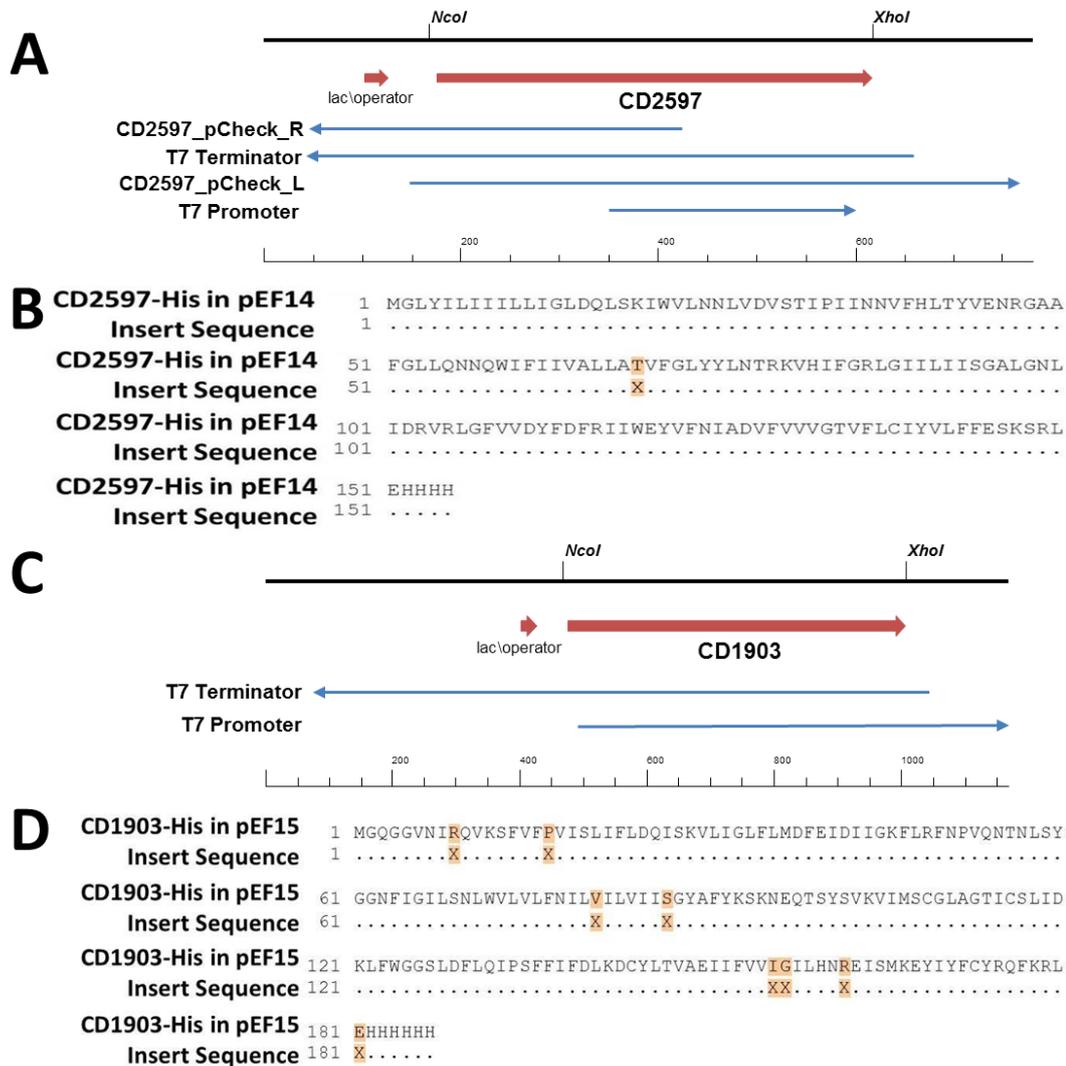


Figure 5.6 Alignment of the DNA sequences of pEF14 and pEF15 to *in silico* references to check for sequence errors

After initial PCR confirmation, plasmid minipreps samples from *E. coli* HB101 holding pEF14 or pEF15 were sequenced to allow identification of sequence errors and screening for SNPs or other mutations. **A:** Pictorial representation of DNA sequencing read alignments to *in silico* pEF14 reference sequence. **B:** Alignment of the translation of the sequencing reads contig to the *in silico* reference amino acid sequence of CD2597 with 3' His tag in pEF14 showing one potentially variant amino acid. **C:** Pictorial representation of DNA sequencing read alignments to *in silico* pEF15 reference sequence; poor sequence traces were returned for reactions with 2597_pCheckL/R primers and did not align. **D:** Alignment of the translation of the sequencing reads contig to the *in silico* reference amino acid sequence of CD1903 with 3' His tag in pEF15 showing 8 potentially variant amino acids.

5.2.4.3. Transformation of pET28a, pEF14 and pEF15 into *E. coli* Rosetta2 (DE3)

Plasmids pEF14, pEF15 and pET28a (for use as an expression-negative control) were transformed into *E. coli* Rosetta2 (DE3) by electroporation. Putative transformants were selected for on media containing appropriate antibiotics and via colony PCR using primers directed against the T7 promoter and T7 terminator regions that flank the multiple insertion site within the plasmids. The results of this PCR are shown in Figure 5.7 and reveal that four from four tested putative pET28a transformants held pET28a; two from four tested putative pEF14 transformants held pEF14 and two from four tested putative pEF15 transformants held pEF15. As a result, the colonies which produced the brightest PCR products were selected to be carried forward into expression experiments: pET28a colony 2, pEF14 colony 1 and pEF15 colony 2.

5.2.4.4. Auto-induction of protein expression from pEF14 and pEF15 and subsequent sequencing analysis

E. coli Rosetta2 (DE3) strains holding pEF14, pEF15 and pET28a were subjected to auto-induction as described for strains holding pBAT4-derived plasmids (Section 5.2.3.2) with *E. coli* C43 (DE3) (BL21) pNIC-KSA1 again included as a positive control for anti-His Western blotting. To check that no mutations had occurred within either pEF14 or pEF15 during auto-induction DNA was extracted from post-auto-induction cultures and sent for sequencing analysis. The results of this analysis are shown in Figure 5.8; no ambiguous amino acids are present in alignments of either set of sequencing results to their corresponding references, suggesting that no mutations had occurred. As such, LspA and LspA2 produced under auto-inducing conditions by these strains would be as identical to the form found in *C. difficile* as possible, with the addition of a His-tag. Alignments of sequencing results contigs to reference sequences showing no sequencing differences are shown in APPENDIX 11. Because the sequences of pEF14 and pEF15 are correct, proteins expressed under auto-inducing conditions which have a His-tag will be correct versions of *C. difficile* LspA and LspA2; Western blotting is required to confirm such expression.

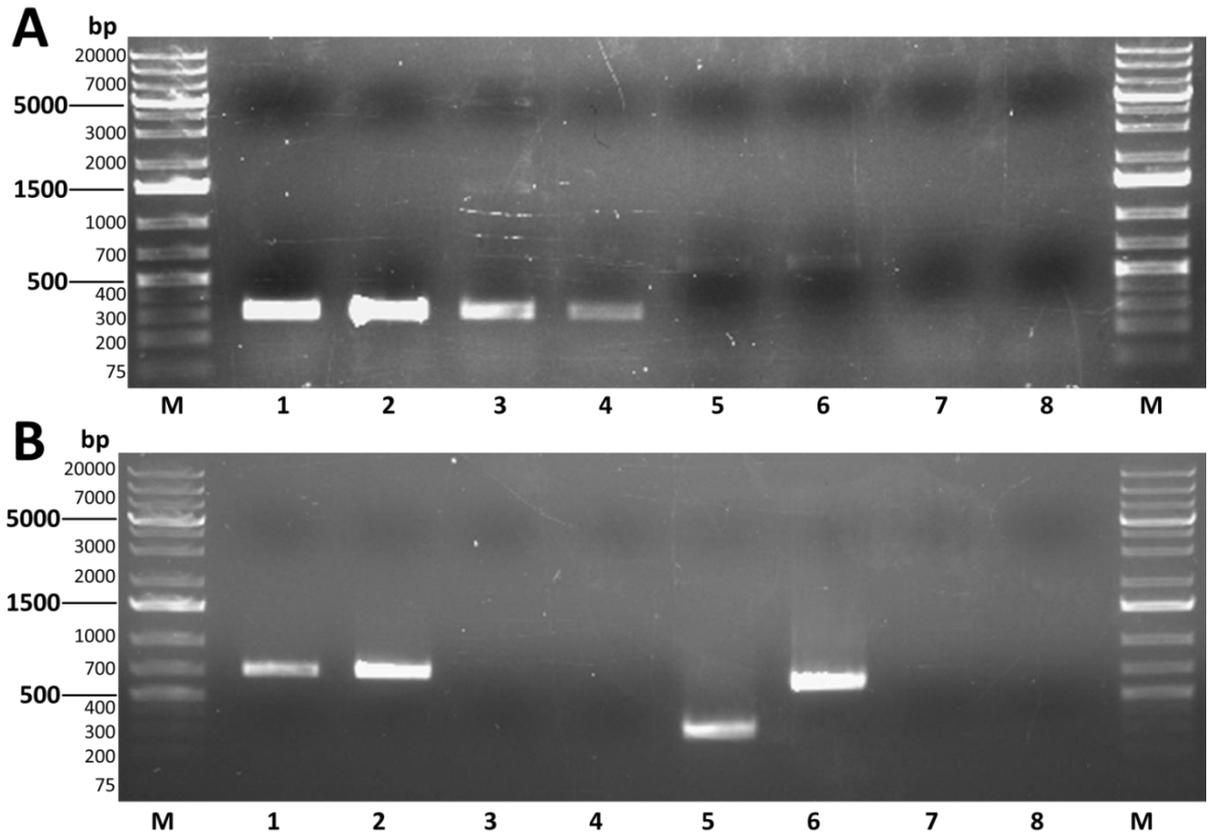


Figure 5.7 Colony PCR to demonstrate transformation of pET28a, pEF14 and pEF15 into *E. coli* Rosetta2 (DE3)

Gel electrophoresis of colony PCR using T7 forward and reverse primers of *E. coli* Rosetta2 (DE3) transformants with pEF28a, pEF14 or pEF15 was performed to confirm successful transformation. **A:** Lanes 1 through 4: pET29a putative transformants. Lanes 5 through 6: pEF14 putative transformants. **B:** Lanes 1 through 4: pEF15 putative transformants. Lane 5: pET28a control. Lane 6: pEF14 control. Lane 7: pEF15 control. Lane 8: negative control.

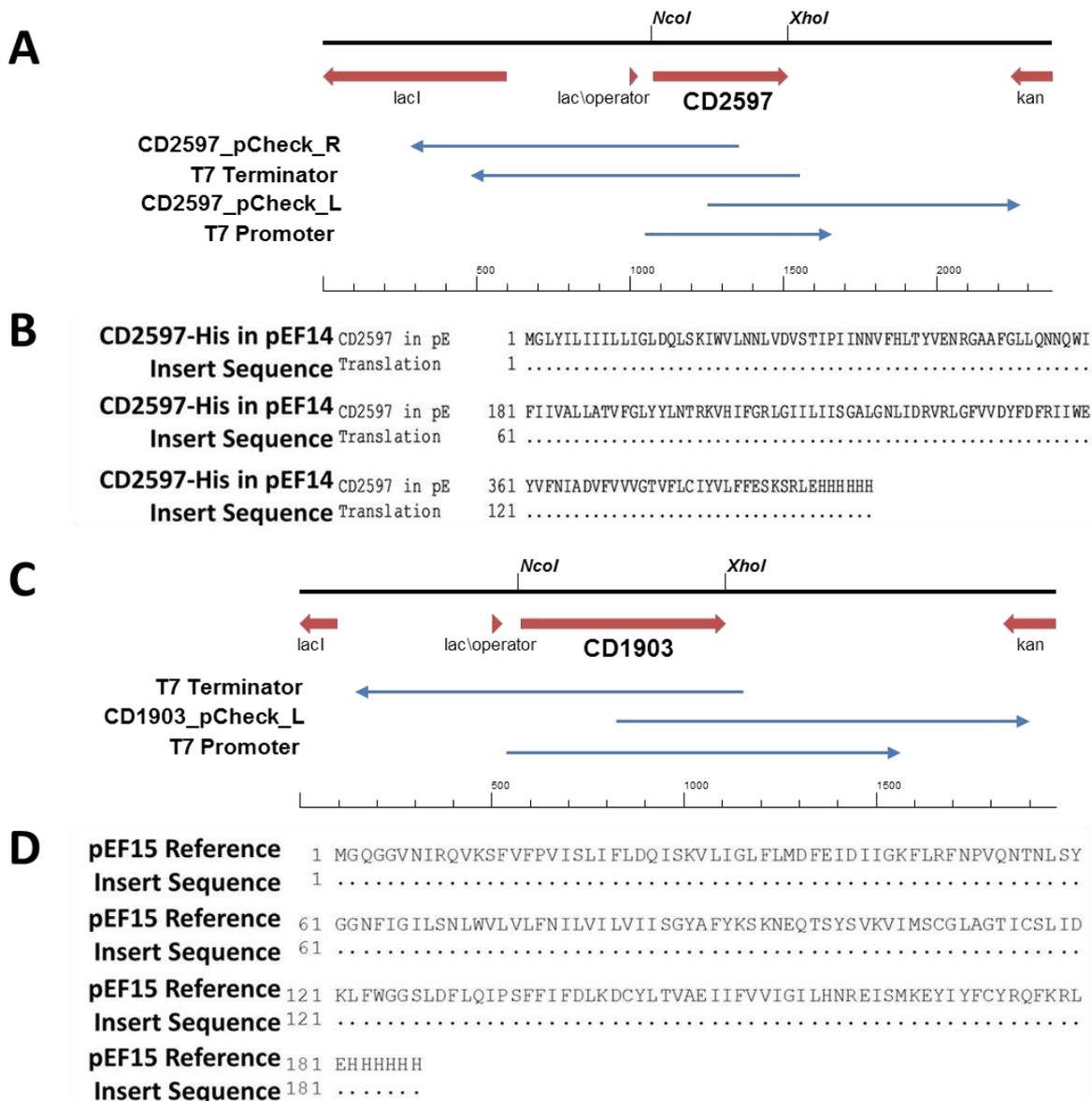


Figure 5.8 Analysis of the DNA sequences of pEF14 and pEF15 to in silico references to check for sequence errors post auto-induction

Due to the potential for mutations in expression plasmid sequence during induction of expression where the expressed protein may have toxic effects on the host cell, post-auto-induction cultures were subjected to plasmid minipreps and samples sent for DNA sequencing to allow checking for mutations. **A:** Pictorial representation of DNA sequencing read alignments to *in silico* pEF14 reference sequence. **B:** Alignment of the translation of the sequence contig to the translation of the *in silico* pEF14 reference showing no amino acid differences. **C:** Pictorial representation of DNA sequencing read alignments to *in silico* pEF15 reference sequence. **D:** Alignment of the translation of the sequence contig to the translation of the *in silico* pEF15 reference showing no amino acid differences.

5.2.4.1. Anti-His Western blotting

To demonstrate that *C. difficile* LspA and LspA2 are expressed under auto-inducing conditions in *E. coli* strains holding pEF14 and pEF15 respectively, Western blotting was performed on crude protein extractions of non-induced and auto-induced cultures. Figure 5.9 shows a safe-stain-treated SDS-PAGE gel of these crude protein extractions (A) and the anti-His Western blot (B). There were no changes in protein expression profiles between *E. coli* Rosetta2 (DE3) strains holding pET28a, pEF14 or pEF15 under either non-inducing or auto-inducing conditions. There is, however, a band which changes intensity between the non-induced and auto-induced samples of *E. coli* C43 (DE3) (BL21) pNIC-KSA1 at approximately 40 kDa (A Lane 4), indicating expression of CD0873 under auto-inducing conditions. Western blotting reveals bands at approximately 18 kDa in Lane 2, 21 kDa in Lane 3 and 40 kDa in Lane 4. These correspond to the expected protein sizes for LspA, LspA2 and CD0873 respectively, indicating successful expression of these *C. difficile* proteins in *E. coli*. Bands are present in both non-induced and auto-induced samples in Lane 2 and Lane 4, suggesting leaky expression in these strains.

It is most desirable for *C. difficile* LspA and LspA2 to be expressed and located in the surface layers of the *E. coli* strain. To test this, samples of auto-induced cultures were subjected to low pH glycine protein extractions to enrich for surface layer proteins (202) before SDS-PAGE analysis and anti-His Western blotting. The Ponceau-S-stained nitrocellulose membrane used in blotting (A) and the Western blot (B) are shown in Figure 5.10. The stained membrane may show some bands at increased intensities in Lanes 2, 3 and 4, compared to the negative pET28a control in Lane 1, that are close to the expected molecular weights. The Western blot reveals very clear bands in Lanes 2, 3 and 4 that are at the expected sizes for *C. difficile* LspA, LspA2 and CD0873 respectively. This indicates that these proteins are all localised to the surface layer of *E. coli* when expressed under auto-inducing conditions.

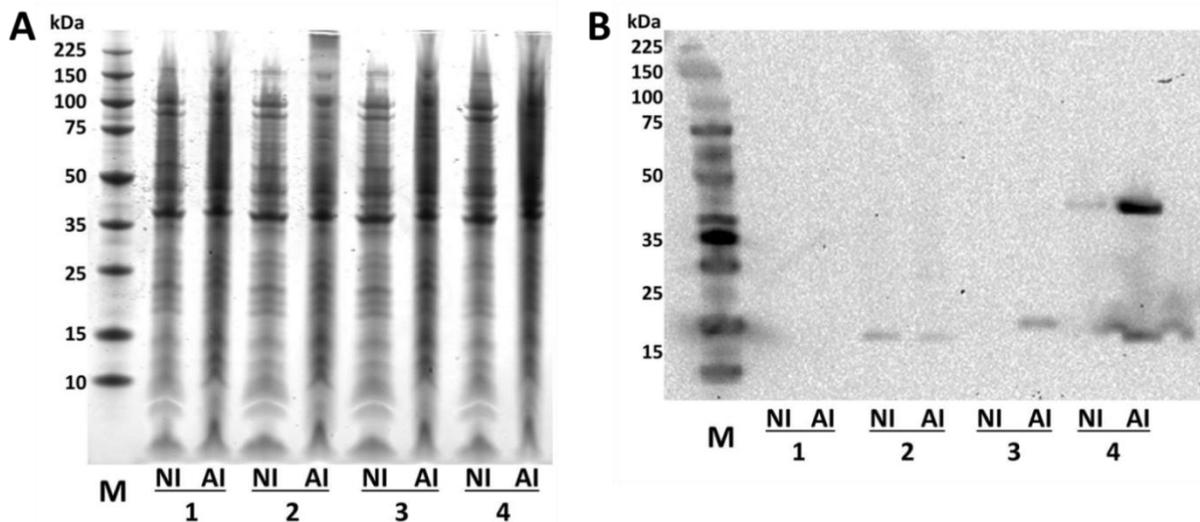


Figure 5.9 Safe-stain-treated SDS-PAGE and anti-His western blot of crude protein extractions of *E. coli* Rosetta2 (DE3) harbouring pET28a, pEF14 or pEF15 and *E. coli* C43 (DE3) (BL21) harbouring pNIC-KSA1 pre- and post-auto-induction

To demonstrate the expression of *C. difficile* CD2597 and CD1903 in *E. coli*, cultures were subjected to auto-induction and protein extracts examined for the presence of the expected proteins. **A:** SDS-PAGE of crude protein extracts of non-induced and auto-induced *E. coli* Rosetta2 (DE3) or C43 (DE3) (BL21) cultures to look for changes in band brightness or novel bands after auto-induction. **B:** Anti-His western blot of the SDS-PAGE gel in A to confirm the presence of proteins containing a His tag. Both images, 1: pET28a, 2: pEF14, 3: pEF15, 4: pNIC-KSA1. NI – Non-induced, AI – Auto-induced.

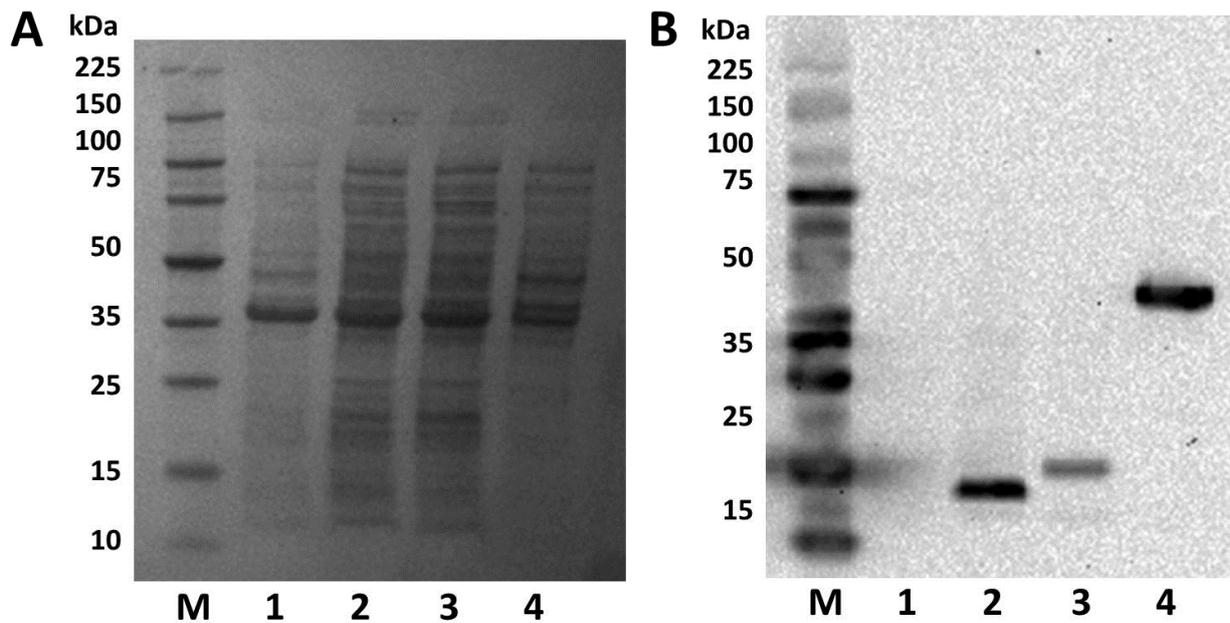


Figure 5.10 *Ponceau-S-stained nitrocellulose membrane and anti-His western blot of low pH glycine protein extractions of E. coli Rosetta2 (DE3) harbouring pET28a, pEF14 or pEF15 and E. coli C43 (DE3) (BL21) harbouring pNIC-KSA1 post-auto-induction*

To demonstrate the localisation of *C. difficile* CD2597 and CD1903 to the surface layers of *E. coli* when expressed via auto-induction of *E. coli* Rosetta2 (DE3) holding pEF14 or pEF15, low pH glycine extractions were performed on auto-induced cultures and examined for the presence of His-tagged proteins by western blotting. **A:** Ponceau-S-stained nitrocellulose membrane after blotting of protein extraction SDS-PAGE. **B:** Anti-His western of the membrane in A. showing very clear expression of His-tagged proteins in all but lane 1. Both images, 1: pET28a, 2: pEF14, 3: pEF15, 4: pNIC-KSA1.

5.3. Discussion

Plasmids for the expression of *C. difficile* LspA and LspA2 in *E. coli* have been generated after two unsuccessful attempts. The first two attempts which used either pQE80 or pBAT4 as the expression plasmid did not produce the desired end result of clear traces on Western blots probed for His-tagged proteins. The first attempt (pQE80) utilised an N-terminal His-tag and the native TTG start codon for both proteins. It was thought that these issues may have contributed to the lack of expression therefore both were changed for the second attempt using pBAT4. In this case, the start codon was changed in two separate cases by both an insertion and a forced mutation to ATG. It was also thought that the induction of expression was not strong enough therefore auto-induction was introduced aiming to produce more *E. coli* cells each producing a higher yield. Again, there was no evidence of expression of either LspA or LspA2 and it was found that there may have been mutations within the *lacI* region of one of the final plasmids, suggesting that expression was being prevented by a non-functional system.

The third and successful attempt utilised pET28a as the expression plasmid; it contains a His-tag downstream of the multiple cloning site, negating the need to include one during cloning. Each stage of cloning was rigorously tested to ensure that no errors in plasmid sequence were present, maximising the probability of success; no errors were found and the plasmids pEF14 (for LspA) and pEF15 (for LspA2) were successfully generated. A final barrier to expression may have been the frequent usage of rare codons within both *lspA* and *lspA2*, therefore pEF14 and pEF15 were transformed into *E. coli* Rosetta2 (DE3) which holds the pRARE2 plasmid, encoding tRNAs for rare codons. Either the use of this strain or the use of pET28a lead to successful expression of *C. difficile* LspA and LspA2. Initial analysis of non-induced and auto-induced samples using SDS-PAGE did not reveal evidence of expression with no high brightness bands visible in auto-induced samples that were not present in non-induced samples. After Western blotting, protein bands were only visible at expected sizes, indicating that the level of expression may not be as high as aimed for. Indeed, the pNIC-KSA1-containing positive control strain exhibited far greater expression of CD0873 which was visible via SDS-PAGE. It may be necessary to investigate improving the level of expression before pEF14 and

pEF15 are used in globomycin protection assays because a low quantity of expressed protein can only lead to a low level of protection which may not lead to significantly greater strain survival than the *E. coli* native protein.

Of interest, it was shown by Western blot of low pH glycine protein extractions of auto-induced cultures that LspA, LspA2 and CD0873 were being localised to the surface layer of *E. coli*. This suggests that they will be well located for functionality in *E. coli*, enabling globomycin protection assays to be performed.

Chapter 6. Discussion

Clostridium difficile presents a significant problem to healthcare in many countries. At present it primarily affects patients within the healthcare setting; however the nature of the disease may be shifting to an increased effect upon otherwise healthy individuals in the community. While there has been a decrease in cases since the introduction of more rigorous infection control measures and improved treatment strategies, the shifting nature of the disease and current lack of a truly effective treatment demand that attention is not diverted. Various routes for novel treatment of *C. difficile* infection and disease exist, including targeting the two primary toxins, TcdA and TcdB, and preventing sporulation or germination, thus removing the transmissible agent of the disease. It is also thought that another key area of infection, adhesion, can be targeted to combat *C. difficile*. Lipoproteins have been shown to play many roles in bacteria (22), including key roles in virulence (18, 19, 24), and are involved in bacterial adhesion (20, 21). As a result, lipoproteins of *C. difficile* and their biosynthesis present a novel area for research to identify new methods for treatment.

The role of lipoprotein signal peptidases in *C. difficile* has been investigated using strains ECF1 and ECF2 which are defective in the genes CD2597 (*lspA*) and CD1903 (*lspA2*) respectively (34-36). These two genes encode proteins LspA and LspA2 which are functional lipoprotein signal peptidases. It has been shown here that both proteins possess properties which identify them as lipoprotein signal peptidases including the presence of four trans-membrane helices in their tertiary structure and the localisation of five amino acid domains which are conserved across similar proteins in many bacteria (29, 30, 177). LspA is similar to functional homologs in other *Clostridia*, however LspA2 does not display similarity to any of the proteins analysed in this study, indicating that it may have been acquired from a different origin during the evolution of *C. difficile* and raising the possibility that it acts upon a subset of lipoproteins.

Strains ECF1 and ECF2 were subjected to phenotypic analysis which revealed a number of differences from and similarities to the 630 Δ *erm* wild type strain. The growth of all three strains is predominantly the same however there is an apparent difference in exponential growth whereby the wild type was observed to enter this phase slightly earlier, resulting in greater observable growth in the

middle of the phase. Examination of the protein profiles of these strains revealed that there was little effect on protein localisation by inactivation of LspA and LspA2. This suggests that a lack of signal peptide cleavage does not alter the localisation of lipoproteins on a cellular fraction level. Conversely, it has previously been shown that inactivation of Lgt causes loss of lipoproteins into the culture filtrate due to the lack of anchoring by acyl-glycerol moieties. Previous work demonstrated that ECF1 and ECF2 were more susceptible to the antimicrobial effects of malachite green when compared to the wild type (34-36). In this study, the effect of hydrogen peroxide on these strains was investigated. It was shown that both ECF1 and ECF2 were significantly less susceptible to this compound than the wild type, with ECF1 the least susceptible of the three. This effect may be due to modulation of lipoprotein functionality by the lack of signal peptide cleavage. Johnston *et al* (168) demonstrated that knocking out the PsaA lipoprotein of *S. pneumoniae* caused a large reduction in percent survival; complete loss of lipoproteins may have a greater effect than disrupted biosynthesis. Lipoproteins and lipoprotein biosynthesis may play different roles in the response of *C. difficile* to these two different stresses, thus the difference in the effects of mutation. For example, malachite green is known to have similar effects to the novel antimicrobial globomycin which directly targets lipoprotein signal peptidase (31), thus one response is directly related to biosynthesis while the other may be related to lipoprotein functionality.

Due to the demonstrated effects of globomycin upon lipoprotein signal peptidases and the use of globomycin protection assays in demonstrating the role of these proteins in bacteria (24, 177); and the potential for use of globomycin or homologs as a treatment, attempts were made to express *C. difficile* LspA and LspA2 in *E. coli*. To that end, various expression plasmids for each protein were generated using three different backbones; pQE80, pBAT4 and pET28a. The first two attempts were not successful and this is thought to be due to a combination of factors including the addition of a His-tag to the proteins, methods for modification of the unusual TTG start codon of both genes and a lack of accounting for rare codon usage in both genes. The final, and successful, attempt using pET28a aimed to overcome these difficulties using an altered cloning methodology and the *E. coli* Rosetta2 (DE3) expression strain

which holds pRARE2, a plasmid encoding tRNAs for rare codons. Expression of both proteins was induced by auto-induction and demonstrated by anti-His Western blotting; no expression was observable by SDS-PAGE alone and the level of expression is not as high as expected. The presence of isolated bands at the expected molecular weights on the Western blot indicates that the proteins were expressed correctly and demonstrate that no other effect was being observed. It was also demonstrated that both LspA and LspA2 were being localised to the surface layer of *E. coli* using a low pH glycine protein extraction, suggesting that the proteins are located suitably for functionality. Strains holding ECF1 and ECF2 are suitable for use in globomycin protection assays where expression of the proteins is driven by auto-induction; however the system may need optimization to ensure a detectable level of protection is achieved.

Some of the areas covered in this study may benefit from further work. Foremost is the use of globomycin protection assays for investigating the role of LspA and LspA2 in *C. difficile*; the tools have been generated but the assay needs conducting. There are further phenotypic assays that would also shed more light on the role of these two proteins. Expansion of the investigation into the effects of hydrogen peroxide upon *C. difficile* strains would be of benefit and should involve a more detailed assessment of the effect of hydrogen peroxide concentration over time. In addition to expansion on the assays performed here, implementation of adhesion assays using tissue culture cells would allow examination of the role of lipoprotein signal peptide cleavage in adhesion of *C. difficile* to host cells. A more detailed study of protein localisation within ECF1 and ECF2 in comparison to 630 Δ *erm* using mass spectrometry may increase clarity on the effect of mutation in *lspA* or *lspA2*. Not only could any changes in lipoprotein localisation be observed but the action of LspA or LspA2 on a given lipoprotein could also be examined, perhaps shedding light upon the differing roles of these two proteins in *C. difficile*.

Concluding Comments

Lipoproteins are thought to play a key role in the virulence and pathogenesis of *C. difficile*. Study of their function and of the effects of disrupting lipoprotein biosynthesis may shed light on this role and demonstrate potential for treatment opportunities. *C. difficile* encodes two functional lipoprotein signal peptidases: LspA and LspA2 which both possess the general properties of proteins of this type. Strains with mutations in *lspA* (ECF1) and *lspA2* (ECF2) have been used to demonstrate that lipoprotein biosynthesis has an important role in the response of *C. difficile* to malachite green and hydrogen peroxide, providing validation of the idea that this process is important in virulence and pathogenicity. Further, both LspA and LspA2 have been successfully expressed in *E. coli* using plasmids pEF14 and pEF15, paving the way for further investigation into the role of both proteins in *C. difficile* by globomycin protection assays.

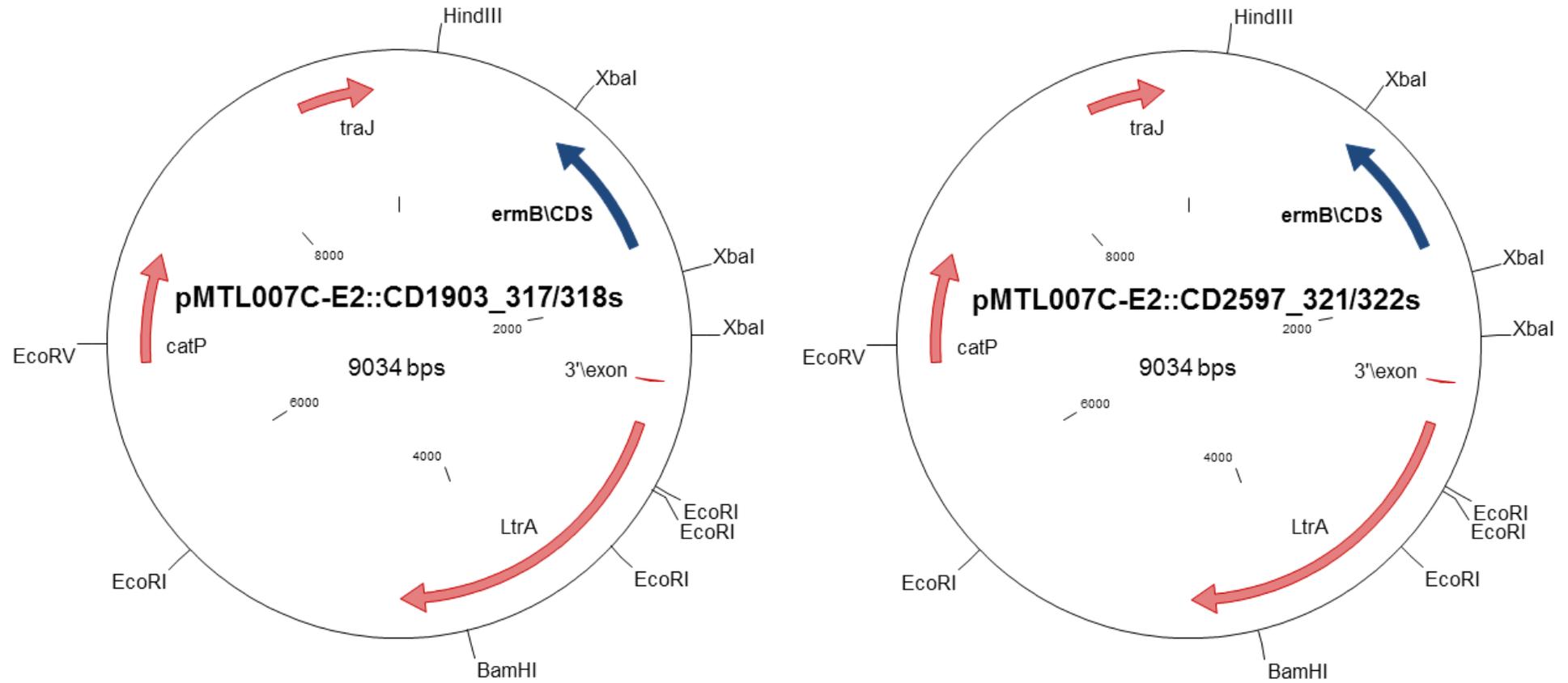
APPENDIX 2 Protein accession numbers for bacterial lipoprotein signal peptidases used in multiple alignments for the generation of **Figure 3.1 WebLogo of aligned lipoprotein signal peptidase from various bacteria, highlighting the conserved functional amino acids** and **Figure 3.2 Phylogenetic tree of lipoprotein signal peptidase genes from various bacteria**. A number of proteins have been removed from online databases or combined with similar entries thus these proteins are listed here with identical accession numbers.

Protein	Accession Number	Protein	Accession Number
Clostridium difficile LspA2	Q187L2	Myxococcus xanthus LspA2	Q1D5G3
Borrelia burgdorferi B31 LspA	O51425	Myxococcus xanthus LspA1	Q1D5E9
Buchnera aphidicola str. APS LspA	P57248	Myxococcus xanthus LspA3	Q1DFD1
Rickettsia bellii LspA	Q1RI47	Myxococcus xanthus LspA4	Q1DFD0
Rickettsia prowazekii LspA	Q9ZDC4	Brucella melitensis bv. 1 str. 16M LspA	Q8YES8
Rickettsia prowazekii str. Madrid E LspA	Q9ZDC4	Mesorhizobium loti MAFF303099 LspA	Q98GR1
Rickettsia typhi LspA	Q68WX1	Sinorhizobium meliloti 1021 LspA	Q92SJ3
Rickettsia Canadensis LspA	A8EZ20	Mycobacterium tuberculosis LspA	A5WMM6
Rickettsia felis LspA	Q4ULU0	Mycobacterium leprae TN LspA	Q9X7E7
Rickettsia akari LspA	A8GNC3	Streptomyces coelicolor A3(2) LspA	Q952X7
Rickettsia rickettsia LspA	H6Q3C1	Cupriavidus metallidurans CH34 LspA	Q1LJB7
Rickettsia conorii LspA	Q92I62	Ralstonia solanacearum GMI1000 LspA	Q8XWL5
Rickettsia conorii str. Malish 7 LspA	Q92I62	Neisseria meningitidis MC58 LspA	P65265

Caulobacter crescentus CB15 LspA	Q9AAA6	Neisseria meningitidis Z2491 LspA	P65264
Chlamydia muridarum Nigg LspA	Q9PJY8	Yersinia pestis LspA	Q8ZIL9
Chlamydia trachomatis LspA	O84413	Enterobacter aerogenes KCTC 2190 LspA	P13514
Chlamydia pneumonia LspA	Q9Z817	Klebsiella pneumonia LspA	Q9RF47
Campylobacter jejuni LspA	Q9PIE1	Escherichia coli K-12 LspA	P00804
Helicobacter pylori 26695 LspA	P25178	Salmonella enterica subsp. enterica serovar Typhi LspA	Q8Z9N1
Helicobacter pylori J99 LspA	Q9ZM23	Salmonella enterica subsp. enterica serovar Typhimurium str. LT2 LspA	Q8ZRY9
Mycoplasma genitalium G37 LspA	Q49401	Vibrio cholerae O1 biovar El Tor str. N16961 LspA	Q9KU46
Mycoplasma pneumonia M129 LspA	P75484	Haemophilus influenzae Rd KW20 LspA	P44975
Clostridium difficile LspA	Q182T8	Pasturella multocida subsp. multocida str. Pm70 LspA	P57959
Clostridium acetobutylicum ATCC 824 LspA	Q97H98	Pseudomonas aeruginosa PAO1 LspA	Q9HVM5
Clostridium botulinum Ba4 str 657 LspA	C3KVB1	Pseudomonas fluorescens LspA	P17942
Clostridium sporogenes ATCC 15579 LspA	J7T0X8	Xyella fastidiosa 9a5c LspA	Q9PAS8
Clostridium tetani E88 LspA	Q894D4	Serratia marescens LspA	O52213
Lactococcus lactis subsp. Cremoris MG1363 LspA	Q48729	Nostoc sp. PCC7120 (Anabaena sp. PCC 7120) LspA	Q8YNI8

Lactococcus lactis subsp. Lactis II1403 LspA	Q9CGU5	Synechocystis sp. PCC 6803 substr. Kazusa LspA	P73540
Streptococcus pyrogenes serotype M1 LspA	Q9A0D2	Deinococcus rediodurans R1 LspA	Q9RRU7
Bacillus halodurans C- 125 LspA	Q9K9V2	Aquifex aeolicus LspA	O67692
Bacillus subtilis LspA	Q45479	Treponema pallidum subsp. pallidum str. Nichols LspA	O83943
Staphylococcus aureus subsp. aureus N315 LspA	P65267	Thermotoga maritime LspA	Q9WYT4
Staphylococcus carnosus subsp. carnosus TM300 LspA	Q59835		

APPENDIX 3 *Clostron* plasmids used to generate *C. difficile* *lspA* and *lspA2* mutants



APPENDIX 4 Statistical analysis of the *C. difficile* growth curves in **Figure 4.2**
Growth curves of wild type and IspA mutant strains of *C. difficile*, using
 GraphPad Prism 6

Comparison of curve fitting models, selecting the Sigmoidal dose-response (variable slope) as the preferred model with no evidence of inadequate model.

	630 Δerm	$\Delta IspA$	$\Delta IspA2$
Comparison of Fits			
Simpler model	Exponential growth equation	Exponential growth equation	Exponential growth equation
Probability it is correct	<0.01%	<0.01%	<0.01%
Alternative model	Sigmoidal dose-response (variable slope)	Sigmoidal dose-response (variable slope)	Sigmoidal dose-response (variable slope)
Probability it is correct	>99.99%	>99.99%	>99.99%
Ratio of probabilities			
Preferred model	Sigmoidal dose-response (variable slope)	Sigmoidal dose-response (variable slope)	Sigmoidal dose-response (variable slope)
Difference in AICc	66.43	93.7	66.53
Exponential growth equation			
Best-fit values			
Y0	0.2241	0.1778	0.1868
k	0.152	0.1731	0.1686
Tau	6.577	5.777	5.931
Doubling Time	4.559	4.004	4.111
Std. Error			
Y0	0.04651	0.03904	0.04123
k	0.02327	0.02396	0.02423
95% Confidence Intervals			
Y0	0.1295 to 0.3187	0.09838 to 0.2572	0.1029 to 0.2706
k	0.1047 to 0.1994	0.1244 to 0.2218	0.1193 to 0.2179
Tau	5.016 to 9.552	4.508 to 8.042	4.590 to 8.382
Doubling Time	3.477 to 6.621	3.125 to 5.574	3.181 to 5.810
Goodness of Fit			
Degrees of Freedom	33	33	33
R square	0.672	0.7225	0.7054
Absolute Sum of Squares	2.028	1.744	1.862
Sy.x	0.2479	0.2299	0.2375
Replicates test for lack of fit			
SD replicates	0.1047	0.05657	0.1007
SD lack of fit	0.403	0.39	0.386
Discrepancy (F)	14.82	47.54	14.69
P value	< 0.0001	< 0.0001	< 0.0001
Evidence of inadequate model?	Yes	Yes	Yes

**Sigmoidal dose-response
(variable slope)**

Best-fit values			
Bottom	0.04743	0.04385	0.03507
Top	0.9886	1.007	0.9947
LogEC50	4.869	5.557	5.309
HillSlope	0.6482	0.6259	0.6066
EC50	73993	360709	203788
Std. Error			
Bottom	0.03129	0.01791	0.02865
Top	0.02725	0.01912	0.02844
LogEC50	0.1398	0.08802	0.1378
HillSlope	0.1196	0.06971	0.1031
95% Confidence Intervals			
Bottom	-0.01639 to 0.1113	0.007329 to 0.08036	-0.02337 to 0.09350
Top	0.9330 to 1.044	0.9684 to 1.046	0.9367 to 1.053
LogEC50	4.584 to 5.154	5.378 to 5.737	5.028 to 5.590
HillSlope	0.4043 to 0.8922	0.4837 to 0.7680	0.3962 to 0.8169
EC50	38379 to 142656	238582 to 545351	106695 to 389239
Goodness of Fit			
Degrees of Freedom	31	31	31
R square	0.9578	0.9836	0.9621
Absolute Sum of Squares	0.2612	0.1031	0.2392
Sy.x	0.09179	0.05766	0.08784
Replicates test for lack of fit			
SD replicates	0.1047	0.05657	0.1007
SD lack of fit	0.04724	0.06024	0.04235
Discrepancy (F)	0.2036	1.134	0.1769
P value	0.991	0.3813	0.9946
Evidence of inadequate model?	No	No	No
Number of points			
Analyzed	35	35	35

Applying the Sigmoidal dose-response (variable slope) curve fitting model to the data sets for all three strains, analysis was conducted to test the suitability of one curve for all data sets, revealing that there is enough difference between the data sets to require a different curve for each.

	630 Δerm	$\Delta lspA$	$\Delta lspA2$	Global (shared)
Comparison of Fits				
Null hypothesis				One curve for all data sets
Alternative hypothesis				Different curve for each data set
P value				0.0063
Conclusion (alpha = 0.05)				Reject null hypothesis
Preferred model				Different curve for each data set
F (DFn, DFd)				2.894 (8,93)
Different curve for each data set				
Best-fit values				
Bottom	0.04743	0.04385	0.03507	
Top	0.9886	1.007	0.9947	
LogEC50	4.869	5.557	5.309	
HillSlope	0.6482	0.6259	0.6066	
EC50	73993	360709	203788	
Std. Error				
Bottom	0.03129	0.01791	0.02865	
Top	0.02725	0.01912	0.02844	
LogEC50	0.1398	0.08802	0.1378	
HillSlope	0.1196	0.06971	0.1031	
95% Confidence Intervals				
Bottom	-0.01639 to 0.1113	0.007329 to 0.08036	-0.02337 to 0.09350	
Top	0.9330 to 1.044	0.9684 to 1.046	0.9367 to 1.053	
LogEC50	4.584 to 5.154	5.378 to 5.737	5.028 to 5.590	
HillSlope	0.4043 to 0.8922	0.4837 to 0.7680	0.3962 to 0.8169	
EC50	38379 to 142656	238582 to 545351	106695 to 389239	
Goodness of Fit				
Degrees of Freedom	31	31	31	
R square	0.9578	0.9836	0.9621	
Absolute Sum of Squares	0.2612	0.1031	0.2392	
Sy.x	0.09179	0.05766	0.08784	
One curve for all data sets				
Best-fit values				
Bottom	0.0411	0.0411	0.0411	0.0411

Top	0.9962	0.9962	0.9962	0.9962
LogEC50	5.238	5.238	5.238	5.238
HillSlope	0.608	0.608	0.608	0.608
EC50	172825	172825	172825	172825
Std. Error				
Bottom	0.01643	0.01643	0.01643	0.01643
Top	0.01596	0.01596	0.01596	0.01596
LogEC50	0.07845	0.07845	0.07845	0.07845
HillSlope	0.05907	0.05907	0.05907	0.05907
95% Confidence Intervals				
Bottom	0.008495 to 0.07370	0.008495 to 0.07370	0.008495 to 0.07370	0.008495 to 0.07370
Top	0.9645 to 1.028	0.9645 to 1.028	0.9645 to 1.028	0.9645 to 1.028
LogEC50	5.082 to 5.393	5.082 to 5.393	5.082 to 5.393	5.082 to 5.393
HillSlope	0.4908 to 0.7251	0.4908 to 0.7251	0.4908 to 0.7251	0.4908 to 0.7251
EC50	120775 to 247306	120775 to 247306	120775 to 247306	120775 to 247306
Goodness of Fit				
Degrees of Freedom				101
R square	0.9432	0.9748	0.9613	0.96
Absolute Sum of Squares	0.351	0.1581	0.2446	0.7537
Sy.x				0.08639
Constraints				
Bottom	Bottom is shared	Bottom is shared	Bottom is shared	
Top	Top is shared	Top is shared	Top is shared	
LogEC50	LogEC50 is shared	LogEC50 is shared	LogEC50 is shared	
HillSlope	HillSlope is shared	HillSlope is shared	HillSlope is shared	
Number of points				
Analyzed	35	35	35	

To identify which data points are different enough from the others for the particular time point to cause a different curve to be required for each data set, a two-way ANOVA with Tukey's multiple comparisons test was performed, comparing the mean of values for each strain at a given time point with the means of the other strains. This test indicates that at 5h and 6h (Rows 6 and 7), 630 Δ *erm* is significantly different to ECF1 (Δ /*spA*) but not ECF2 (Δ /*spA2*).

Within each row,
compare columns
(simple effects
within rows)

Number of families 11
Number of comparisons per family 3
Alpha 0.05

Tukey's multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary	Adjusted P Value	Adjusted P Value
Row 1						
630 Δ <i>erm</i> vs. Δ / <i>spA</i>	0	-0.1762 to 0.1762	No	ns	> 0.9999	> 0.9999
630 Δ <i>erm</i> vs. Δ / <i>spA2</i>	0.003333	-0.1729 to 0.1795	No	ns	0.9989	0.9989
Δ / <i>spA</i> vs. Δ / <i>spA2</i>	0.003333	-0.1729 to 0.1795	No	ns	0.9989	0.9989
Row 2						
630 Δ <i>erm</i> vs. Δ / <i>spA</i>	0.006667	-0.1695 to 0.1829	No	ns	0.9955	0.9955
630 Δ <i>erm</i> vs. Δ / <i>spA2</i>	0.01	-0.1662 to 0.1862	No	ns	0.9898	0.9898
Δ / <i>spA</i> vs. Δ / <i>spA2</i>	0.003333	-0.1729 to 0.1795	No	ns	0.9989	0.9989
Row 3						
630 Δ <i>erm</i> vs. Δ / <i>spA</i>	0.02	-0.1562 to 0.1962	No	ns	0.9600	0.9600
630 Δ <i>erm</i> vs. Δ / <i>spA2</i>	0.01667	-0.1595 to 0.1929	No	ns	0.9721	0.9721
Δ / <i>spA</i> vs. Δ / <i>spA2</i>	-0.003333	-0.1795 to 0.1729	No	ns	0.9989	0.9989
Row 4						
630 Δ <i>erm</i> vs. Δ / <i>spA</i>	0.04	-0.1362 to 0.2162	No	ns	0.8497	0.8497
630 Δ <i>erm</i> vs. Δ / <i>spA2</i>	0.04	-0.1362 to 0.2162	No	ns	0.8497	0.8497
Δ / <i>spA</i> vs. Δ / <i>spA2</i>	2.48E-09	-0.1762 to 0.1762	No	ns	> 0.9999	> 0.9999
Row 5						
630 Δ <i>erm</i> vs. Δ / <i>spA</i>	0.1067	-0.06953 to 0.2829	No	ns	0.3208	0.3208

630 Δerm vs. $\Delta IspA2$	0.09667	-0.07953 to 0.2729	No	ns	0.3917	0.3917		
$\Delta IspA$ vs. $\Delta IspA2$	-0.01	-0.1862 to 0.1662	No	ns	0.9898	0.9898		
Row 6								
630 Δerm vs. $\Delta IspA$	0.1967	0.02047 to 0.3729	Yes	*	0.0251	0.0251		
630 Δerm vs. $\Delta IspA2$	0.1167	-0.05953 to 0.2929	No	ns	0.2581	0.2581		
$\Delta IspA$ vs. $\Delta IspA2$	-0.08	-0.2562 to 0.09619	No	ns	0.5244	0.5244		
Row 7								
630 Δerm vs. $\Delta IspA$	0.2	0.02381 to 0.3762	Yes	*	0.0223	0.0223		
630 Δerm vs. $\Delta IspA2$	0.1467	-0.02953 to 0.3229	No	ns	0.1212	0.1212		
$\Delta IspA$ vs. $\Delta IspA2$	-0.05333	-0.2295 to 0.1229	No	ns	0.7491	0.7491		
Row 8								
630 Δerm vs. $\Delta IspA$	0.06667	-0.1095 to 0.2429	No	ns	0.6377	0.6377		
630 Δerm vs. $\Delta IspA2$	0.04667	-0.1295 to 0.2229	No	ns	0.8014	0.8014		
$\Delta IspA$ vs. $\Delta IspA2$	-0.02	-0.1962 to 0.1562	No	ns	0.9600	0.9600		
Row 9								
630 Δerm vs. $\Delta IspA$	-0.04	-0.2162 to 0.1362	No	ns	0.8497	0.8497		
630 Δerm vs. $\Delta IspA2$	0.006667	-0.1829 to 0.1695	No	ns	0.9955	0.9955		
$\Delta IspA$ vs. $\Delta IspA2$	0.03333	-0.1429 to 0.2095	No	ns	0.8930	0.8930		
Row 10								
630 Δerm vs. $\Delta IspA$	-0.02333	-0.1995 to 0.1529	No	ns	0.9460	0.9460		
630 Δerm vs. $\Delta IspA2$	0.003333	-0.1795 to 0.1729	No	ns	0.9989	0.9989		
$\Delta IspA$ vs. $\Delta IspA2$	0.02	-0.1562 to 0.1962	No	ns	0.9600	0.9600		
Row 11								
630 Δerm vs. $\Delta IspA$	0	-0.1762 to 0.1762	No	ns	> 0.9999	> 0.9999		
630 Δerm vs. $\Delta IspA2$	0	-0.1762 to 0.1762	No	ns	> 0.9999	> 0.9999		
$\Delta IspA$ vs. $\Delta IspA2$	0	-0.1762 to 0.1762	No	ns	> 0.9999	> 0.9999		

Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	N1	N2	q	DF
Row 1								
630 Δerm vs. $\Delta IspA$	0.03667	0.03667	0	0.07348	3	3	0	66
630 Δerm vs. $\Delta IspA2$	0.03667	0.03333	0.003333	0.07348	3	3	0.06415	66

Δ/spA vs. $\Delta/spA2$	0.03667	0.03333	0.003333	0.07348	3	3	0.06415	66
Row 2								
630 Δerm vs. Δ/spA	0.04	0.03333	0.006667	0.07348	3	3	0.1283	66
630 Δerm vs. $\Delta/spA2$	0.04	0.03	0.01	0.07348	3	3	0.1925	66
Δ/spA vs. $\Delta/spA2$	0.03333	0.03	0.003333	0.07348	3	3	0.06415	66
Row 3								
630 Δerm vs. Δ/spA	0.06333	0.04333	0.02	0.07348	3	3	0.3849	66
630 Δerm vs. $\Delta/spA2$	0.06333	0.04667	0.01667	0.07348	3	3	0.3208	66
Δ/spA vs. $\Delta/spA2$	0.04333	0.04667	-0.003333	0.07348	3	3	0.06415	66
Row 4								
630 Δerm vs. Δ/spA	0.12	0.08	0.04	0.07348	3	3	0.7698	66
630 Δerm vs. $\Delta/spA2$	0.12	0.08	0.04	0.07348	3	3	0.7698	66
Δ/spA vs. $\Delta/spA2$	0.08	0.08	2.484E-09	0.07348	3	3	4.78E-08	66
Row 5								
630 Δerm vs. Δ/spA	0.26	0.1533	0.1067	0.07348	3	3	2.053	66
630 Δerm vs. $\Delta/spA2$	0.26	0.1633	0.09667	0.07348	3	3	1.86	66
Δ/spA vs. $\Delta/spA2$	0.1533	0.1633	-0.01	0.07348	3	3	0.1925	66
Row 6								
630 Δerm vs. Δ/spA	0.5433	0.3467	0.1967	0.07348	3	3	3.785	66
630 Δerm vs. $\Delta/spA2$	0.5433	0.4267	0.1167	0.07348	3	3	2.245	66
Δ/spA vs. $\Delta/spA2$	0.3467	0.4267	-0.08	0.07348	3	3	1.54	66
Row 7								
630 Δerm vs. Δ/spA	0.85	0.65	0.2	0.07348	3	3	3.849	66
630 Δerm vs. $\Delta/spA2$	0.85	0.7033	0.1467	0.07348	3	3	2.823	66
Δ/spA vs. $\Delta/spA2$	0.65	0.7033	-0.05333	0.07348	3	3	1.026	66
Row 8								
630 Δerm vs. Δ/spA	0.9767	0.91	0.06667	0.07348	3	3	1.283	66
630 Δerm vs. $\Delta/spA2$	0.9767	0.93	0.04667	0.07348	3	3	0.8981	66
Δ/spA vs. $\Delta/spA2$	0.91	0.93	-0.02	0.07348	3	3	0.3849	66
Row 9								
630 Δerm vs. Δ/spA	1	1.04	-0.04	0.07348	3	3	0.7698	66
630 Δerm vs. $\Delta/spA2$	1	1.007	-0.006667	0.07348	3	3	0.1283	66
Δ/spA vs. $\Delta/spA2$	1.04	1.007	0.03333	0.07348	3	3	0.6415	66
Row 10								

630 Δerm vs. $\Delta lspA$	1.003	1.027	-0.02333	0.07348	3	3	0.4491	66
630 Δerm vs. $\Delta lspA2$	1.003	1.007	-0.003333	0.07348	3	3	0.06415	66
$\Delta lspA$ vs. $\Delta lspA2$	1.027	1.007	0.02	0.07348	3	3	0.3849	66
Row 11								
630 Δerm vs. $\Delta lspA$	0.98	0.98	0	0.07348	3	3	0	66
630 Δerm vs. $\Delta lspA2$	0.98	0.98	0	0.07348	3	3	0	66
$\Delta lspA$ vs. $\Delta lspA2$	0.98	0.98	0	0.07348	3	3	0	66

APPENDIX 5 Statistical analysis of *C. difficile* survival under hydrogen peroxide stress using GraphPad Prism 6

Statistical analysis for **Figure 4.4 Effect of hydrogen peroxide concentration on survival of *C. difficile* strains**, using GraphPad Prism 6 two-way ANOVA followed by Tukey's multiple comparisons test to compare the differences between the survival of each strain at each concentration.

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	4.365	< 0.0001	****	Yes
Time	81.34	< 0.0001	****	Yes
Column Factor	0.3871	0.1692	ns	No
Subjects (matching)	2.426	0.6721	ns	No

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	11329	8	1416	F (8, 96) = 4.561	P < 0.0001
Time	211078	4	52770	F (4, 96) = 170.0	P < 0.0001
Column Factor	1005	2	502.3	F (2, 24) = 1.915	P = 0.1692
Subjects (matching)	6296	24	262.3	F (24, 96) = 0.8450	P = 0.6721
Residual	29804	96	310.5		

Number of missing values	0
--------------------------	---

Within each row, compare columns (simple effects within rows)

Number of families	5
Number of comparisons per family	3
Alpha	0.05

Tukey's multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary	Adjusted P Value
0					
630 Derm vs. DlspA	12.72	-6.688 to 32.12	No	ns	0.2692
630 Derm vs. DlspA2	10.56	-8.841 to 29.97	No	ns	0.4026
DlspA vs. DlspA2	-2.153	-21.56 to 17.25	No	ns	0.9625
0.01					
630 Derm vs. DlspA	5.289	-14.11 to 24.69	No	ns	0.7945
630 Derm vs. DlspA2	-18.31	-37.72 to 1.090	No	ns	0.0687
DlspA vs. DlspA2	-23.6	-43.01 to -4.199	Yes	*	0.0127

0.05					
630 Derm vs. DlspA	22.63	3.228 to 42.03	Yes	*	0.0178
630 Derm vs. DlspA2	19.29	-0.1147 to 38.69	No	ns	0.0517
DlspA vs. DlspA2	-3.342	-22.75 to 16.06	No	ns	0.9121

0.1					
630 Derm vs. DlspA	-29.76	-49.16 to -10.35	Yes	**	0.0012
630 Derm vs. DlspA2	-33.44	-52.85 to -14.04	Yes	***	0.0002
DlspA vs. DlspA2	-3.684	-23.09 to 15.72	No	ns	0.8942

1					
630 Derm vs. DlspA	0.006667	-19.40 to 19.41	No	ns	> 0.9999
630 Derm vs. DlspA2	-0.01	-19.41 to 19.39	No	ns	> 0.9999
DlspA vs. DlspA2	-0.01667	-19.42 to 19.39	No	ns	> 0.9999

Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	N1	N2	q	DF
0								
630 Derm vs. DlspA	104.2	91.45	12.72	8.176	9	9	2.199	120
630 Derm vs. DlspA2	104.2	93.61	10.56	8.176	9	9	1.827	120
DlspA vs. DlspA2	91.45	93.61	-2.153	8.176	9	9	0.3725	120
0.01								
630 Derm vs. DlspA	93.06	87.77	5.289	8.176	9	9	0.9148	120
630 Derm vs. DlspA2	93.06	111.4	-18.31	8.176	9	9	3.168	120
DlspA vs. DlspA2	87.77	111.4	-23.6	8.176	9	9	4.082	120
0.05								
630 Derm vs. DlspA	105.1	82.44	22.63	8.176	9	9	3.914	120
630 Derm vs. DlspA2	105.1	85.78	19.29	8.176	9	9	3.336	120
DlspA vs. DlspA2	82.44	85.78	-3.342	8.176	9	9	0.5781	120
0.1								
630 Derm vs. DlspA	14.25	44.01	-29.76	8.176	9	9	5.147	120
630 Derm vs. DlspA2	14.25	47.7	-33.44	8.176	9	9	5.784	120
DlspA vs. DlspA2	44.01	47.7	-3.684	8.176	9	9	0.6373	120
1								
630 Derm vs. DlspA	0.006667	0	0.006667	8.176	9	9	0.00115	120
630 Derm vs. DlspA2	0.006667	0.01667	-0.01	8.176	9	9	0.00173	120
DlspA vs. DlspA2	0	0.01667	-0.01667	8.176	9	9	0.00288	120

Statistical analysis for **Figure 4.5 Further investigation of the effect of hydrogen peroxide concentration on survival of *C. difficile* strains**, using GraphPad Prism 6 two-way ANOVA followed by Tukey's multiple comparisons test to compare the differences between the survival of each strain at each concentration.

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	8.204	< 0.0001	****	Yes
Time	26.34	< 0.0001	****	Yes
Column Factor	0.575	0.4684	ns	No
Subjects (matching)	29.28	0.0045	**	Yes

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	117770	4	29443	F (4, 156) = 8.987	P < 0.0001
Time	378133	2	189067	F (2, 156) = 57.71	P < 0.0001
Column Factor	8254	2	4127	F (2, 78) = 0.7658	P = 0.4684
Subjects (matching)	420362	78	5389	F (78, 156) = 1.645	P = 0.0045
Residual	511062	156	3276		

Within each row, compare columns (simple effects within rows)

Number of families	3
Number of comparisons per family	3
Alpha	0.05

Tukey's multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary	Adjusted P Value
0.1					
630 Δerm vs. $\Delta lspA$	60.57	20.06 to 101.1	Yes	**	0.0015
630 Δerm vs. $\Delta lspA2$	34.65	-5.848 to 75.16	No	ns	0.1101
$\Delta lspA$ vs. $\Delta lspA2$	-25.91	-66.41 to 14.59	No	ns	0.2885
0.5					
630 Δerm vs. $\Delta lspA$	-65.06	-105.6 to -24.56	Yes	***	0.0006
630 Δerm vs. $\Delta lspA2$	-0.02185	-40.52 to 40.48	No	ns	> 0.9999
$\Delta lspA$ vs. $\Delta lspA2$	65.04	24.53 to 105.5	Yes	***	0.0006
1					
630 Δerm vs. $\Delta lspA$	0.01889	-40.48 to 40.52	No	ns	> 0.9999
630 Δerm vs. $\Delta lspA2$	0.01852	-40.48 to 40.52	No	ns	> 0.9999
$\Delta lspA$ vs. $\Delta lspA2$	-0.00037	-40.50 to 40.50	No	ns	> 0.9999

Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	N1	N2	q	DF
0.1								
630 Δerm vs. Δ/spA	124.1	63.57	60.57	17.17	27	27	4.988	234
630 Δerm vs. $\Delta/spA2$	124.1	89.48	34.65	17.17	27	27	2.854	234
Δ/spA vs. $\Delta/spA2$	63.57	89.48	-25.91	17.17	27	27	2.134	234
0.5								
630 Δerm vs. Δ/spA	0.002222	65.06	-65.06	17.17	27	27	5.358	234
630 Δerm vs. $\Delta/spA2$	0.002222	0.02407	-0.02185	17.17	27	27	0.0018	234
Δ/spA vs. $\Delta/spA2$	65.06	0.02407	65.04	17.17	27	27	5.356	234
1								
630 Δerm vs. Δ/spA	0.01889	0	0.01889	17.17	27	27	0.001556	234
630 Δerm vs. $\Delta/spA2$	0.01889	0.0003704	0.01852	17.17	27	27	0.001525	234
Δ/spA vs. $\Delta/spA2$	0	0.0003704	-0.0003704	17.17	27	27	3.05E-05	234

Statistical analysis for **Figure 4.6 Effect of incubation time with hydrogen peroxide on survival of *C. difficile* strains**, using GraphPad Prism 6 two-way ANOVA followed by Sidak's multiple comparisons test to compare the difference in survival of each strain after the two time periods.

Source of Variation	% of total variation	P value	P value summary	Significant?
Interaction	2.764	0.3811	ns	No
Time	28.44	0.0001	***	Yes
Column Factor	2.764	0.3811	ns	No
Subjects (matching)	33.02	0.5	ns	No

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	0.000004466	2	0.000002233	F (2, 24) = 1.005	P = 0.3811
Time	0.00004594	1	0.00004594	F (1, 24) = 20.67	P = 0.0001
Column Factor	0.000004466	2	0.000002233	F (2, 24) = 1.005	P = 0.3811
Subjects (matching)	0.00005335	24	0.000002223	F (24, 24) = 1.000	P = 0.5000
Residual	0.00005335	24	0.000002223		

Number of missing values	0
--------------------------	---

Compare each cell mean with the other cell mean in that column.

Number of families	1
Number of comparisons per family	3
Alpha	0.05

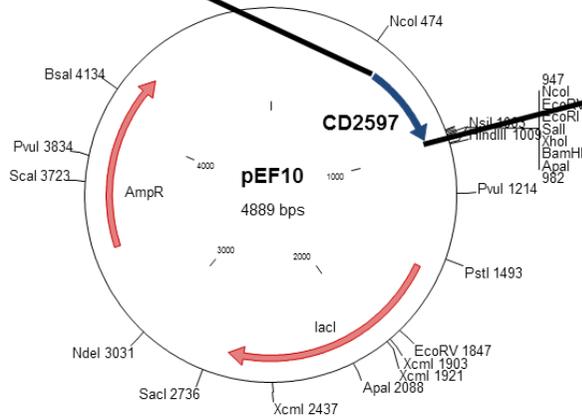
Sidak's multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary	Adjusted P Value
30 - 45					
630 Δ erm	-0.001057	-0.002860 to 0.0007466	No	ns	0.3764
Δ /spA	-0.002414	-0.004217 to - 0.0006103	Yes	**	0.0065
Δ /spA2	-0.002063	-0.003867 to - 0.0002600	Yes	*	0.0215

Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	N1	N2	t	DF
30 - 45								
630 Δ erm	0	0.001057	-0.001057	0.0007028	9	9	1.504	24
Δ /spA	0	0.002414	-0.002414	0.0007028	9	9	3.434	24
Δ /spA2	0	0.002063	-0.002063	0.0007028	9	9	2.936	24

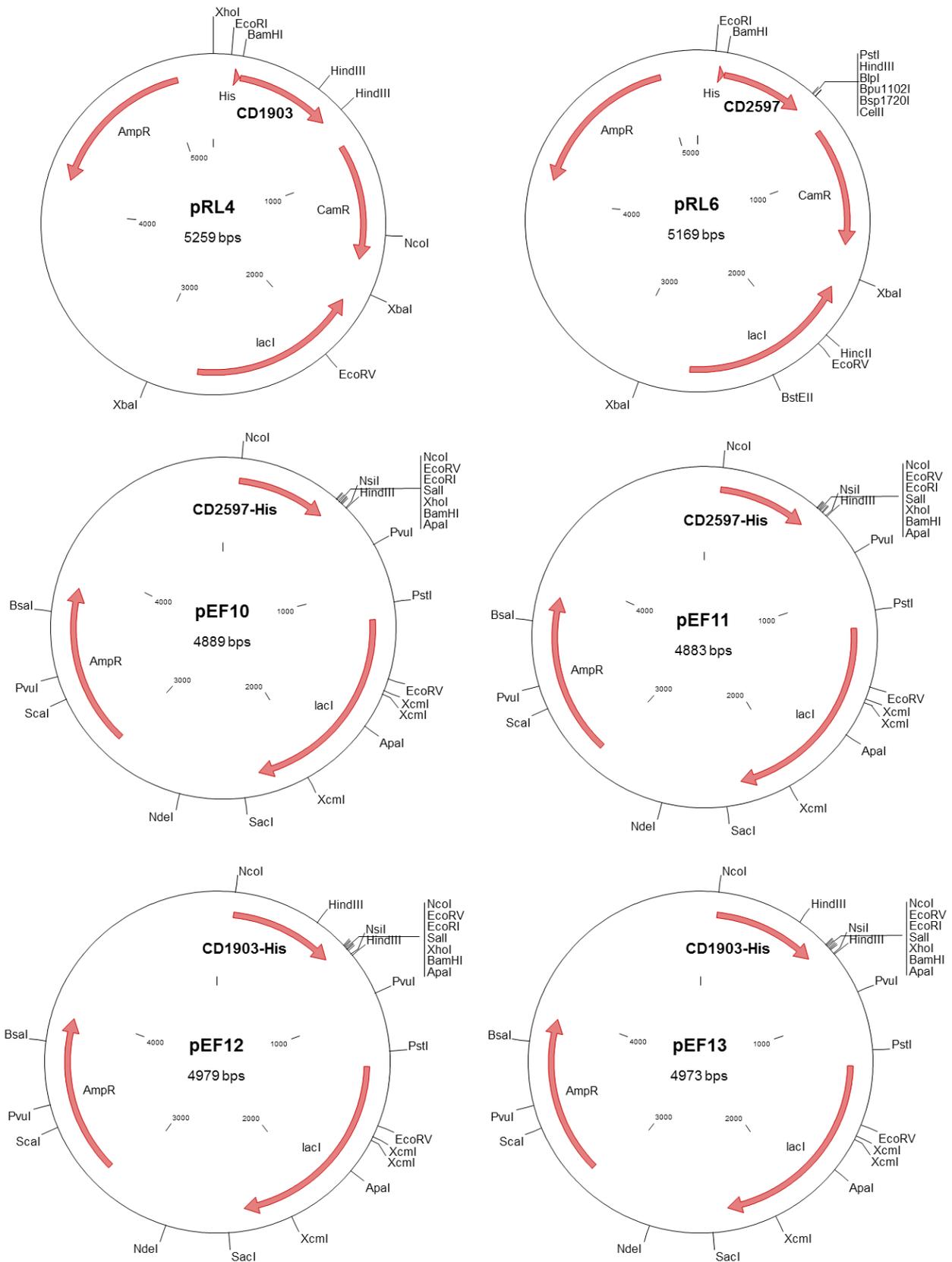
APPENDIX 6 Cloning strategy for expression of *C. difficile* *LspA* and *LspA2* in *E. coli* using vectors *pQE80* and *pBAT4*

To allow for detection of expressed protein the addition of a His tag was required (5' end for *pQE80* and 3' end for *pBAT4*), and amplification primers were designed to have this effect. Due to the uncommon TTG start codon for both *LspA* and *LspA2*, the cloning strategy for *pBAT4* was to introduce an ATG start codon via forced cloning with the amplification primers plus an *Nco1* restriction site (C^CATGG). Two methods were used: the deletion of two bases to generate a different first pair (*pEF11* and *pEF13*) or the insertion of 10 bases to extend the start of the gene (*pEF10* and *pEF12*).

DNA	Sequence	bp	aa	Mol Weight
CD2597	G A A A A A T C A T T G C T A T A T A T T A A T A A T A . . . A G T A A A A G T A G G T G A E K S L L Y I L I I . . . S K S R -	446	148	17002
pEF10	T C C A T G G C T A T G C T A T A T A T A T T A A T A A T A . . . A G T A A A A G T A G G C A C C A C C A C C A C C A C T A A S M A M L Y I L I I . . . S K S R H H H H H H -	470	156	18045.2
pEF11	G A T A T A T C C A T G G C T T A T A T A T T A A T A A T A . . . A G T A A A A G T A G G C A C C A C C A C C A C C A C T A A D I S M A Y I L I I . . . S K S R H H H H H H -	464	154	17800.8
pRL6	A A A T T A A C T A T G A G A G G A T C G C A T C A C C A T C A C C A T C A C G G A T C C T G G C T A T A T A T T A A T A A T A . . . A G T A A A A G T A G G T G A K L T M R G S H H H H H H G S L L Y I L I I . . . S K S R -	480	160	18400.5
CD1903	G A A T G A C T T T T G C A A G G A G G T G T T A A T A T C . . . C A G T T T A A A A G A T A A E - L L Q G G V N I . . . Q F H R -	536	178	20400.9
pEF12	T C C A T G G C T A T G C A A G G A G G T G T T A A T A T C . . . C A G T T T A A A A G A C A C C A C C A C C A C C A C T A A S M A M Q G G V N I . . . Q F K R H H H H H H -	560	186	21444.1
pEF13	T C C A T G G C T G G A G G T G T T A A T A T C . . . C A G T T T A A A A G A C A C C A C C A C C A C C A C T A A S M A G G V N I . . . Q F K R H H H H H H -	546	183	20982.5
pRL4	A A A T T A A C T A T G A G A G G A T C G C A T C A C C A T C A C C A T C A C G G A T C C T T G C A A G G A G G T G T T C A T A T C . . . C A G T T T A A A A G A T A A K L T M R G S H H H H H H G S L Q G G V N I . . . Q F K R -	572	190	21799.4



APPENDIX 7 Plasmids generated for expression of *C. difficile* *LspA* and *LspA2* in *E. coli* that did not result in detectable expression



APPENDIX 8 DNA sequencing data analysis for pEF10

DNA samples of pBAT4-derived expression plasmids were sent for DNA sequencing to ascertain the cause of a lack of protein expression. Only the sample for pEF10 returned useable sequence results which were combined into the data labelled Sample below. While the CD2597 sequence was correct (bold face with start and stop codons in capitals), there were many potential mutations in both the region after CD2597 and in *lacI* (all capitals), possibly explaining the lack of expression.

pEF10	1	taatacgaactcactataggggaattgtgagcggataacaattcctctagaaaataatttgtttaactttaagaaaggagatataatcc ATGgc
Consensus	428
pEF10	91	taatgctataatataataataaattctactcataggttagaccaactgtctaaaataatgggtattgaaataattgggtggaagtatcaac
Consensus	518
pEF10	181	aataccaataataataatgtaattcatttaacttatgtogaaaaatagaggtgcagcatttggattattacaaaataatcaatggatatt
Consensus	608
pEF10	271	tataattgttgcaattacttgcaacagttatggactatactatcttaatacaaggaaagtaacataattggaaaggtgggaattatatt
Consensus	698
pEF10	361	aattatactgggtgcaatgggaaatcctaattgataagtagacgattaggtttgttagtagattacttcgactttagaattatagggaaata
Consensus	788
pEF10	451	tgatataatagctgataatgttagttaggaactgtgtttttagtataatgttttattttttaaagtaaagtaggaccca
Consensus	878
pEF10	541	ccaccaccaccac TAA ccatggatatacgaattcgtgacactcgagggatccgggcccctcagatgcggccgatgcataagcttgagtat
Consensus	968
pEF10	631	tctatagtgacacataatccagcttgatcoggtgtgtaacaaagcccgaagggaagctgagttggctgctgaccgctgagcaataa
Consensus	1058
pEF10	721	ctagcataacccttggggcctctaaacgggtcttgagggttttttctgctgaaaggaggaaactataatccggataaccctggcgtaatagcg
Consensus	1148
pEF10	811	aagaggcccgcaccgatcgccctcccaacagttgocgacgctgaatggcgaaatggaaattgtaaacgtaataattttgttaaaattcgc
Consensus	1238
pEF10	901	gttaaatatttggtaaatcagctcatttttaaccaataggccgaaatcggcaaaatcccttataaaatcaaaagaatagaccgagatagg
Consensus	1328
pEF10	991	gttgagtgtgttccagtttggaacaagagtcacatttaaagaacgtggaactccacgtaaaaggcgaaaaaccgctctatcagggcgga
Consensus	1418
pEF10	1081	tggccactacagcttgcatgctcaggtcoggaagcataaagttaaaagcctgggtgcctaagagtgagctaacctacattaattgc
Consensus	1508
pEF10	1171	gttgcgc TCAC TC CCCGCTTCCAGTCGGAAACCTGTCGTGCCAGCTG-CATTAAATGAATCGCCAACGC CGGGAGAGCGGTT TGC
Consensus	1598
pEF10	1260	GTATTGGCCGCCAGGTGGTTTCTTTTCCACAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCTTGCCCTGAGAGAGTTGCAGC
Consensus	1688
pEF10	1350	AAGCGTCCACGCTGGTTGCCCCAGCGCCGAAATCCGTGTTGATGGTTCACGGGGGATATAACAAGACCTGCTTCGGTATCG
Consensus	1778

APPENDIX 9 Rare Codon Caltor analysis of CD2597 and CD1903 genes as found in plasmids pEF10, pEF11, pEF12, pEF13, pEF14 and pEF15, showing considerable rare codon usage in all cases.

CD2597 in pEF10

ATG GCT ATG **CUA** TAT **AUA** TTA **AUA** **AUA** ATT **CUA** CTC **AUA** GGT
 TTA GAC CAA CTG TCT AAA **AUA** TGG GTA TTG AAT AAT TTG GTG
 GAT GTA TCA ACA **AUA** CCA **AUA** **AUA** AAT AAT GTA TTT CAT TTA
 ACT TAT GTC GAA AAT **AGA** GGT GCA GCA TTT **GGA** TTA TTA CAA
 AAT AAT CAA TGG **AUA** TTT **AUA** ATT GTT GCA TTA CTT GCA ACA
 GTA TTT **GGA** **CUA** TAC TAT CTT AAT ACA **AGG** AAA GTA CAT **AUA**
 TTT **GGA** **AGG** TTG **GGA** ATT **AUA** TTA ATT **AUA** TCT GGT GCA TTG
GGA AAT **CUA** ATT GAT **AGA** GTA **CGA** TTA GGT TTT GTA GTA GAT
 TAC TTC GAC TTT **AGA** ATT **AUA** TGG GAA TAT GTA TTC AAT **AUA**
 GCT GAT GTA TTT GTA GTT GTA **GGA** ACT GTG TTT TTA TGT **AUA**
 TAT GTT TTA TTT TTT GAA AGT AAA AGT **AGG** CAC CAC CAC CAC
 CAC CAC TAA

Amino Acid	Rare Codon	Frequency of Occurrence
Arginine	CGA	1
	CGG	0
	AGG	3
	AGA	3
Glycine	GGA	6
	GGG	0
Isoleucine	AUA	16
Leucine	CUA	4
Proline	CCC	0
Threonine	ACG	0

CD2597 in pEF11

ATG GCT TAT **AUA** TTA **AUA** **AUA** ATT **CUA** CTC **AUA** GGT TTA GAC
 CAA CTG TCT AAA **AUA** TGG GTA TTG AAT AAT TTG GTG GAT GTA
 TCA ACA **AUA** CCA **AUA** **AUA** AAT AAT GTA TTT CAT TTA ACT TAT
 GTC GAA AAT **AGA** GGT GCA GCA TTT **GGA** TTA TTA CAA AAT AAT
 CAA TGG **AUA** TTT **AUA** ATT GTT GCA TTA CTT GCA ACA GTA TTT
GGA **CUA** TAC TAT CTT AAT ACA **AGG** AAA GTA CAT **AUA** TTT **GGA**
AGG TTG **GGA** ATT **AUA** TTA ATT **AUA** TCT GGT GCA TTG **GGA** AAT
CUA ATT GAT **AGA** GTA **CGA** TTA GGT TTT GTA GTA GAT TAC TTC
 GAC TTT **AGA** ATT **AUA** TGG GAA TAT GTA TTC AAT **AUA** GCT GAT
 GTA TTT GTA GTT GTA **GGA** ACT GTG TTT TTA TGT **AUA** TAT GTT
 TTA TTT TTT GAA AGT AAA AGT **AGG** CAC CAC CAC CAC CAC CAC
 TAA

Amino Acid	Rare Codon	Frequency of Occurrence
Arginine	CGA	1
	CGG	0
	AGG	3
	AGA	3
Glycine	GGA	6
	GGG	0
Isoleucine	AUA	16
Leucine	CUA	3
Proline	CCC	0
Threonine	ACG	0

CD1903 in pEF12

ATG CAA **GGA** GGT GTT AAT ATC **AGG** CAA GTA AAG TCT TTT GTT
 TTT CCA GTT **AUA** TCC TTA **AUA** TTC TTA GAC CAA ATT AGC AAA
 GTT CTT **AUA** **GGA** TTA TTC TTA ATG GAC TTT GAA ATT GAT **AUA**
 ATT **GGG** AAA TTT TTA **AGA** TTC AAT CCT GTT CAA AAT ACA AAT
CUA TCT TAT **GGG** **GGA** AAC TTT ATT GGT ATT **CUA** TCT AAT TTA
 TGG GTA TTG GTA TTG TTT AAC ATT TTA GTT **AUA** TTA GTT ATT
AUA TCT **GGA** TAT GCT TTT TAT AAA TCA AAA AAT GAA CAA ACA
 AGC TAT TCA GTA AAA GTA ATT ATG TCT TGT **GGA** CTT GCT GGT
 ACA **AUA** TGT AGC TTG **AUA** GAT AAA TTA TTT TGG **GGA** **GGA** AGT
 TTA GAT TTT TTG CAG **AUA** CCA AGC TTT TTT ATT TTC GAC TTA
 AAA GAC TGC TAC CTT ACT GTT GCA GAA **AUA** **AUA** TTT GTT GTC
AUA **GGA** ATT TTG CAT AAT **AGA** GAA **AUA** TCA ATG AAA GAA TAC
AUA TAT TTT TGC TAT CGT CAG TTT AAA **AGA** CAC CAC CAC CAC
 CAC CAC TAA

Amino Acid	Rare Codon	Frequency of Occurrence
Arginine	CGA	0
	CGG	0
	AGG	1
	AGA	3
Glycine	GGA	8
	GGG	2
Isoleucine	AUA	14
Leucine	CUA	2
Proline	CCC	0
Threonine	ACG	0

CD1903 in pEF13

ATG GCT **GGA** GGT GTT AAT ATC **AGG** CAA GTA AAG TCT TTT GTT
 TTT CCA GTT **AUA** TCC TTA **AUA** TTC TTA GAC CAA ATT AGC AAA
 GTT CTT **AUA** **GGA** TTA TTC TTA ATG GAC TTT GAA ATT GAT **AUA**
 ATT **GGG** AAA TTT TTA **AGA** TTC AAT CCT GTT CAA AAT ACA AAT
CUA TCT TAT **GGG** **GGA** AAC TTT ATT GGT ATT **CUA** TCT AAT TTA
 TGG GTA TTG GTA TTG TTT AAC ATT TTA GTT **AUA** TTA GTT ATT
AUA TCT **GGA** TAT GCT TTT TAT AAA TCA AAA AAT GAA CAA ACA
 AGC TAT TCA GTA AAA GTA ATT ATG TCT TGT **GGA** CTT GCT GGT
 ACA **AUA** TGT AGC TTG **AUA** GAT AAA TTA TTT TGG **GGA** **GGA** AGT
 TTA GAT TTT TTG CAG **AUA** CCA AGC TTT TTT ATT TTC GAC TTA
 AAA GAC TGC TAC CTT ACT GTT GCA GAA **AUA** **AUA** TTT GTT GTC
AUA **GGA** ATT TTG CAT AAT **AGA** GAA **AUA** TCA ATG AAA GAA TAC
AUA TAT TTT TGC TAT CGT CAG TTT AAA **AGA** CAC CAC CAC CAC
 CAC CAC TAA

Amino Acid	Rare Codon	Frequency of Occurrence
Arginine	CGA	0
	CGG	0
	AGG	1
	AGA	3
Glycine	GGA	8
	GGG	2
Isoleucine	AUA	14
Leucine	CUA	2
Proline	CCC	0
Threonine	ACG	0

CD2597 in pET28a

ATG GGC **CUA** TAT **AUA** TTA **AUA** **AUA** ATT **CUA** CTC **AUA** GGT TTA
 GAC CAA CTG TCT AAA **AUA** TGG GTA TTG AAT AAT TTG GTG GAT
 GTA TCA ACA **AUA** CCA **AUA** **AUA** AAT AAT GTA TTT CAT TTA ACT
 TAT GTC GAA AAT **AGA** GGT GCA GCA TTT **GGA** TTA TTA CAA AAT
 AAT CAA TGG **AUA** TTT **AUA** ATT GTT GCA TTA CTT GCA ACA GTA
 TTT **GGA** **CUA** TAC TAT CTT AAT ACA **AGG** AAA GTA CAT **AUA** TTT
GGA **AGG** TTG **GGA** ATT **AUA** TTA ATT **AUA** TCT GGT GCA TTG **GGA**
 AAT **CUA** ATT GAT **AGA** GTA **CGA** TTA GGT TTT GTA GTA GAT TAC
 TTC GAC TTT **AGA** ATT **AUA** TGG GAA TAT GTA TTC AAT **AUA** GCT
 GAT GTA TTT GTA GTT GTA **GGA** ACT GTG TTT TTA TGT **AUA** TAT
 GTT TTA TTT TTT GAA AGT AAA AGT **AGG** CTC GAG CAC CAC CAC
 CAC CAC CAC TGA

Amino Acid	Rare Codon	Frequency of Occurrence
Arginine	CGA	1
	CGG	0
	AGG	3
Glycine	AGA	3
	GGA	6
Glycine	GGG	0
	AUA	16
Leucine	CUA	4
Proline	CCC	0
Threonine	ACG	0

CD1903 in pET28a

ATG GGC CAA **GGA** GGT GTT AAT ATC **AGG** CAA GTA AAG TCT TTT
 GTT TTT CCA GTT **AUA** TCC TTA **AUA** TTC TTA GAC CAA ATT AGC
 AAA GTT CTT **AUA** **GGA** TTA TTC TTA ATG GAC TTT GAA ATT GAT
AUA ATT **GGG** AAA TTT TTA **AGA** TTC AAT CCT GTT CAA AAT ACA
 AAT **CUA** TCT TAT **GGG** **GGA** AAC TTT ATT GGT ATT **CUA** TCT AAT
 TTA TGG GTA TTG GTA TTG TTT AAC ATT TTA GTT **AUA** TTA GTT
 ATT **AUA** TCT **GGA** TAT GCT TTT TAT AAA TCA AAA AAT GAA CAA
 ACA AGC TAT TCA GTA AAA GTA ATT ATG TCT TGT **GGA** CTT GCT
 GGT ACA **AUA** TGT AGC TTG **AUA** GAT AAA TTA TTT TGG **GGA** **GGA**
 AGT TTA GAT TTT TTG CAG **AUA** CCA AGC TTT TTT ATT TTC GAC
 TTA AAA GAC TGC TAC CTT ACT GTT GCA GAA **AUA** **AUA** TTT GTT
 GTC **AUA** **GGA** ATT TTG CAT AAT **AGA** GAA **AUA** TCA ATG AAA GAA
 TAC **AUA** TAT TTT TGC TAT CGT CAG TTT AAA **AGA** CTC GAG CAC
 CAC CAC CAC CAC CAC TGA

Amino Acid	Rare Codon	Frequency of Occurrence
Arginine	CGA	0
	CGG	0
	AGG	1
	AGA	3
Glycine	GGA	8
	GGG	2
Isoleucine	AUA	14
Leucine	CUA	2
Proline	CCC	0
Threonine	ACG	0

APPENDIX 10 Alignments of DNA sequencing results for pEF14 and pEF15 to in silico references, further to the information in Figure 5.6 Alignment of the DNA sequences of pEF14 and pEF15 to in silico references to check for sequence errors

Sequence alignment for pEF14. The start and stop codons (ATG and TGA) and represented in capitals, with the His tag (6Xcac) in bold and larger case. In addition, the source of the single potentially variable amino acid seen in the protein alignment is indicated in capitals at pEF14 reference base 5272. The variant amino acid is a result of two sequencing traces (2597_pCheckL and T7R) returning an ambiguous base with the other two traces returning the expected base; it was thus assumed that the base was correct in the plasmid.

```

pEF14 Reference 5041 tttgtttaactttaagaaggagataaccATGggcctatatatattaataataattctac
2597_pCheckL      tttgttttnnctttaagaaggagathhccATGggcctatatatattaataataattctgc
T7R              tttatttaactttaagaaggaaatataccATGggcctatatatattaataataattctac
T7F              tttaacttttaaanaaggagataaccATGggcctatatatattaataataattctac

pEF14 Reference 5101 tcataggttttagaccaactgtctaaaatatgggtattgaataatttgggtggatgtatcaa
2597_pCheckL      tcataggttttagaccaactgtctagaatagggtattgaataatttgggtggacgtatcaa
T7R              tcataggttttagaccaactgtctaaaatatgggtattgaataatttgggtggatgtatcaa
T7F              tcataggttttagaccaactgtctaaaatatgggtattgaataatttgggtggatgtatcaa

pEF14 Reference 5161 caataccaataataaataatgtatttcatttaacttatgtcgaataatagaggtgcagcat
2597_pCheckL      caataccantaatnaaataatgtatttcatttaacttaggtcgaacatagaggtgcagcat
T7R              caataccaataataaataatgtatttcatttaacatatgtcgaataatagaggtgcagcat
T7F              caataccaataataaataatgtatttcatttaacttatgtcgaataatagaggtgcagcat

pEF14 Reference 5221 ttggattattacaaaataatcaatggatatttataattgttgcattacttgcAcagtat
2597_pCheckL      ttggattattacaaaataatcaatggatactgataattgttgcattacttgcNcaggt
T7R              ttggattattacaaaataatcaatggatattaataattgttgcattacttgcNcagtat
T7F              ttggattattacaaaataatcaatggatatttataattgttgcattacttgcAcagtat
2596_pCheckR      tataattnntgattacttgcAcagtat

pEF14 Reference 5281 ttggactatactatcttaataacaaggaaagtacatatatttggaaaggttgggaattatat
2597_pCheckL      ttggactatcctatcttgataacaaggaaagtacatatat
T7R              ttgggctatactatgaaataacaaggaaagtacatatatttggaaaggttgggaattatat
T7F              ttggactatactatcttaataacaaggaaagtacatatatttggaaaggttgggaattatat
2596_pCheckR      ttggactatactatcttaataacaaggaaagtacatatatttggaaaggttgggaattatat

pEF14 Reference 5341 taattatatctggtgcattgggaaatctaattgatagagtacgattaggtttttagtag
T7R              taattatatctggtgcattgggaaatctaattgatagagtacgattaggtttttagtag
T7F              taattatatctggtgcattgggaaatctaattgatagagtacgattaggtttttagtag
2596_pCheckR      taattatatctggtgcattgggaaatctaattgatagagtacgattaggtttttagag

pEF14 Reference 5401 attacttcgactttagaattatatgggaaatgtattcaatatagctgatgtattttag
T7R              attacttcgactttagaattatatgggaaatgtattcaatatagntgatgtattttag
T7F              attacttcgactttagaattatatgggaaatgtattcaatatagctgatgtattttag
2596_pCheckR      attacttcgactttagaattatatgggaaatgtattcaatatagctgatgtatttgn

pEF14 Reference 5461 ttgtaggaactgtgTTTTTatgtatatatgTTTTTatTTTTTgaaagtaaaagtaggctcg
T7R              ttgtaggaactgtgTTTTTatgtatatatgTTTTTatTTTTTgaaagtaaaagtaggctcg
T7F              ttgtaggaactgtgTTTTTatgtatatatgTTTTTatTTTTTgaaagtaaaagtaggctcg
2596_pCheckR      ttgtttggaactgtgTTTTTacgtatatatgtcacattttgagaaaag

pEF14 Reference 5521 agcaccaccaccaccaccATGAgatccggctgctaacaagcccgaaaggaagctgagt
T7R              agcaccaccgccaccaccanTGNgatccggctgctaac
T7F              agcaccaccaccaccaccATGAatccggctgctaacaagcccgaaaggaagctgagt

```

Sequence alignment for pEF15. The start and stop codons (ATG and TGA) and represented in capitals, with the His tag (6Xcac) in bold and larger case. While DNA sequencing was performed with 4 primers, neither pEF15_pCheckR nor pEF15_pCheckL reactions returned sequence traces that could be aligned to the *in silico* reference. The sources of the potentially variant amino acids seen in the alignment in Figure 5.6 are where one sequence trace has returned a different amino acid to the other (red bases). In all cases, one sequence trace returned the expected base, thus it is not expected that there are any sequence differences in the plasmid.

pEF15 Reference	5039	atTTTgTTTaaCTTTaagaaggagatatacc ATG ggccaaggaggtgTTaataTcaggca
T7R		atTTTgTTTaaCTTTaagaaggagatatacc ATG ggccaaggaggtgTTaataTcaggca
T7F		actTTaagaaggagatatacc ATG ggccaaggaggtgTTaataTcnggca
pEF15 Reference	5099	agTaaagTcTTTTgTTTTccagTTataTcCTTaaTattCTTgaccaaattagcaaagT
T7R		agTaaagTcTTTTgTTTTccagTTataTcCTTaaTattCTTgaccaaattagcaaagT
T7F		agTaaagTcTTTTgTTTTccngTTataTcCTTaaTattCTTgaccaaattagcaaagT
pEF15 Reference	5159	TcTTataggattattcTTaatggactTTgaaattgatataaTTgggaaatTTTaaagatt
T7R		TcTTataggattattcTTaatggactTTgaaattgatataaTTgggaaatTTTaaagatt
T7F		TcTTataggattattcTTaatggactTTgaaattgatataaTTgggaaatTTTaaagatt
pEF15 Reference	5219	caatcctgTTcAAAatacaaatctatctTatgggggaaactTTattggtattctatctaa
T7R		caatcctgTTcAAAatacaaatctatctTatgggggaaactTTattggtattctatctaa
T7F		caatcctgTTcAAAatacaaatctatctTatgggggaaactTTattggtattctatctaa
pEF15 Reference	5279	TTtatgggtattggtattgTTTaaCattTTtagTTatattagTTattataTctggatattgc
T7R		TTtatgggtattggtattgTTTaaCattTTtagTTatattagTTattataTctggatattgc
T7F		TTtatgggtattggtattgTTTaaCattTTaCttatattagTTattataTcnggatattgc
pEF15 Reference	5339	TTTTataaatcaAAAAatgaacaaacaagctattcagTaaaagTaatTatgTctTgtgg
T7R		TTTTataaatcaAAAAatgaacaaacaagctattcagTaaaagTaatTatgTctTgtgg
T7F		TTTTataaatcaAAAAatgaacaaacaagctattcagTaaaagTaatTatgTctTgtgg
pEF15 Reference	5399	actTgctggtacaatatgtagctTgatagataaaattattTTggggaggaagTTtagattt
T7R		actTgctggtacaatatgtagctTgatagataaaattattTTggggaggaagTTtagattt
T7F		actTgctggtacaatatgtagctTgatagataaaattattTTggggaggaagTTtagattt
pEF15 Reference	5459	TTTgcagataccaagcTTTTTTattTTTcgactTaaaagactgctacTTactgTTgcaga
T7R		TTTgcagataccaagcTTTTTTattTTTcgactTaaaagactgctacTTactgTTgcaga
T7F		TTTgcagataccaagcTTTTTTattTTTcgactTaaaagactgctacTTactgTTgcaga
pEF15 Reference	5519	aataatattTgTgtcataggaattTTTgcataatagagaaatatcaatgaaagaatacat
T7R		aataatattTgTgtcataggaattTTTgcataatagagaaatatcaatgaaagaatacat
T7F		aataatattTgTgtcannngaattTTTgcataatnagaaatatcaatgaaagaatacat
pEF15 Reference	5579	atattTTTgctatcgTcagTtTaaaagactcgag caccaccaccaccaccACTGA gatcc
T7R		atattTTTgctatcgTcagTtTaaaagactcgag caccaccaccaccaccACTGA ga
T7F		atattTTTgctatcgTcagTtTaaaagactcag caccaccaccaccaccACTGA tatcc

Sequence alignment for pEF15. The sequence alignment is displayed in reverse, thus the start codon is represented by CAT and the stop codon by TCA, both in larger font than the rest of the alignment. The His tag is coded for by 6 repeated GTG codons. The Insert Sequence is the contig of combined DNA sequencing results. There are no DNA sequence differences between the insert and the *in silico* reference, leading to a lack of amino acid sequence differences, indicating successful correct cloning.

```

pEF15 Reference      1 ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCAA
Insert Sequence 1209 .....

pEF15 Reference      61 GGGGTTATGCTAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTT
Insert Sequence 1149 .....

pEF15 Reference      121 TGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGGTGCTCGAGTCTTTTAAACTGACGAT
Insert Sequence 1089 .....

pEF15 Reference      181 AGCAAAAATATATGTATTCTTTCATTGATATTTCTCTATTATGCAAAATTCCTATGACAA
Insert Sequence 1029 .....

pEF15 Reference      241 CAAATATTATTTCTGCAACAGTAAGGTAGCAGTCTTTTAAAGTCGAAAATAAAAAAGCTTG
Insert Sequence 969 .....

pEF15 Reference      301 GTATCTGCAAAAAATCTAAACTTCCTCCCCAAAATAATTTATCTATCAAGCTACATATTG
Insert Sequence 909 .....

pEF15 Reference      361 TACCAGCAAGTCCACAAGACATAATTACTTTTACTGAATAGCTTGTTTGTTTCATTTTTTG
Insert Sequence 849 .....

pEF15 Reference      421 ATTTATAAAAAGCATATCCAGATATAATAACTAATAACTAAAATGTTAAACAATACCA
Insert Sequence 789 .....

pEF15 Reference      481 ATACCCATAAATTAGATAGAATACCAATAAAGTTTCCCCATAAGATAGATTTGTATTTT
Insert Sequence 729 .....

pEF15 Reference      541 GAACAGGATTGAATCTTAAAAATTTCCCAATTATATCAATTTCAAAGTCCATTAAGAATA
Insert Sequence 669 .....

pEF15 Reference      601 ATCCTATAAGAACTTTGCTAATTTGGTCTAAGAATATTAAGGATATAACTGGAAAAACAA
Insert Sequence 609 .....

pEF15 Reference      661 AAGACTTTACTTGCTGATATTAACACCTCCTTGGCCCATGGTATATCTCCTTCTTAAAG
Insert Sequence 549 .....

pEF15 Reference      721 TTAACAAAATTTTCTAGAGGGGAATTGTTATCCGCTCACAATCCCCATAGTGAGT
Insert Sequence 489 .....

```

APPENDIX 12 Permission to reuse material from Paetzel M, Karla A, Strynadka NC, Dalbey RE. Signal peptidases. Chemical reviews. 2002 Dec; 102(12):4549-80 (29), as reprinted for **Figure 1.4 Schematic of type II signal peptidase structure and proteolytic mechanism**

Rightslink® by Copyright Clearance Center

https://s100.copyright.com/AppDispatchServlet



RightsLink®

Home

Create Account

Help



Title: Signal Peptidases
Author: Mark Paetzel,†, Andrew Karla,‡, Natalie C. J. Strynadka,† and, and Ross E. Dalbey*,‡
Publication: Chemical Reviews
Publisher: American Chemical Society
Date: Dec 1, 2002
Copyright © 2002, American Chemical Society

User ID
<input type="text"/>
Password
<input type="text"/>
<input type="checkbox"/> Enable Auto Login
<input type="button" value="LOGIN"/>
Forgot Password/User ID?
If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already a RightsLink user or want to learn more?

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

Copyright © 2014 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#)
Comments? We would like to hear from you. E-mail us at customer care@copyright.com

Bibliography

1. *Clostridium difficile*: Welcome Trust Sanger Institute; 2011; 27/02/2011. Available from: <http://www.sanger.ac.uk/resources/downloads/bacteria/clostridium-difficile.html>.
2. Kutty PK, Woods CW, Sena AC, Benoit SR, Naggie S, Frederick J, et al. Risk factors for and estimated incidence of community-associated *Clostridium difficile* infection, North Carolina, USA. *Emerging infectious diseases*. 2010 Feb;16(2):197-204.
3. Hughes GJ, Nickerson E, Enoch DA, Ahluwalia J, Wilkinson C, Ayers R, et al. Impact of cleaning and other interventions on the reduction of hospital-acquired *Clostridium difficile* infections in two hospitals in England assessed using a breakpoint model. *Journal of Hospital Infection*. 2013;84(3):227-34.
4. Suzuki H, Senda J, Yamashita K, Tokuda Y, Kanosaka Y, Kotaki N, et al. Impact of intensive infection control team activities on the acquisition of methicillin-resistant *Staphylococcus aureus*, drug-resistant *Pseudomonas aeruginosa* and the incidence of *Clostridium difficile*-associated disease. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*. 2013 Dec;19:1047-52.
5. HPS Weekly Report. HPS Scotland, 2013 01/05/2013. Report No.: 2013/18.
6. Tattevin P, Buffet-Bataillon S, Donnio P-Y, Revest M, Michelet C. *Clostridium difficile* infections: do we know the real dimensions of the problem? *International Journal of Antimicrobial Agents*. 2013;42, Supplement 1(0):S36-S40.
7. Hensgens MPM, Keessen EC, Squire MM, Riley TV, Koene MGJ, de Boer E, et al. *Clostridium difficile* infection in the community: a zoonotic disease? *Clinical Microbiology and Infection*. 2012;18(7):635-45.
8. Dial S, Kezouh A, Dascal A, Barkun A, Suissa S. Patterns of antibiotic use and risk of hospital admission because of *Clostridium difficile* infection. *CMAJ*. 2008 Oct 7;179(8):767-72.
9. Gerding DN, Muto CA, Owens RCJ. Treatment of *Clostridium difficile* infection. *Clin Infect Dis*. 2008;46:S32-S42.
10. Heinlen L, Ballard JD. *Clostridium difficile* infection. *The American journal of the medical sciences*. 2010 Sep;340(3):247-52.
11. Walk ST, Young VB. Emerging insights into antibiotic-associated diarrhea and *Clostridium difficile* infection through the lens of microbial ecology. *Interdisciplinary perspectives on infectious diseases*. 2008;2008:125081.
12. Hookman P, Barkin JS. *Clostridium difficile* associated infection, diarrhea and colitis. *World J Gastroenterol*. 2009;15(13):1554-80.

13. Kolling G, Wu M, Guerrant RL. Enteric pathogens through life stages. *Frontiers in cellular and infection microbiology*. 2012;2:114.
14. Xia Q, Williams T, Husted D, Price P, Morrison M, Yu Z. Quantitative analysis of intestinal bacterial populations from term infants fed formula supplemented with fructo-oligosaccharides. *Journal of pediatric gastroenterology and nutrition*. 2012 Sep;55(3):314-20.
15. Enoch DA, Butler MJ, Pai S, Aliyu SH, Karas JA. *Clostridium difficile* in children: Colonisation and disease. *Journal of Infection*. 2011;63(2):105-13.
16. Butt E, Foster J, Keedwell E, Bell J, Titball R, Bhangu A, et al. Derivation and validation of a simple, accurate and robust prediction rule for risk of mortality in patients with *Clostridium difficile* infection. *BMC infectious diseases*. 2013;13(1):316.
17. Rupnik M, Wilcox MH, Gerding DN. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nature reviews Microbiology*. 2009 Jul;7(7):526-36.
18. Mei JM, Nourbakhsh F, Ford CW, Holden DW. Identification of *Staphylococcus aureus* virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. *Mol Microbiol*. 1998 Oct;26:399-407.
19. Tidhar A, Flashner Y, Cohen S, Levi Y, Zauberman A, Gur D, et al. The NlpD lipoprotein is a novel *Yersinia pestis* virulence factor essential for the development of plague. *PLoS one*. 2009;4:e7023.
20. Sutcliffe IC, Russell RR. Lipoproteins of gram-positive bacteria. *Journal of bacteriology*. 1995 Mar;177:1123-8.
21. Kovacs-Simon A, Leuzzi R, Kasendra M, Minton N, Titball RW, Michell SL. Lipoprotein CD0873 is a Novel Adhesin of *Clostridium difficile*. *The Journal of infectious diseases*. 2014 Jan 29;E pub.
22. Kovacs-Simon A, Titball RW, Michell SL. Lipoproteins of bacterial pathogens. *Infect Immun*. 2011;79(2):548-61.
23. Alloing G, de Philip P, Claverys JP. Three highly homologous membrane-bound lipoproteins participate in oligopeptide transport by the Ami system of the gram-positive *Streptococcus pneumoniae*. *Journal of molecular biology*. 1994 Aug 5;241:44-58.
24. Khandavilli S, Homer KA, Yuste J, Basavanna S, Mitchell T, Brown JS. Maturation of *Streptococcus pneumoniae* lipoproteins by a type II signal peptidase is required for ABC transporter function and full virulence. *Mol Microbiol*. 2007 Feb;67:541-57.
25. Vyas NK, Vyas MN, Quioco FA. Crystal structure of *M tuberculosis* ABC phosphate transport receptor: specificity and charge compensation dominated by ion-dipole interactions. *Structure (London, England : 1993)*. 2003 Jul;11:765-74.

26. Sander PR, M.; Walker, B.; Pamini, S.K.; Kroppenstedt, R.M.; Ehlers, S.; Keller, C.; Keeble, J.R.; Hagemeyer, M.; Colston, M.J.; Springer, B.; Böttger, E.C. Lipoprotein processing is required for virulence of *Mycobacterium tuberculosis*. *Mol Microbiol*. 2004;52(6):1543-52.
27. Rezwani M, Grau T, Tschumi A, Sander P. Lipoprotein synthesis in mycobacteria. *Microbiology*. 2007;153:652-8.
28. Okugawa S, Mahtab M, Pomerantsev AP, Sastalla I, Crown D, Gupta PK, et al. Lipoprotein biosynthesis by prolipoprotein diacylglycerol transferase is required for efficient spore germination and full virulence of *Bacillus anthracis*. *Mol Microbiol*. 2012;83(1):96-109.
29. Paetzel M, Karla A, Strynadka NC, Dalbey RE. Signal peptidases. *Chemical reviews*. 2002 Dec;102(12):4549-80.
30. Tjalsma H, Zanen G, Venema G, Bron S, van Dijk JM. The potential active site of the lipoprotein-specific (type II) signal peptidase of *Bacillus subtilis*. *The Journal of biological chemistry*. 1999 Oct 1;274:28191-7.
31. Dev IK, Harvey RJ, Ray PH. Inhibition of prolipoprotein signal peptidase by globomycin. *J Biol Chem*. 1985;260:5891-4.
32. Kiho T, Nakayama M, Yasuda K, Miyakoshi S, Inukai M, Kogen H. Synthesis and antimicrobial activity of novel globomycin analogues. *Bioorganic & medicinal chemistry letters*. 2003 Jul 21;13(14):2315-8.
33. Xiao Y, Gerth K, Muller R, Wall D. Myxobacterium-produced antibiotic TA (Myxovirescin) Inhibits Type II Signal Peptidase. *Antimicrobial Agents and Chemotherapy*. 2012;56(4):2014-21.
34. Farries EC. *Clostridium difficile* *IspA* and *IspA2* mutants display increased susceptibility to malachite green, implicating lipoprotein signal peptidase II as a potential drug target: University of Exeter; 2012.
35. Farries EC, Kovacs-Simon A, Titball R, Minton N, Mitchell SL. Poster: *Clostridium difficile* encodes two *IspA* genes, both of which are involved in lipoprotein processing. SGM Spring Conference 2012; March 26-29; Dublin 2012.
36. Farries EC, Kovacs-Simon A, Titball R, Minton N, Mitchell SL. Poster: Phenotypic characterisation of *IspA* mutants of *Clostridium difficile*. SGM Spring Conference 2013; Manchester 2013. p. Poster 87.
37. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, et al. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat Genet*. 2013 Jan;45(1):109-13.
38. Dale N, Gerding MD. Global epidemiology of *Clostridium difficile* infection in 2010. *Infection Control and Hospital Epidemiology*. 2010;31(S1):S32-S4.
39. Kuijper EJ, Coignard B, Tull P. Emergence of *Clostridium difficile*-associated disease in North America and Europe. *Clinical microbiology and*

infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2006 Oct;12 Suppl 6:2-18.

40. Pépin J, Valiquette L, Cossette B. Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. CMAJ. 2005;173:1037-42.

41. Taori SK, Wroe A, Poxton IR. Clostridium difficile infections in SE Scotland: mortality and recurrence in a region without PCR ribotype 027. Journal of medical microbiology. 2013 Jun 20;E pub.

42. Alcalá L, Martín A, Marín M, Sánchez-Somolinos M, Catalán P, Peláez T, et al. The undiagnosed cases of Clostridium difficile infection in a whole nation: where is the problem? Clinical Microbiology and Infection. 2012;18(7):E204-E13.

43. Summary points on Clostridium difficile infection (CDI) London: Public Health England; 2013 [updated 08 January 2014]. Available from: http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1278944283388.

44. Jones AM, Kuijper EJ, Wilcox MH. Clostridium difficile: a European perspective. The Journal of infection. 2013 Feb;66(2):115-28.

45. Jury LA, Sitzlar B, Kundrapu S, Cadnum JL, Summers KM, Muganda CP, et al. Outpatient healthcare settings and transmission of Clostridium difficile. PloS one. 2013;8:e70175.

46. Chitnis AS, Holzbauer SM, Belflower RM, Winston LG, Bamberg WM, Lyons C, et al. Epidemiology of community-associated Clostridium difficile infection, 2009 through 2011. JAMA internal medicine. 2013 Jul 22;173(14):1359-67.

47. Hirshon JM, Thompson AD, Limbago B, McDonald LC, Bonkosky M, Heimer R, et al. Clostridium difficile infection in outpatients, Maryland and Connecticut, USA, 2002-2007. Emerging infectious diseases. 2011 Oct;17(10):1946-9.

48. Goorhuis A, Debast SB, van Leengoed LA, Harmanus C, Notermans DW, Bergwerff AA, et al. Clostridium difficile PCR ribotype 078: an emerging strain in humans and in pigs? Journal of clinical microbiology. 2008 Mar;46(3):1157.

49. Schoster A, Staempfli HR, Arroyo LG, Reid-Smith RJ, Janecko N, Shewen PE, et al. Longitudinal study of Clostridium difficile and antimicrobial susceptibility of Escherichia coli in healthy horses in a community setting. Veterinary Microbiology. 2012;159(3-4):364-70.

50. Kalchayanand N, Arthur TM, Bosilevac JM, Brichta-Harhay DM, Shackelford SD, Wells JE, et al. Isolation and characterization of Clostridium difficile associated with beef cattle and commercially produced ground beef. Journal of food protection. 2013 Feb;76:256-64.

51. Visser M, Saphri S, Olson N, Du T, Mulvey MR, Alfa MJ. Detection of *Clostridium difficile* in retail ground meat products in Manitoba. *The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie medicale / AMMI Canada*. 2013 Spring;23:28-30.
52. Hoover DG, Rodriguez-Palacios A. Transmission of *Clostridium difficile* in foods. *Infectious disease clinics of North America*. 2013 Sep;27:675-85.
53. Kotila SM, Pitkanen T, Brazier J, Eerola E, Jalava J, Kuusi M, et al. *Clostridium difficile* contamination of public tap water distribution system during a waterborne outbreak in Finland. *Scandinavian journal of public health*. 2013 Jul;41:541-5.
54. Bartlett JG. Antibiotic-associated diarrhea. *Clin Infect Dis*. 1992;15:573-81.
55. *Clostridium difficile*: Map of Medicine; 2013 [updated 30-Sep-2013]; Accessed: 17-Dec-2013. Available from: http://healthguides.mapofmedicine.com/choices/map/clostridium_difficile1.html.
56. Bartlett JG. Narrative Review: The new epidemic of *Clostridium difficile*-associated enteric disease. *Annals of Internal Medicine*. 2006;145(10):758-64.
57. Young VB, Schmidt TM. Antibiotic-associated diarrhea accompanied by large-scale alterations in the composition of the fecal microbiota. *Journal of clinical microbiology*. 2004 Mar;42(3):1203-6.
58. Stewardson AJ, Huttner B, Harbarth S. At least it won't hurt: the personal risks of antibiotic exposure. *Current Opinion in Pharmacology*. 2011;11(5):446-52.
59. Chervonsky AV. Intestinal commensals: influence on immune system and tolerance to pathogens. *Current Opinion in Immunology*. 2012;24(3):255-60.
60. Hasegawa M, Kamada N, Jiao Y, Liu MZ, Nunez G, Inohara N. Protective role of commensals against *Clostridium difficile* infection via an IL-1beta-mediated positive-feedback loop. *Journal of immunology (Baltimore, Md : 1950)*. 2012 Sep 15;189(6):3085-91.
61. Aldape MJ, Packham AE, Nute DW, Bryant AE, Stevens DL. Effects of ciprofloxacin on the expression and production of exotoxins by *Clostridium difficile*. *Journal of medical microbiology*. 2013 May;62:741-7.
62. Dial S, Delaney JAC, Barkun AN, Suissa S. Use of gastric acid-suppressive agents and the risk of community-acquired *Clostridium difficile*-associated disease. *JAMA: The Journal of the American Medical Association*. 2005;294(23):2989-95.
63. Eckmann C, Wasserman M, Latif F, Roberts G, Beriot-Mathiot A. Increased hospital length of stay attributable to *Clostridium difficile* infection in patients with four co-morbidities: an analysis of hospital episode statistics in four

European countries. The European journal of health economics : HEPAC : health economics in prevention and care. 2013 Oct;14(5):835-46.

64. Issa M, Vijayapal A, Graham MB, Beaulieu DB, Otterson MF, Lundeen S, et al. Impact of *Clostridium difficile* on inflammatory bowel disease. Clin Gastroenterol Hepatol. 2007 Mar;5(3):345-51.

65. Ananthakrishnan AN, Issa M, Binion DG. Clostridium difficile and inflammatory bowel disease. The Medical clinics of North America. 2010 Jan;94(1):135-53.

66. Halvorson SA, Cedfeldt AS, Hunter AJ. Fulminant, non-antibiotic associated Clostridium difficile colitis following Salmonella gastroenteritis. Journal of general internal medicine. 2011;26(1):95-7.

67. Denève C, Janoir C, Poliane I, Fantinato C, Collingnon A. New trends in *Clostridium difficile* virulence and pathogenesis. Int J Antimicrob Agents. 2009;33(S1):S24-S8.

68. Wheeldon LJ, Worthington T, Hilton AC, Elliott TS, Lambert PA. Physical and chemical factors influencing the germination of Clostridium difficile spores. Journal of applied microbiology. 2008 Dec;105(6):2223-30.

69. Sorg JA, Sonenshein AL. Bile salts and glycine as cogermnants for *Clostridium difficile* spores. Journal of bacteriology. 2008 Apr;190(7):2505-12.

70. Calabi E, Calabi F, Phillips AD, Fairweather NF. Binding of Clostridium difficile surface layer proteins to gastrointestinal tissues. Infect Immun. 2002 Oct;70(10):5770-8.

71. Kovacs-Simon A. Characterisation of lipoproteins of Clostridium difficile and their role in virulence: University of Exeter; 2013.

72. Heeg D, Burns DA, Cartman ST, Minton NP. Spores of *Clostridium difficile* clinical isolates display a diverse germination response to bile salts. PloS one. 2012;7:e32381.

73. Burns DA, Heap JT, Minton NP. The diverse sporulation characteristics of Clostridium difficile clinical isolates are not associated with type. Anaerobe. 2010 Dec;16:618-22.

74. Cohen SH, Gerding DN, Johnson S, Kelly CP, Loo VG, McDonald LC, et al. Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA). Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America. 2010 May;31(5):431-55.

75. Wilcox MH. Updated guidance on the management and treatment of *Clostridium difficile* infection - 1317138914904. In: England PH, editor. 2013.

76. Weinsch C, Parschalk B, Hasenhüdl M, Hirschl AM, Graninger W. Comparison of vancomycin, teicoplanin, metronidazole and fusidic acid for the

treatment of *Clostridium difficile*-associated diarrhea. Clin Infect Dis. 1996;22:813-8.

77. Allen CA, Babakhani F, Sears P, Nguyen L, Sorg JA. Both fidaxomicin and vancomycin inhibit outgrowth of *Clostridium difficile* spores. Antimicrobial Agents and Chemotherapy. 2012 Jan;57:664-7.

78. Lamontagne F, Labbé AC, Haeck O, Lesur O, Lalancette M, Patino C, et al. Impact of emergency colectomy on survival of patients with fulminant *Clostridium difficile* colitis during an epidemic caused by a hypervirulent strain. Ann Surg. 2007;245(2):267-72.

79. Parkes GC, Sanderson JD, Whelan K. The mechanisms and efficacy of probiotics in the prevention of *Clostridium difficile*-associated diarrhoea. The Lancet Infectious Diseases. 2009;9(4):237-44.

80. Marcu L, Olver I. Tirapazamine: From Bench to Clinical Trials. Current Clinical Pharmacology. 2006;1:71-9.

81. Shah Z, Mahbuba R, Turcotte B. The anti-cancer drug tirapazamine has antimicrobial activity against *Escherichia coli*, *Staphylococcus aureus*, and *Clostridium difficile*. FEMS Microbiology Letters. 2013 2013/10;347(1):61-9.

82. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. New England Journal of Medicine. 2013;368(5):407-15.

83. Burke KE, Lamont JT. Fecal transplantation for recurrent *Clostridium difficile* infection in older adults: A review. Journal of the American Geriatrics Society. 2013 2013/07;61(8):1394-8.

84. Aas J, Gessert CE, Bakken JS. Recurrent *Clostridium difficile* colitis: case series involving 18 patients treated with donor stool administered via a nasogastric tube. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2003 Mar 1;36(5):580-5.

85. Gough E, Shaikh H, Manges AR. Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent *Clostridium difficile* infection. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2011 Nov;53(10):994-1002.

86. Khoruts A, Dicksved J, Jansson JK, Sadowsky MJ. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent *Clostridium difficile*-associated diarrhea. Journal of clinical gastroenterology. 2010 May-Jun;44(5):354-60.

87. Villano SA, Seiberling M, Tatarowicz W, Monnot-Chase E, Gerding DN. Evaluation of an oral suspension of VP20621, spores of nontoxigenic *Clostridium difficile* strain M3, in healthy subjects. Antimicrob Agents Chemother. 2012 Oct;56:5224-9.

88. Nagaro KJ, Phillips ST, Cheknis AK, Sambol SP, Zukowski WE, Johnson S, et al. Nontoxigenic *Clostridium difficile* Protects Hamsters against Challenge

with Historic and Epidemic Strains of Toxigenic BI/NAP1/027 *C. difficile*. *Antimicrob Agents Chemother*. 2013 Nov;57:5266-70.

89. Merrigan MM, Sambol SP, Johnson S, Gerding DN. New approach to the management of *Clostridium difficile* infection: colonisation with non-toxigenic *C. difficile* during daily ampicillin or ceftriaxone administration. *International Journal of Antimicrobial Agents*. 2009;33, Supplement 1(0):S46-S50.

90. Crowther GS, Baines SD, Todhunter SL, Freeman J, Chilton CH, Wilcox MH. Evaluation of NVB302 versus vancomycin activity in an in vitro human gut model of *Clostridium difficile* infection. *The Journal of antimicrobial chemotherapy*. 2012 Jan;68:168-76.

91. Goldstein EJ, Citron DM, Tyrrell KL, Merriam CV. Comparative in-vitro activity of SMT19969, a new antimicrobial agent, against *Clostridium difficile* and 350 Gram-positive and negative aerobic and anaerobic intestinal flora isolates. *Antimicrob Agents Chemother*. 2013 October;57(10):4872-6.

92. Inukai M, Nakajima M, Osawa M, Haneishi T, Arai M. Globomycin, a new peptide antibiotic with spheroplast-forming activity. II. Isolation and physico-chemical and biological characterization. *J Antibiot (Tokyo)*. 1978 May;31(5):421-5.

93. Kiho T, Nakayama M, Yasuda K, Miyakoshi S, Inukai M, Kogen H. Structure-activity relationships of globomycin analogues as antibiotics. *Bioorg Med Chem*. 2004;12:337-61.

94. Inukai M, Takeuchi M, Shimizu K, Arai M. Mechanism of action of globomycin. *J Antibiot (Tokyo)*. 1978 Nov;31(11):1203-5.

95. Maynard-Smith M, Ahern H, McGlashan J, Nugent P, Ling R, Denton H, et al. Recombinant antigens based on toxins A and B of *Clostridium difficile* that evoke a potent toxin-neutralising immune response. *Vaccine*. 2014;32(6):700-5.

96. Yan W, Shin KS, Wang SJ, Xiang H, Divers T, McDonough S, et al. Equine hyperimmune serum protects mice against *Clostridium difficile* spore challenge. *Journal of veterinary science*. 2013 2013/10/18;E pub.

97. Lane JA, Mehra RK, Carrington SD, Hickey RM. The food glycome: A source of protection against pathogen colonization in the gastrointestinal tract. *International Journal of Food Microbiology*. 2010;142(1-2):1-13.

98. Koleva PT, Valcheva RS, Sun X, Ganzle MG, Dieleman LA. Inulin and fructo-oligosaccharides have divergent effects on colitis and commensal microbiota in HLA-B27 transgenic rats. *The British journal of nutrition*. 2012 Nov 14;108(9):1633-43.

99. Kohler P, Bregenzer-Witteck A, Rafeiner P, Schlegel M. Presumably hospital-transmitted *Clostridium difficile* infections based on epidemiological linkage. *Swiss medical weekly*. 2013;143(w13824).

100. Doan L, Forrest H, Fakis A, Craig J, Claxton L, Khare M. Clinical and cost effectiveness of eight disinfection methods for terminal disinfection of

hospital isolation rooms contaminated with *Clostridium difficile* 027. *Journal of Hospital Infection*. 2012;82(2):114-21.

101. Speight S, Moy A, Macken S, Chitnis R, Hoffman PN, Davies A, et al. Evaluation of the sporicidal activity of different chemical disinfectants used in hospitals against *Clostridium difficile*. *The Journal of hospital infection*. 2011 Sep;79:18-22.

102. Vohra P, Poxton IR. Efficacy of decontaminants and disinfectants against *Clostridium difficile*. *Journal of medical microbiology*. 2011 Aug;60:1218-24.

103. Zhang A, Nerandzic MM, Kundrapu S, Donskey CJ. Does organic material on hospital surfaces reduce the effectiveness of hypochlorite and UV radiation for disinfection of *Clostridium difficile*? *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America*. 2013 Oct;34:1106-8.

104. Galvin S, Cahill O, O'Connor N, Cafolla AA, Daniels S, Humphreys H. The antimicrobial effects of helium and helium-air plasma on *Staphylococcus aureus* and *Clostridium difficile*. *Letters in applied microbiology*. 2013 Aug;57:83-90.

105. Barbut F, Yezli S, Otter JA. Activity in vitro of hydrogen peroxide vapour against *Clostridium difficile* spores. *Journal of Hospital Infection*. 2012;80(1):85-7.

106. Malik DJ. The elephant in the room: on the routine use of hydrogen peroxide vapour decontamination systems in health care. *The Journal of hospital infection*. 2013 Apr;83:354-5.

107. Barbut F, Yezli S, Otter JA. Hydrogen peroxide room disinfection: there is no elephant in the room but there's plenty of evidence in the trunk. *The Journal of hospital infection*. 2013 Apr;83:355-6.

108. Hammond GA, Johnson JL. The toxigenic element of *Clostridium difficile* strain VPI 10463. *Microbial Pathogenesis*. 1995;19(4):203-13.

109. Cohen SH, Tang YJ, Silva J, Jr. Analysis of the pathogenicity locus in *Clostridium difficile* strains. *The Journal of infectious diseases*. 2000 Feb;181(2):659-63.

110. Kuehne SA, Collery MM, Kelly ML, Cartman ST, Cockayne A, Minton NP. Importance of Toxin A, Toxin B, and CDT in virulence of an epidemic *Clostridium difficile* Strain. *The Journal of infectious diseases*. 2014 2014/01/01;209(1):83-6.

111. Perelle S, Gilbert M, Bourlioux P, Corthier G, Popoff MR. Production of a complete binary toxin (Actin-specific ADP-Ribosyltransferase) by *Clostridium difficile* CD196. *Infect, Immun*. 1997;65:1402-7.

112. Lyerly DM, Saum KE, MacDonald DK, Wilkins TD. Effects of *Clostridium difficile* toxins given intragastrically to animals. *Infect Immun*. 1985 Feb;47(2):349-52.

113. Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phumoonna T, et al. Toxin B is essential for virulence of *Clostridium difficile*. *Nature*. 2009 Apr 30;458(7242):1176-9.
114. Shin B, Kuak EY, Yoo SJ. Emerging toxin A-B+ variant of *Clostridium difficile* responsible for pseudomembranous colitis at a tertiary care hospital in Korea. *Diag Microbiol Infect Dis*. 2008;60:333-7.
115. Limaye AP, Turgeon DK, Cooksin BT, Fritsche TR. Pseudomembranous colitis caused by a Toxin A- B+ strain of *Clostridium difficile*. *J Clin Microbiol*. 2000;38(4):1696-7.
116. Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, Minton NP. The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature*. 2010;467:711-3.
117. von Eichel-Streiber C, Sauerborn M, Kuramitsu HK. Evidence for a modular structure of the holologous repetitive C-terminal carbohydrate-binding sites of *Clostridium difficile* toxins and *Streptococcus mutans* glucosyltransferases. *J Bacteriol*. 1992;174:6707-10.
118. Voth DE, Ballard JD. *Clostridium difficile* toxins: mechanism of action and role in disease. *Clin Microbiol Rev*. 2005;18:247-63.
119. Jank T, Giesemann T, Arktories K. Rho-glucosylating *Clostridium difficile* toxins A and B: new insights into structure and function. *Glycobiology*. 2007;17(4):15R-22R.
120. Papatheodorou P, Hornuss D, Nolke T, Hemmasi S, Castonguay J, Picchianti M, et al. *Clostridium difficile* binary toxin CDT induces clustering of the lipolysis-stimulated lipoprotein receptor into lipid rafts. *mBio*. 2013;4(3):e00244-13.
121. Popoff MR, Bouvet P. Genetic characteristics of toxigenic *Clostridia* and toxin gene evolution. *Toxicon*. 2013 Dec;75:63-89.
122. Govind R, Dupuy B. Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE. *PLoS Pathog*. 2012;8(6):e1002727.
123. Tan KS, Wee BY, Song KP. Evidence for holin function of tcdE gene in the pathogenicity of *Clostridium difficile*. *Journal of medical microbiology*. 2001 Jul;50(7):613-9.
124. Olling A, Seehase S, Minton NP, Tatge H, Schroter S, Kohlscheen S, et al. Release of TcdA and TcdB from *Clostridium difficile* cdi 630 is not affected by functional inactivation of the tcdE gene. *Microbial Pathology*. 2011 Jan;52:92-100.
125. McKee RW, Mangalea MR, Purcell EB, Borchardt EK, Tamayo R. The second messenger c-di-GMP regulates *Clostridium difficile* toxin production by controlling expression of sigD. *Journal of bacteriology*. 2013 Sep 13;E pub.
126. Carter GP, Douce GR, Govind R, Howarth PM, Mackin KE, Spencer J, et al. The anti-sigma factor TcdC modulates hypervirulence in an epidemic

- BI/NAP1/027 clinical isolate of *Clostridium difficile*. PLoS Pathog. 2011 Oct;7:e1002317.
127. Bakker D, Smits WK, Kuijper EJ, Corver J. TcdC does not significantly repress toxin expression in *Clostridium difficile* 630DeltaErm. PloS one. 2012;7:e43247.
128. Cartman ST, Kelly ML, Heeg D, Heap JT, Minton NP. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. Applied and environmental microbiology. 2012 Jul;78(13):4683-90.
129. Aubry A, Hussack G, Chen W, KuoLee R, Twine SM, Fulton KM, et al. Modulation of toxin production by the flagellar regulon in *Clostridium difficile*. Infect Immun. 2012 Oct;80:3521-32.
130. Martin MJ, Clare S, Goulding D, Faulds-Pain A, Barquist L, Browne HP, et al. The agr locus regulates virulence and colonization genes in *Clostridium difficile* 027. Journal of bacteriology. 2013 Aug;195:3672-81.
131. Rothenbacher FP, Suzuki M, Hurley JM, Montville TJ, Kirn TJ, Ouyang M, et al. *Clostridium difficile* MazF toxin exhibits selective, not global, mRNA cleavage. Journal of bacteriology. 2012 Jul;194(13):3464-74.
132. Rupnik M, Avesani V, Janc M, von Eichel-Streiber C, Delmée M. A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. J Clin Microbiol. 1998;36:2240-7.
133. Rupnik M. *Clostridium difficile* toxinotyping. Methods in molecular biology (Clifton, NJ). 2010;646:67-76.
134. Rupnik M. Toxinotyping of *Clostridium difficile* [updated August 2011]; Accessed: 12/12/2013. Available from: <http://www.mf.uni-mb.si/tox/>.
135. Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, et al. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. Lancet. 2005;366:1079-84.
136. Rupnik M. Heterogeneity of large clostridial toxins: importance of *Clostridium difficile* toxinotypes. FEMS microbiology reviews. 2008 May;32:541-55.
137. Fimlaid KA, Bond JP, Schutz KC, Putnam EE, Leung JM, Lawley TD, et al. Global analysis of the sporulation pathway of *Clostridium difficile*. PLoS genetics. 2013 Aug;9:e1003660.
138. Saujet L, Pereira FC, Serrano M, Soutourina O, Monot M, Shelyakin PV, et al. Genome-wide analysis of cell type-specific gene transcription during spore formation in *Clostridium difficile*. PLoS genetics. 2013 Oct;9:e1003756.
139. Saujet L, Monot M, Dupuy B, Soutourina O, Martin-Verstraete I. The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and

virulence factor expression in *Clostridium difficile*. Journal of bacteriology. 2011 Jul;193:3186-96.

140. Pereira FC, Saujet L, Tome AR, Serrano M, Monot M, Couture-Tosi E, et al. The spore differentiation pathway in the enteric pathogen *Clostridium difficile*. PLoS genetics. 2013 Oct;9:e1003782.

141. Dembek M, Stabler RA, Witney AA, Wren BW, Fairweather NF. Transcriptional analysis of temporal gene expression in germinating *Clostridium difficile* 630 endospores. PloS one. 2013;8:e64011.

142. Burns DA, Heap JT, Minton NP. SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate. Journal of bacteriology. 2010 Feb;192(3):657-64.

143. Adams CM, Eckenroth BE, Putnam EE, Doublet S, Shen A. Structural and functional analysis of the CspB protease required for *Clostridium* spore germination. PLoS Pathog. 2013 Feb;9:e1003165.

144. Permpoonpattana P, Tolls EH, Nadem R, Tan S, Brisson A, Cutting SM. Surface layers of *Clostridium difficile* endospores. Journal of bacteriology. 2011 Dec;193:6461-70.

145. Permpoonpattana P, Phetcharaburanin J, Mikelsone A, Dembek M, Tan S, Brisson MC, et al. Functional characterization of *Clostridium difficile* spore coat proteins. Journal of bacteriology. 2013 Apr;195:1492-503.

146. Sebahia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, et al. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet. 2006 Jul;38:779-86.

147. Giel JL, Sorg JA, Sonenshein AL, Zhu J. Metabolism of bile salts in mice influences spore germination in *Clostridium difficile*. PloS one. 2010;5(1):e8740.

148. Francis MB, Allen CA, Shrestha R, Sorg JA. Bile acid recognition by the *Clostridium difficile* germinant receptor, CspC, is important for establishing infection. PLoS Pathog. 2013 May;9:e1003356.

149. Ramirez N, Liggins M, Abel-Santos E. Kinetic evidence for the presence of putative germination receptors in *Clostridium difficile* spores. Journal of bacteriology. 2010 Aug;192:4215-22.

150. Sorg JA, Sonenshein AL. Inhibiting the initiation of *Clostridium difficile* spore germination using analogs of chenodeoxycholic acid, a bile acid. Journal of bacteriology. 2010 Oct;192:4983-90.

151. Burns DA, Heap JT, Minton NP. *Clostridium difficile* spore germination: an update. Research in microbiology. 2010 Nov;161:730-4.

152. Burns DA, Minton NP. Sporulation studies in *Clostridium difficile*. J Microbiol Methods. 2011 Nov;87:133-8.

153. Burns DA, Heeg D, Cartman ST, Minton NP. Reconsidering the sporulation characteristics of hypervirulent *Clostridium difficile* BI/NAP1/027. *PloS one*. 2011;6:e24894.
154. Li Y, Jin K, Ghosh S, Devarakonda P, Carlson K, Davis A, et al. Structural and Functional Analysis of the GerD Spore Germination Protein of *Bacillus* Species. *Journal of molecular biology*. 2014 Feb 12;E pub.
155. Paredes-Sabja D, Setlow P, Sarker MR. Germination of spores of *Bacillales* and *Clostridiales* species: mechanisms and proteins involved. *Trends in microbiology*. 2010 Feb;19:85-94.
156. Pelczar PL, Setlow P. Localization of the germination protein GerD to the inner membrane in *Bacillus subtilis* spores. *Journal of bacteriology*. 2008 Aug;190:5635-41.
157. Mongkolthanaruk W, Robinson C, Moir A. Localization of the GerD spore germination protein in the *Bacillus subtilis* spore. *Microbiology*. 2009 Apr;155:1146-51.
158. Merrigan MM, Venugopal A, Roxas JL, Anwar F, Mallozzi MJ, Roxas BA, et al. Surface-layer protein A (SlpA) is a major contributor to host-cell adherence of *Clostridium difficile*. *PloS one*. 2013;8:e78404.
159. Dingle TC, Mulvey GL, Armstrong GD. Mutagenic analysis of the *Clostridium difficile* flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters. *Infect Immun*. 2011 Oct;79(10):4061-7.
160. Baban ST, Kuehne SA, Barketi-Klai A, Cartman ST, Kelly ML, Hardie KR, et al. The role of flagella in *Clostridium difficile* pathogenesis: comparison between a non-epidemic and an epidemic strain. *PloS one*. 2013;8(9):e73026.
161. Fagan RP, Fairweather NF. Biogenesis and functions of bacterial S-layers. *Nature reviews Microbiology*. 2014 Mar;12:211-22.
162. Fagan RP, Fairweather NF. *Clostridium difficile* Has Two Parallel and Essential Sec Secretion Systems. *The Journal of biological chemistry*. 2011 Aug 5;286:27483-93.
163. Ryan A, Lynch M, Smith SM, Amu S, Nel HJ, McCoy CE, et al. A role for TLR4 in *Clostridium difficile* infection and the recognition of surface layer proteins. *PLoS Pathog*. 2011 Jun;7:e1002076.
164. Tulli L, Marchi S, Petracca R, Shaw HA, Fairweather NF, Scarselli M, et al. CbpA: a novel surface exposed adhesin of *Clostridium difficile* targeting human collagen. *Cellular microbiology*. 2013 Oct;15:1674-87.
165. Reynolds CB, Emerson JE, de la Riva L, Fagan RP, Fairweather NF. The *Clostridium difficile* cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. *PLoS Pathog*. 2011 Apr;7:e1002024.
166. Spigaglia P, Barketi-Klai A, Collignon A, Mastrantonio P, Barbanti F, Rupnik M, et al. Surface-layer (S-layer) of human and animal *Clostridium*

difficile strains and their behaviour in adherence to epithelial cells and intestinal colonization. *Journal of medical microbiology*. 2013 Sep;62:1386-93.

167. Grin I, Hartmann MD, Sauer G, Hernandez Alvarez B, Schutz M, Madlung J, et al. A Trimeric Lipoprotein assists in Trimeric Autotransporter Biogenesis in Enterobacteria. *The Journal of biological chemistry*. 2013 Dec 25;E pub.

168. Johnston JW, Myers LE, Ochs MM, Benjamin WHJ, Briles DE, Hollingshead SK. Lipoprotein PsaA in virulence of *Streptococcus pneumoniae*: surface accessibility and role in protection from superoxide. *Infect Immun*. 2004;72(10):5858-67.

169. Brautigam CA, Ouyang Z, Deka RK, Norgard MV. Sequence, biophysical, and structural analyses of the PstS lipoprotein (BB0215) from *Borrelia burgdorferi* reveal a likely binding component of an ABC-type phosphate transporter. *Protein Sci*. 2013 Feb;23:200-12.

170. Sutcliffe IC, Harrington DJ. Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes. *Microbiology*. 2002 Jul;148(Pt 7):2065-77.

171. de Greef A, Hamilton A, Sutcliffe IC, Buys H, van Alpen L, Smith HE. Lipoprotein signal peptidase of *Streptococcus suis* serotype 2. *Microbiology*. 2003;149(1399-1407).

172. Bray BA, Sutcliffe IC, Harrington DJ. Impact of *lgt* mutation on lipoprotein biosynthesis and *in vitro* phenotypes of *Streptococcus agalactiae*. *Microbiology*. 2009;155:1451-8.

173. Tjalsma HK, V.P.; Prágai, Z.; Wu, H.; Meima, R.; Venema, G.; Bron, S.; Sarvas, M.; van Dijl, J.M. The role of lipoprotein processing by Signal Peptidase II in the Gram-positive eubacterium *Bacillus subtilis*. *J Biol Chem*. 1999;274(3):1698-707.

174. Yakushi T, Tajima T, Matsuyama S, Tokuda H. Lethality of the covalent linkage between mislocalized major outer membrane lipoprotein and the peptidoglycan of *Escherichia coli*. *Journal of bacteriology*. 1997 May;179(9):2857-62.

175. Denham EL, Ward PN, Leigh JA. In the absence of Lgt, lipoproteins are shed from *Streptococcus uberis* independently of Lsp. *Microbiology*. 2009;155:134-41.

176. Arimoto T, Igarashi T. Role of prolipoprotein diacylglycerol transferase (Lgt) and lipoprotein-specific signal peptidase II (LspA) in localisation and physiological function of lipoprotein MsmE in *Streptococcus mutans*. *Oral Microbiology Immunology*. 2008;23:515-9.

177. Rahman MS, Ceraul SM, Dreher-Lesnack SM, Beier MS, Azad AF. The *lspA* gene, encoding the type II signal peptidase of *Rickettsia typhi*: transcriptional and functional analysis. *Journal of bacteriology*. 2007 Jan;189(2):336-41.

178. Dalbey RE, Wang P, van Dijk JM. Membrane proteases in the bacterial protein secretion and quality control pathway. *Microbiol Mol Biol Rev.* 2012 Jun;76:311-30.
179. Hussain M, Ichihara S, Mizushima S. Accumulation of glyceride-containing precursor of the outer membrane lipoprotein in the cytoplasmic membrane of *Escherichia coli* treated with globomycin. *The Journal of biological chemistry.* 1980 Apr 25;255(8):3707-12.
180. Banaiee N, Jacobs WR, Ernst JD. LspA-independent action of globomycin on *Mycobacterium tuberculosis*. *J Antimicrob Chemother.* 2007;60(2):414-6.
181. Kiho TN, M.; Yasuda, K.; Miyakoshi, S.; Inukai, M.; Kogen, H. Structure-activity relationships of globomycin analogues as antibiotics. *Bioorg Med Chem.* 2004;12:337-61.
182. Gelman E, McKinney JD, Dhar N. Malachite green interferes with postantibiotic recovery of mycobacteria. *Antimicrob Agents Chemother.* 2012 Jul;56(7):3610-4.
183. Alderman DJ. Malachite green: a review. *Journal of Fish Diseases.* 1985;8(3):289-98.
184. Banaei N, Kincaid EZ, Grace Lin SY, Desmond E, Jacobs WRJ, Ernst JD. Lipoprotein processing is essential for resistance of *Mycobacterium tuberculosis* to malachite green. *Antimicrob Agents Chemother.* 2009;53(9):3799-802.
185. Bianco MV, Blanco FC, Imperiale B, Forrellad MA, Rocha RV, Klepp LI, et al. Role of P27 -P55 operon from *Mycobacterium tuberculosis* in the resistance to toxic compounds. *BMC infectious diseases.* 2011;11:195.
186. Henderson ALS, T.C.; Heinze, T.M.; Cerniglia, C.E. Reduction of malachite green to leucomalachite green by intestinal bacteria. *Appl Environ Microbiol.* 1997;63(10):4099-101.
187. Jun W, Liguang L, Hongwei D, Lijuan J, Qiong Z, Zhongbo W, et al. Lipoproteins Binding Malachite Green to Slow the Decolorization of Malachite Green in *Pseudomonas* sp. JT-1. *Microbes Environ.* 2011;26(1):81-3.
188. Xiao Y, Wall D. Genetic redundancy, proximity and functionality of *LspA*, the target of antibiotic TA, in the *Myxococcus xanthus* producer strain. *Journal of bacteriology.* 2014 Jan 3;E pub.
189. Williams D, Li J. Total synthesis of Myxovirescin A1. *Tetrahedron Letters.* 1994;35(29):5113-6.
190. Onishi N, Izaki K, Takahashi H. A macrocyclic antibiotic M230B produced by *Myxococcus xanthus* isolation and characterisation. *J Antibiotics.* 1984 January 1984;37(1):13-0.

191. Hecht V, Vorlop J, Kalbitz H, Gerth K, Lehmann J. Vortex chamber for in situ recovery of the antibiotic Myxovirescin A in continuous cultivation. *Biotechnol Bioeng.* 1987;29(2):222-7.
192. Purdy DOK, T.A.; Elmore, M.; Herbert, M.; McLeod, A.; Bokori-Brown, M.; Ostrowski, A.; Minton, N.P. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol Microbiol.* 2002;46(2):439-52.
193. Keuhne SAH, J.T.; Cooksley, C.M.; Catman, S.T.; Minton, N.P. Clostron-mediated engineering of *Clostridium*. *Methods Mol Biol.* 2011;765:389-407.
194. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. The Clostron: A universal gene knock-out system for the genus *Clostridium*. *J Microbiol Methods.* 2007;70:452-64.
195. Wood-Helie SJ, Dalton HP, Shadony S. Hydrophobic and adherence properties of *Clostridium difficile*. *Eur J Clin Microbiol.* 1986;5(4):441-5.
196. Hussain HA, Roberts, A.P., Mullany P. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Derm) and demonstration that the conjugative transposon Tn916DE enters the genome of this strain at multiple sites. *J Med Microbiol.* 2005;54(2):137-41.
197. Purdy D, O'Keeffe TA, Elmore M, Herbert M, McLeod A, Bokori-Brown M, et al. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol Microbiol.* 2002 Oct;46(2):439-52.
198. Peränen J, Rikkonen M, Hyvönen M, Kääriäinen L. T7 vectors with modified T7/lac promoter for expression of proteins in *Escherichia coli*. *Anal Biochem.* 1996;236:371-3.
199. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome research.* 2004 Jun;14(6):1188-90.
200. Buchan DW, Ward, S.M., Lobley, A.E., Nugent, T.C., Bryson, K., Jones, D.T. Protein annotation and modelling servers at University College London. *Nucl Acids Res.* 2010;38 Suppl:W563-W8.
201. Miles AA, Misra SS, Irwin JO. The estimation of the bactericidal power of the blood. *The Journal of hygiene.* 1938 Nov;38:732-49.
202. Wright A, Wait R, Begum S, Crossett B, Nagy J, Brown K, et al. Proteomic analysis of cell surface proteins from *Clostridium difficile*. *Proteomics.* 2005 Jun;5(9):2443-52.
203. Bordier C. Phase separation of integral membrane proteins in Triton X-114 solution. *J Biol Chem.* 1981;256(4):1605-7.
204. Ng E. Rare Codon Caltor, Programmed by Edmund Ng. 2001.

205. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification*. 2005 May;41(1):207-34.
206. Leskela S, Wahlstrom E, Kontinen VP, Sarvas M. Lipid modification of prelipoproteins is dispensable for growth but essential for efficient protein secretion in *Bacillus subtilis*: characterization of the Lgt gene. *Mol Microbiol*. 1999 Feb;31(4):1075-85.
207. Lawley TD, Clare S, Deakin LJ, Goulding D, Yen JL, Raisen C, et al. Use of purified *Clostridium difficile* spores to facilitate evaluation of health care disinfection regimens. *Applied and environmental microbiology*. 2010 Oct;76:6895-900.
208. Yang G, Zhou B, Wang J, He X, Sun X, Nie W, et al. Expression of recombinant *Clostridium difficile* toxin A and B in *Bacillus megaterium*. *BMC microbiology*. 2008 2008-11-06;8:192.
209. Makoff AJ, Oxer MD, Romanos MA, Fairweather NF, Ballantine S. Expression of tetanus toxin fragment C in *E. coli*: high level expression by removing rare codons. *Nucleic acids research*. 1989 Dec 25;17:10191-202.